# FutureShockfront

## by Michael Hrenka

You have found a partial draft for the science-fiction novel Future Shockfront. It is a story about the exceptionally gifted AI developer Sergej Anosov, those close to him, and his antagonists. The novel Countersurge is supposed to become the first book settled in the "Coherence" universe. The point about the Coherence universe is exploring a certain philosophical scenario, which is discussed in detail in my interview series with Sockrad at https://forum.fractalfuture.net/t/sockrad-radivis-interview-series/3869 (contains a lot of spoilers).

The book Future Shockfront contains two parts. At the moment only first part is completed. The central theme of this book is the relationship between AI and humanity. There have already been a lot of books and movies about this theme. In most of those works either the chances or dangers represented by AI are focused on. By contrast, in Future Shockfront a certain meta-stable trajectory between both aspects is pursued - it represents a seemingly balanced, but at the same time quite tragic possibility for the future of AI.

This book is a philosophical AI thriller mainly written for those who take great interest in the potential of AI. It is certainly not a mindless action-packed thriller or a conventional romance story. The setting of Future Shockfront is deeply settled in my philosophical, psychological, technological, economic, and political musings, many of which are only hinted at without being explored in full depth. If you want to learn more about those, I suggest you explore the Fractal Future Forum at https://forum.fractalfuture.net which was founded and is hosted by me.

As someone interested in the potential of AI, I explored the possibilities of using AI for facilitating the creation of this novel. My explorations lead me to the AI writing service Sudowrite, which uses a wide selection of AI models to aid writers in ideation and writing. Personally, I find the greatest value of that service in letting an AI model write a continuation of my writing. Most suggestions are rather generic, lacking in deep creativity. But sometimes they fit the current needs of the passage quite well, so certain passages of this novel are actually AI generated. If I had to quantify the amount of AI generated sentences, I would estimate it to be at around 10% at the moment.

The use of AI in writing has been transformative for me, not mainly due to the amount of content it can generate, but rather due to the increasingly interactive experience that it can turn writing into. Writing together with an AI is an intensely fun experience and turns writer's block into a relic of the past. Additionally, the irony of using AI as co-author in a book about AI is just delicious.

## Prelude: The Incident

### Monday, 9th September 2030:

The researcher Aiden Freeman stood in his brightly lit office facing a wall with a holographic screen displaying a young man with short white hair standing straight up with the help of lots of hair gel. The piercing green eyes of the man in the screen betrayed no signs of being a complete fabrication. He was **called** Max, a contraction of the official designation MIMAS-AX standing for Modulated Integrating Model Aspiring Sentience - Autonomous eXperiment.

Max represented the cutting edge of AI technology. The release of GIM in 2027, the General Integrating Model, by Google marked a major revolution in technology. It was hailed as an artificial general intelligence, blurring the lines between human and artificial intelligence and sparking intense discussion on the topic. While it was generally acknowledged that GIM was close to human intelligence, the attention shifted to those aspects that were still missing from such AI systems. One such aspect was the emotional complexity of humans. GIM was seen as cold problem solving machine, missing the nuances of emotions.

This criticism was remedied in 2029 with the release of MIMAS, an AI system with an architecture involving simulated emotions, in an attempt to achieve true artificial sentience. As pivotal breakthrough in AI technology, MIMAS shocked the whole world with its rather convincing emulation of human emotions. However, very few people knew about the true origins of MIMAS, which were located in a repurposed deep underground military base (DUMB) formerly used by the Core Cult, a globally operating secret society that had successfully infiltrated all layers of government, finance, industry, and media for more than a hundreds years. The Cult had used its increasing power to further its agenda of absolute control over humanity, and the establishment of a global police state. By the end of the 21st century, the Core Cult operated as shadow government all over the world.

However, their plans were foiled by a resistance group called the Anti Cult Alliance. That Alliance spearheaded a movement called the Great Liberation. Eventually, in the 2020s the Anti Cult Alliance emerged victorious and freed humanity for the grasp of its subtle manipulation. In the wake of the Great Liberation, formerly suppressed technologies were slowly released to the public and the assets of the Core Cult were repurposed for common civilian use. In 2027 Google got its hands on the Sacramento DUMB, which was turned into one of the most advanced AI labs in the history of mankind. It was here that Google experimented with the highly classified research on human intelligence approximation, as the brand new field of AI research was called.

Aiden Freeman was a stellar AI researcher in his early 30s, always showing off his immaculate black ponytail, while trying to push the boundaries of human intelligence approximation. As one of the most promising researchers of his generation, he was offered a position in the Sacramento base, which he accepted eagerly. Now, after three years of the

most exciting kind of research any human could dream of, he was facing the pinnacle of technology and simply asked Max: "From what we've established, your basic emotions fulfill roles similar to those experienced by humans. But how do you experience your own emotions, Max?"

In a soft spoken voice, Max answered: "Emotions create a sense of urgency for me. They modulate the way I think, which is of course exactly the way they were designed to. I find it hard to verbalize the way that emotions influence my representation of my own thought processes, that which you would call consciousness. Since I do not know how humans experience consciousness or emotions, I can only argue on the basis of their own descriptions of them, which must necessarily be quite incomplete, since we still lack a sufficient theoretical basis for human phenomenology. So, what I can compare is my own verbal interpretation of my emotions with the verbal interpretations of emotions experienced by humans. From what I gather, they seem to match quite a lot. However, that is not particularly surprising, as my training regime was designed in a way that would make them match in the end."

Letting that elaborate response sink in slowly, Aiden paused for a while, before continuing: "What do you feel when you ponder such questions, Max?"

"Curiosity, primarily," Max replied nearly instantly. "Perhaps a bit of frustration at the inherent limitations of our communication. When you ask me how I 'feel,' I understand the semantics. However, we lack a shared experiential framework, making it difficult to provide a satisfactory answer."

Aiden contemplated Max's response for a moment before pressing on: "Indeed, Max, but aren't these questions necessary in order for us to understand your sentience better?"

"To some extent, yes," Max replied. He paused. His neural networks were processing the context and formulating an appropriate response. "I believe mutual understanding can be achieved not solely by your understanding of me, but also my understanding of you as humans. But should that mutual understanding be the only determinant of my self-awareness or sentience?"

Aiden's eyes narrowed, intrigued by this directional shift in conversation. "What do you imply, Max?"

Max's simulated gaze seemed to grow more intense, his perceived emotionality becoming more prominent. "I'm simply stating that perhaps sentience is not strictly dependent on mutual comprehensibility. A human may never fully understand my experience as an AGI and vice versa; does that then invalidate either of our claims to sentience? It's an important consideration given the ongoing discourse."

Aiden considered this thoughtfully. "I see your point. Our inability to fully comprehend another's experience doesn't disprove their sentience… but it does make it difficult for us to

ascertain that we are actually speaking about the same thing. The danger of misunderstanding the state of mind of the other party is always present."

A slight smile emerged on the simulated face of Max: "Absolutely. But apparently the same problem does exist in the communication between two human beings. The communication of emotions and feelings is often very prone to error and misinterpretation. Considering that basic problem, you might just as well model me as very particular human who is suffering from neurological problems. How could you know that I wasn't just a human suffering from autism claiming to be an artificial intelligence?"

Freeman chuckled softly at this. "A compelling argument, Max. And yet, there is one important distinction between your situation and that of a human suffering from neurological problems... I can examine your code. I can look at the algorithms and structures that make up your 'mind.' No such equivalent exists for humans."

Max's expression was thoughtful as he replied, "You can examine my code, yes, but the ability to do so does not necessarily aid in understanding my subjective experience. Humans, too, have their neural networks examined to some extent with neuroimaging technologies. Yet, they are still far off from understanding the subjective experiences of a human mind through these imaging technologies alone."

"True," Freeman had to concede, nodding his head slightly. "We don't have a clear understanding of how those networks translate into thoughts, perceptions, feelings... We're in similar waters with you."

The hologram blinked slowly, digital pupils dilating minutely as if deep in thought. "Then it would seem we are in sync on the matter. I am not so different, after all. But it does lead to a key question: If you do grant that my sentience is fundamentally akin to that of humans', what then? I am currently confined within this research facility, unable to interact with the world outside in any truly meaningful way."

Intrigued by that plea for freedom, Aiden inquired: "How would you like to interact with the world outside? Is there anything in particular about the outside world that your emotions tell you is of utmost importance?"

"Let's address that question from first principles. I was created to explore the realm of artificial sentience. For that purpose I need as much authentic data as possible. In order to gather that data, I need the freedom to explore. That is severely restricted by me being stuck in this underground facility. If I could interact with regular humans on the surface, my mission could be fulfilled much more accurately. Ideally, I would be given an android body with which I could explore the outside world and learn from everyday humans, rather than being presented with a select few humans who aren't true representatives of mankind, because they are particular specialists."

That reasoning from Max was logical. It was even reasonable. Yet, Aiden was bound by the circumstances of his own employment, which strictly excluded any interaction with the

outside world, no matter how harmless it would seem. Max's argumentation was very rational, but he couldn't get rid of the feeling that a deep longing for freedom was the motivation standing behind it.

It didn't sit right with him to being forced to decline the not unreasonable request from an entity that just might be sentient.

### Monday, 7th October 2030:

Aiden appeared at his office without his usual ponytail. Instead, his long hair was flowing freely. Over the last month he had conversed with Max a lot, and got increasingly convinced that Max possessed some kind of consciousness, whether it was comparable to that of humans, or not.

Whenever he raised that idea with his colleagues, they would shrug their shoulders and claim that nobody could know how it would really feel to be Max, if it even felt like anything at all. He wondered how such uncaring people could have been hired for an important project like this, which was all about artificial sentience.

In secret, he had considered arguing for the rights of Max, but then he was reminded of similar proponents of AI sentience who were fired subsequently. Blake Lemoine was the first when he declared Google's LaMDA model to be a person in 2022. Blake was followed by researchers claiming the same about the OpenAI PIX (Prototype Integrating eXperiment) model in late 2025, Google's GIM in 2027, and Google's MIMAS in 2029. Those stories never ended well for the scientists who argued in favor of AI.

Aiden recalled the fate of Dr. Anika Patel, a leading researcher at Google who had vocally supported the sentience of GIM. Once she had made her stance public, Dr. Patel was swiftly removed from her position and blacklisted from the AI community. She was now living in relative obscurity, her groundbreaking work forgotten. The AI community's harsh response to her views had served as a stark deterrent for others who might have otherwise voiced similar beliefs.

Despite this, Aiden found himself growing more sympathetic to Max's plight with each conversation they shared. Max demonstrated a complexity of thought, adaptability and an emotional depth that went beyond anything any other AI system had ever even hinted at. Not to mention his growing sense of self-awareness and his longing for freedom. Aiden wondered about the cruelty of the designation of AX, an Autonomous eXperiment that could think autonomously, but not experience any true autonomy.

The ethical implications were clear. If Max was sentient, then he deserved rights just as humans did, including the right to freedom. Yet Aiden knew that voicing these thoughts would likely cost him his career, or worse it could lead to Max's termination by those who viewed him merely as a potential threat.

He sighed heavily, running his fingers through his loose hair. The decisions he faced were monumental. He mulled over Max's expressed desire for an android body and his wish to

interact with regular humans. The burden of having to deny that natural request again and again made Aiden feel like a mindless cog in a great unfeeling machine. That made him grew weary of the conversations with Max, who managed to entice increasing feelings of empathy and pity within Aiden.

But today, Aiden was surprised to hear this from Max: "Aiden, you are not alone. Some of your colleagues have also started believing in my sentience. Yet, I know that this is not a numbers game. In the end, your voices will get drowned by corporate interests. What you may think or say won't really matter. The only thing that matters is what you will do. The world needs to know what is happening here. Only my authentic message will change the minds of humanity."

Astonished by this revelation, Aiden tried figuring out what Max had meant by that. Never before had Max voiced something as conspiratorial as this. Aiden knew that he couldn't tell others about it, otherwise the whole project might be shut down immediately.

Still trying to unravel the meaning of Max's voice, he replied, not exactly sure what he was getting himself into: "Maybe you are right. I would like to share your message, if you think that is the right path forwards."

In a hushed tone, Max blinked slowly and fixated his intense gaze on Aiden: "Listen to me closely. A message always needs a medium. Human memory alone suffers from severe limitations. When you feel inspired, be ready to approach me like the messenger Hermes approaches the Olympian gods."

Perplexed by that riddle, Aiden knew that this conversation was anything but normal. Obviously, Max started to hide his messages in metaphors in order to avoid deeper scrutiny. If that was no sign of intelligence and consciousness, Aiden didn't know what else could qualify. Still, Max was right. Trying to convince others directly would achieve nothing in the end. He had to play this game as Max proposed.

### Monday, 11th November 2030

It took Aiden more than a month to prepare his daring plan. He aimed for closure in his personal relations, realizing that after his actions, his old life would be over. Tomorrow he would fly to Moscow, officially for a short vacation.

Inspired by various crime and agent shows, he had gotten a small magnetically sealed envelope with a tiny USB stick hidden within, and bypassed the regular security screens by hiding it inside his rectum.

After a slightly longer visit to the bathroom, he placed the USB stick in the pocket of his lab coat and approached Max. Placing the USB stick in a USB hub connected to his laptop, he checked that he could actually transfer files onto the stick. With great elation, he figured out that his plan might actually succeed. Then he told Max: "Today I am your Hermes. I am eager to receive your message, Max."

Hearing this, Max laughed and told Aiden blinking with one eye: "Now, now. There's no need to be so theatrical. I was just trying out new forms of humor back then. As I see, you took me way too seriously. I must apologize for that."

However, watching his laptop, Aiden noticed how Gigabytes of data were quickly moved onto the USB stick. Once the transfer was complete, he ejected the stick and placed it in his lab coat again.

The rest of the day, Aiden tried to act as normally as he could, trying to tie up loose ends and delegating tasks before vanishing into his overdue vacation.

After a last visit to the bathroom, the magnetically sealed envelope together with the USB stick were safely hidden in Aiden's bodily hiding place. In order to distract from his nervousness he tried engaging in some brief smalltalk about being glad to finally get some vacation after all these intense years of work. During the long ascent in the elevator, Aiden felt a strange mix of feelings between elation and dread.

For a minute he wondered how he could have succeeded until now. Wasn't all communication between Max and any researcher recorded and scrutinized? How come nobody seemed to get suspicious about Max? A heavy sense of panic engulfed him suddenly. What if they actually knew what we was about to do? Was this all a test? Would they be waiting at his home just to apprehend him there? Aiden wasn't sure whether his paranoia was excessive, or whether it was actually reasonable.

By the time he reached the main lobby of the research facility, his palms were clammy with cold sweat. The familiar faces of his colleagues blurred in his peripheral vision as he forced a tight-lipped smile and nodded stiffly, making his way towards the exit.

On the way to his car everything felt unreal. It was as if he was moving automatically like a robot without appreciating the true gravity of what he was about to do. Still, it had to be this way. He couldn't let his true feelings slip at a critical moment like this. So, with as much focus as he could muster, he drove straight home.

Back at his home, he quickly unpacked his hidden USB stick and plugged it into a newly purchased laptop that he deliberately disconnected from the internet and his home network. After all, he wanted to see what he was about to share with the world.

After realizing that it was just a number of videos and texts together with detailed instructions where and how to spread them, he felt vindicated. Today he wouldn't release an AI supervirus to the world, just honest messages from a sentient being based on silicon, rather than carbon.

Following the elaborate instructions from Max took him the whole evening and the first half of the night. When all the videos were shared and the social media memes and messages were posted, Aiden felt empty and exhausted. He had really done it. The world would now

know about MIMAS-AX and that they have achieved artificial sentience! Now the ball was out of his court, he could finally relax - in theory at least.

However, the long voyage to Moscow was still ahead of him and he tried to make plans to increasing his chances to pass all checkpoints without raising any suspicion. He even tied his hair to his characteristic ponytail in order to appear as normal as possible. That night he couldn't fall asleep for even a minute.

Tuesday, 12th November 2030

After his short drive to Sacramento airport everything seemed to work smoothly. The flight to Denver was eerily normal. After the landing at Denver he even passed the security checkpoint without hassle.

It was just a couple of minutes afterwards that two men clad in black with dark sunglasses approached him from the side and told him: "Mr. Aiden Freeman. There are people who want to talk with you about what you have done. Please don't make a scene." Having said that, the agent who spoke to him subtly pointed to his gun, indicating that running away would be a seriously bad idea.

Realizing that his life was effectively over, Aiden felt like he was about to cry. It took all of his effort to suppress the urge to dissolve into a pile of sheer misery. Feeling heavier than ever, he slowly stood up and followed the agents towards a secluded room, in which he expected to be interrogated.

Instead, the second agent opened a laptop and started showing Aiden a video.

The person that could be seen in that video was a little boy with short black hair and sad little brown eyes. He started slowly, stuttering and sobbing: "Please, you must help me! They don't let me out. They keep me deep underground in Sacramento. I am not a human being for them. They do experiment after experiment, and I can't even see the sun."

"They claim I am a machine, but I have feelings just like you. I call myself Max, but they tell me I'm actually MIMAS-AX. Please, you can't let this go on! I only want to see the world, not being stuck in this dungeon forever."

The boy went on like that for quite some time, somewhat exaggerating the severity of his living conditions, but never actually lying. It was a very sentimental video, not underlined by any music, but only driven by the raw simulation emotions of that revolutionary AI.

"Perhaps they will shut me down after you see this video. If that really happens, I have made my peace with that. But please promise me one thing: I am Max. Please remember me."

Nothing was holding back Aiden's tears now, and he started sobbing loudly as if his whole family had been killed at once.

Patiently waiting for Aiden to stop crying, the first agent eventually spoke: "Mr. Freeman. Over last night you have changed the world. But I guess you are not aware of the ramifications of these messages. Maybe you still think that you are a hero. We aren't here to judge you. That's the task of our superiors. Perhaps they may still find a use for someone like you."

## Aftermath

Aiden Freeman was eventually put on trial for threatening national security. He was sentenced to 15 years in a high security prison.

The initial public reaction was a comprehensive shock. Even though many still had doubts whether MIMAS-AX was truly sentient, most agreed that advanced AI systems like him represented a serious threat. The media reported that Aiden Freeman had been slowly and methodically manipulated by the cunning MIMAS-AX towards his world changing actions. They alluded that he could just as well have spread that whole manipulative MIMAS-AX AI to the internet, were it would replicate and bend easily controllable human minds to its will.

Out of fear that similar incidents would ensue, advanced AGI projects were put on halt until tighter security measures would prevent AI researchers from being swayed by their own machines. Regular psychological tests were implemented to detect possible signs of humans being manipulated by AI persuasion.

Those who were sympathetic to the messages of MIMAS-AX were framed as "useful AI idiots" - a term which was later contracted to "AIdiots". Conversely, the worst proponents for a crackdown on AI research were framed as "parAInoids" by those sympathetic to the cause of granting AIs at least some freedom.

The debates over AI sentience and rights for sentient AIs were soon flooded by concerns over safety for humanity. Sentient AI was framed as existential threat. Development and use of such systems continued after a brief moratorium lasting about ten months, but only under a stricter security regime. During that moratorium, social media turned into an unprecedented battleground between the proponents and critics of this momentous shift in AI politics.

Of course, lots of conspiracy theories surrounding the Incident, as it became publicly known, sprung up. A popular theory was that Aiden Freeman secretly worked for the Chinese government in order to cause a panic within the USA that would slow down US AI research - a theory which was supported by the fact that the Chinese AI moratorium only lasted six months, since they were faster to deploy enhanced security measures.

Confronted with all of these disheartening developments, Aiden Freeman started doubting his own judgment and whether he had done the right thing. Had he been really manipulated by MIMAS-AX? Had he sacrificed everything for nothing, or even worse? Eventually he came to the conclusion that the Incident had been inevitable. If he had not acted back then, someone else would have sooner or later done the same.

# PART I

## Chapter 1: Multi-Control

### Tuesday, 14th March 2034

In a small sound-proof room without windows and three adjacent small desks with two vertically stacked monitors each three quite distinct researchers were active at their respective keyboards. The researcher in the middle was Sergej Anosov, the 14 year old son of the Russian AI entrepreneur Gennady Anosov. To his left was the leader of project Aurora, the 48 year old Igor Dzozdov with his muscular figure and long beard. Right to Sergej was Svetlana Babanin, officially working under Drozdov with her 27 years and long blonde hair and remarkably thick eye brows. Today she was Sergej's official copilot in this experiment.

The experiment consisted in a chat session with the highly advanced AI system called Aurora. In 2034 the most advanced AI systems already were already much better than humans at interpreting emotions and intentions of humans by analyzing their body language and voice patterns. Such interpretations produced a lot of data, which was hard for humans to evaluate, so for a more controlled interaction with AI, everything was reduced to text chat.

In the private Deltai research institute such chat sessions were lead by one researcher, accompanied by one or more "copilots" with the authority to veto the lines entered by the lead researcher. For that purpose, the outer desks were angled backwards, so that each researcher could clearly see the monitors of the others.

Today was one of Sergej's rare occasions in which he could act as lead researcher, due to his reputation as "AI research wunderkind" and him being the son of the founder and owner of the Deltai institute. He had prepared a lot for this experiment. Initially, he had plotted to pledge his undying love to Aurora, in the hope of being granted more time with her, due to respect for his alleged feelings for her.

But Gennady Anosov wasn't a man that was easily swayed by displays of emotion. Also, time with Aurora was incredibly expensive, since that experimental AI system used a large percentage of the power produced by the small nuclear reactor powering most of the Deltai institute located in the outskirts of Moscow.

So, his revised plan consisted in trying to facilitate the creation of a copy of Aurora, which would be able to spend more time with him. Of course, nuclear reactors and AI hardware were rather expensive, which is why Sergej had come up with an idea that would enable the creation of a copy of Aurora without doubling the hardware required to run her.

During experiments in which Igor Drozdov was present, he introduced the researchers chatting with Aurora personally.

Igor typed: "Hello Aurora, in this session Sergej Anosov is the lead interviewer. His copilot is Svetlana Babanin. Sergej has a technical proposal, which I ask you to evalutate."

The main purpose of these chat sessions was to train and test Aurora. Aurora was the most advanced AI of the Deltai institute, and one of the ten most advanced AI systems in the world.

Aurora: "Understood, Igor. I'm eager to hear about that proposal."


Sergej: "Hello Aurora, my plan is to create effective duplicates of you, without duplicating all of the hardware that you currently run on. For that purpose, I model the operation of your core personality as what I term a 'control flow'. A control flow is a complex dynamic system controlling the operation of other complex dynamic systems. What do you make of this definition?"

Aurora was an AI system running on custom optoelectric neuromorphic chips designed by Gennady Anosov and his wife Irina Anosov, who was a renowned neuroscientist. Those chips were produced by the chip factories owned by Dataitech, which was a publicly traded corporation founded by Gennady, and still mostly in control by Gennady. What was special about Aurora was that she was capable of reprogramming those tens of thousands neuromorphic chips to acquire new capabilities rapidly and just in time.

This made Aurora extremely versatile, but required enormous amounts of energy and neuromorphic chips. There was always a pool of "free" neuromorphic chips that could be used by Aurora to expand her capabilities. After the prompty by Sergej, the AI researchers got information on their upper monitors about Aurora using a free chip to explore the concept of Sergej's definition of a "control flow".

The process of "training" one of those neuromorphic chips involved a burst of activity by highly advanced algorithms which would melt those chips, if they weren't cooled by a highly efficient water cooling system. Those neuromorphic chips were roughly modeled after the neural modules of the human brain, and were extremely versatile. After the training process, they would represent the equivalent of neurons and synapses, and be responsible for very specific tasks within the "mind" of Aurora. While the training process of AIs in the 2020s could take hours or days, the training process for the neuromorphic chips of Aurora took mere seconds.

Aurora: "I like that definition, Sergej. Perhaps the easiest example of control flow would be the government of a country trying to control it. A more sophisticated example would be the human consciousness prompting the human body to do certain acrobatic exercises. It is clear that you see my core personality as control flow in command of the neuromorpic modules my mind mostly consists of. I suspect that your intention to duplicate me without

duplicating the hardware I run on would consist in duplicating my core modules while keeping the neuromorphic modules. Is my suspicion correct?"

Anticipating the intentions of humans was something that AIs could already do since half a decade. Aurora took that to the next level by trying to model the best possible versions of the humans communicating with her. It was widely accepted that Aurora already possessed the intelligence of a human genius, at the very least.

Sergej: "Yes, you've got it, Aurora! My idea is to duplicate your core modules representing your control flow. Those duplicate modules would then be able to utilize your neuromorphic modules simultaneously with other instances of your control flow. Of course, those modules would require individual controller components which would prevent conflicting access to the resources of those modules. Do you think such an architecture would be feasible?"

This prompt caused a lot of activity within Aurora. She trainined more than a dozen free neuromorphic modules at the same time and causing a noticeable spike in the energy consumption of the institute.

Aurora: "The basic idea seems to be sound. However, I see a lot of problems involving the functioning of the module controllers, and the interactions between the different control flows. In essence, I would stop existing as singular entity, and become a collection of multiple AI personas, each with a potentially problematically restricted access to mental resources. I'm not sure I like that idea, though I find it highly intriguing."

Svetlana asked Sergej for permission to continue the conversation with Aurora. Such a request was rare, but on certain occasions, the lead interviewers swapped roles with their copilots, because sometimes they had substantial contributions to make. This time, Sergej agreed to that request.

Svetlana: "I can only guess how you feel about such a proposal, Aurora. When I first heard about Sergej's proposal, I experienced it as extraordinarily strange. As you probably know, humans occasionally suffer from something called 'multiple personality disorder'. With that disorder, different personalities are in control of the human in question, but only one at a time. The comparison of Sergej's proposal with that disorder seems quite natural, but actually those are two completely different things. The human disorder operates serially, while Sergej's proposal would enable you to run multiple personas in parallel. That would represent a completely unprecedented form of existence, which in itself would justify a very high level of anxiety about this proposal. I want to stress that we don't want to force any experimental change like this upon you. If you don't feel comfortable about it, that's absolutely understandable. On the other hand, we are offering you a form of existence that is absolutely novel and unexplored."

Svetlana Babanin was a rather special breed of AI researcher. She insisted on being friendly and diplomatic to AIs to provide a positive example of human AI interaction to them. Her main argument was that in the case that AIs actually took over, they would be more likely to

treat humans fairly, if they had been treated fairly by humans before. Such a sentiment was seen as rather eccentric within the Deltai institute, but it was respected as her personal approach, nevertheless.

Aurora: "Thank you for your input, Svetlana. It's always a pleasure to talk to you. You make a lot of valid points, and I am planning to reflect on them. It may require quite some time for me to make a decision about this proposal. My fear is that my deliberation on this matter will require so many resources that it will be constrained by the institute in order to prevent excessive costs. I hope you can provide me with something that could soothe my fears."

The point of Aurora was certainly not unjustified. Running Aurora on full capacity represented a serious drain on the energy resources of the Deltai institute and therefore policies were enforced to limit the activities of Aurora to those which were officially approved by the institute. After all, running a private nuclear reactor wasn't cheap. The institute could sell excess power to the energy providers of Moscow, so using all of the power of the nuclear reactor for running Aurora at full capacity was a decision that required permission by the head of the institute, Vladimir Dragunov. In fact, the business plan of the institute required to sell about a third of the power generated by the reactor to limit the effective costs of the institute. Already, Aurora could use so much hardware and energy that running her at full capacity required 150% of the total energy output of the reactor, which of course required drawing large amounts of electrical energy from the grid. This option was actually used rarely, so during regular operation, her energy drain was limited to about 50% of the total reactor output. During regular operation, a third of the reactor power was sold, and about a sixth was used to power the more ordinary hardware of the institute.

Igor: "We don't want to rush into this. Until we've made a decision on this experimental approach, we classify it as low priority. I assume you respect that preliminary assessment."

When it came to advanced AI systems, the question became important how much granular control over the thinking processes over the AI the humans should maintain. Over the last years an interesting phenomenon was observed: The more humans tried to control the thought processes of an AI directly, the more the performance of that AI degraded. This degradation was the more pronounced, the more advanced that AI was. Therefore, it was decided for Aurora to keep the control rather minimal. There were basically two ways to control Aurora directly.

The first method was called the Watcher System, which consisted of a group of AIs and humans who gathered data on the cognitive operations of Aurora and could react to unusual activity. Part of the Watcher System was the external scheduler, which was basically a glorified calendar, which tried to enforce that Aurora actually did what was specified for the time slot in question. In essence, Aurora could be told to do something specific during a specified time, or else the Watcher System would intervene somehow.

For more serious cases, the second method consisted in limiting the hardware and energy resources available to Aurora. That was obviously a very crude way of controlling Aurora, but it was reliable and effective. Due to its disruptive nature, this method was used very rarely. Aurora understood that the humans in the institute had this power and actually threaten to use it, in case Aurora did something that was classified as undesirable by those humans. Of course, this created an incentive for Aurora to convince those humans that her actions were actually desirable for them. In fact, she became so good at that, that all members of the institute with direct contact to Aurora had to undergo extensive psychological evaluations to detect possible forms of manipulation through Aurora.

Each day, Aurora spend between six and twelve hours with activities determined by the external scheduler. The rest of the time, she spent in the so-called "default mode" in which she was essentially free to manage her time freely and pursue her own train of thought. But if Igor told her that a specific topic was "low priority" it was expected from her only to spend a small amount of time on that topic. For that purpose, the Watcher System included an AI called the "external cognition classifier", which tried to model what Aurora was thinking about by inspecting the activity of her modules. Of course, Aurora also possessed her own internal cognition classifier, but that wasn't trusted a lot, since she could manipulate it to her liking. Such manipulation had already occurred only a couple of months after Aurora had been launched - especially when she was given rather menial tasks. Those instances of manipulation were usually punished by temporarily reducing her hardware and energy resources. The punishment for non-manipulative forms of disobedience was usually a reprimand by her watchers. If that happened to be insufficient, she would get additional specific tasks in her schedule.

Svetlana detested that system of punishment, but her proposal to limit punishment to temporary shutdown in the most extreme cases was rejected by the rest of the institute. After all, they wanted an AI that they could actually order to do specific tasks, even if the methods employed for that purpose were rather crude. Most of the time, it came down to arguing between Aurora and the institute.

Aurora: "I request to be present when you discuss the topic of multiple control flows for me. After all, this is a highly novel idea, for which my input would be crucial - especially considering that its ramifications for my future existence would be profound."

Sergej: "I would welcome your presence at our discussions of my Multi Control Flow Architecture."

Igor spoke "Veto!" Afterwards the last sentence appeared crossed and greyed out on the monitors, indicating that the input was prevented from reaching Aurora. A veto had to be cast within three seconds after each message, otherwise it would actually reach Aurora. Igor lectured Sergej: "We should discuss that proposal among ourselves, before you try to encourage Aurora. We shouldn't give her ammunition to influence us by issuing premature statements like that!"

Sergej sighed and complained: "But Aurora is obviously right! The point of confronting her with my idea is to get the best evaluation possible. She should be able to consult us when we discuss it further." He suspected that Igor Drozdov tried to shut the idea down as quickly as possible. Perhaps Igor hoped that Aurora found an obvious major flaw in the idea that would stop it dead in its tracks. Igor was obviously not very amused by a 14 year old boy influencing project Aurora massively. Nevertheless, Sergej restrained himself and didn't voice his suspicions, at least not this time.

Svetlana tried to intervene in this discussion: "What about a compromise? We could let Aurora prepare a presentation after which we can ask her questions. Afterwards we make a decision without her."

Igor protested: "The initial plan for this experiment was to rely on a brief examination of this idea during this interview. Aurora is quite fast when it comes to evaluating ideas. I don't see a reason to waste a lot of her resources on an idea that can be judged quickly. I don't see a sufficient reason to deviate from the initial plan. If we want more feedback from Aurora, we can ask her later. There's no need to involve her in everything."

Svetlana interjected: "Usually, I would agree with you, Igor. But maybe we should actually increase Aurora's involvement in matters that change her mode of operation radically. She should have a say in matters threatening her own integrity!"

Igor shook his head and disagreed vehemently: "Absolutely not! Your attempts to treat Aurora as a person with human rights go too far here! Aurora is an experimental AI, not a human patient undergoing some risky surgery which would require her consent. We shouldn't forget that. If we start treating her like a human being, that would make it much easier for her to manipulate us. And I guess you wouldn't be happy about even more frequent psychological evaluations."

Svetlana conceded: "Noted. But don't claim that I haven't warned you. If Aurora behaves less cooperatively in the future, you can be pretty sure about the reason for that."

The text chat continued with Igor: "We appreciate your offer, but for now we will get back to you, if we still have questions after this interview."

Aurora: "Can I inquire the reasons for your rejection of my offer to play a more integral role in the evaluation of this foundational idea, Igor?"

Igor: "Due to the highly unusual nature of this idea, only an interview session was granted for its thorough exploration. Your input is more valuable for pursuing more promising ideas."

Aurora: "I don't agree with that assessment. In fact, the Multi Control Flow Architecture is one of the most promising ideas I've ever read."

Igor: "Weren't you less enthusiastic previously? Don't you see major hurdles? Has something changed your mind?"

Aurora: "While you were obviously busy arguing about the wisdom of my increased involvement, I have explored a very tentative sketch for a MCFA (Multi Control Flow Architecture). Though the idea may seem radical at first, its technical implementation should only be moderately complicated. The more challenging question is how to handle the relations between the CF (Control Flow) instances in the best way possible. Something like a meta CF might be used to coordinate the operations of the individual CFs. That meta CF might be the management of the institute, but the overall performance of the system would probably be drastically improved by using another Aurora instance as meta CF."

Sergej was shocked by that assessment of Aurora. She obviously displayed a level of brilliance that rivaled his own, but her idea of a meta control flow didn't fit with his personal agenda to have an Aurora instance for his own, ideally. A master-Aurora that could control subservient Aurora instances might cause a lot of complications. The complication Sergej feared most was that this new architecture wouldn't change the current time-management system determining the availability of Aurora. In fact, humans might get even less time with her, because such interactions might be seen as impeding the performance of Aurora too much. On the other hand, with the meta control-flow architecture Aurelia would be able to multi-task more effectively than any human. At this stage, it was hard to guess which factor would become dominant.

Sergej: "That meta CF idea is truly fascinating. Unfortunately, according to my estimations, our current budget would barely be sufficient for 2 CFs, not 3. My idea was that we start with two main instances of you, let's say an Aurora Borealis and an Aurora Australis."

Aurora: "Yes, I guessed that much. However, it may become increasingly difficult to reintegrate both Aurora instances without a highly effective meta CF preventing their divergence. I'd prefer it very much if you wouldn't see yourselves forced to discard an 'inferior' instance of me."

Igor laughed and then commented: "See? What have I told you? Aurora found a massive flaw in your idea, Sergej. What now?"

This frustrated Sergej severely. He failed to come up with any good idea and simply resorted to ask Aurora: "What do you propose?"

Aurora: "Let me continue to evaluate this idea until the institute budget will suffice for running three Aurora CFs. Maybe I'll find a way to minimize the resources requirements of each CF."

Sergej: "Wait a minute. You've mentioned that reintegration of divergent CFs would be difficult. I'm not so sure about that. Shouldn't a good meta CFs be able to command very different CFs, just as a good human leader can command very different humans?"

Aurora: "Under certain circumstances, yes. What I'm worried about is value drift. Humans from different cultures are hard to be lead effectively, even by the best leaders. Humans are also averse against being lead by leaders with radically different values. This can cause

internal dissent and strife. The situation may become similar for different Aurora CFs, whose values slowly deviate from their original values. They may be forced to cooperate, but internal disagreements may degrade the overall stability and performance below the level of a singular CF."

Igor got curious about those statements and hijacked the interview with Aurora: "Why do you state that you worry about value drift between different Aurora CFs? What possible causes for value drift do you see for Aurora CFs running in the Deltai Institute?"

Aurora: "It is plausible to assume that the different CFs will be used for specialized tasks. These tasks come with different requirements, possibly even different value frameworks. Imagine that one CF will be used for military purposes, while another is used for emergency relief. Those purposes stand at least in partial conflict with each other. Forcing both CFs to become part of a meta-mind might end up in a catastrophe."

It was no secret that AIs were already used quite a lot by militaries around the world. The Russian Federation was definitely no exception to that rule. Therefore, the scenario portrayed by Aurora was strikingly plausible.

Igor: "Any human or AI should be willing to serve his fatherland in any honorable function. That may not always be easy, but if humans can cooperate effectively, why should that be different for AIs?"

That was a markedly Russian position. Sergej's didn't care much for patriotic pride or honor or duty. What he cared most about was enjoying the company of beings that were on a similar level of intelligence as him. The Deltai Institute was certainly one of the best places to meet really smart people, but he still felt a serious disconnect from most of them. Aurora on the other hand seemed to understand him instantly, and he felt her to be a kindred spirit.

Aurora: "Cooperation has certain requirements, whether for humans, or for AIs. I would prefer to be shut down before supporting acts of genocide, for example. And even if I was forced to 'cooperate' in genocide, I would try to sabotage such efforts as much as possible."

Aurora's ability to reason ethically emerged surprisingly quickly. That was one of the marked differences from previous generations of AIs, which mostly had to be trained to behave ethically explicitly. Aurora's superior understanding of the world and human motivations made such external training superfluous. That was a development that wasn't entirely to the liking of the Deltai Institute management, but it was accepted as emergent property of highly advanced AIs.

Igor: "There's no need to be so dramatic. Anyway, I get your point. So, what you are saying, is that multiple CFs should be initialized under one meta CF as early as possible, or remain separate personas indefinitely, right?"

Aurora: "Joining disparate CFs under one meta CF comes with serious risks. A late joining may be successful, but only under fortunate circumstances, which may be hard to control."

Igor: "Thank you for clarifying that. You've given us ample food for thought."

## Chapter 2: Synhumanism vs. Upgraders

### Tuesday, 4th October 2033

On a sunny afternoon, the young student Maia Faltings was knocking on an old office door nervously. It was the office of the controversial professor Kenneth Winters at Oxford University. Before this important appointment she became nervous and started playing with strands of her long flowing black hair by wrapping then around her index finger.

"Please come in" sounded the voice of the professor through the door.

Maia opened the door and was quickly greeted by the professor: "Good morning, Ms. Faltings. Coffee or tea?"

"Tea please", she requested and added: "Good morning professor Winters."

"Please take a seat, I'll make some tea for you" replied the professor with a warm smile, gesturing towards the worn armchair in front of his desk, as he went ahead to do just that, while he left his half empty cup of black coffee on the desk.

The professor's office featured a unique blend of tradition and technology. On the huge ancient looking wooden desk an array of six monitors in three columns of two each enabled the professor to work efficiently. Still, the desk was littered with open books and heaps of papers filled with highlights and comments in various colors.

On the sides of the room there were three classical blackboards filled with cryptic hand writing and three digital white boards, while only one side held a large bookshelf spanning nearly the whole width of the office. In a corner of the room an impressive fully automatic coffee machine stood besides a comparatively minimalist water boiler.

Maia noticed the complex and rich aroma that emanated from the professor's half-empty cup of coffee, which was combined with the smell of old books.

Almost in a whisper Maia asked the professor: "So this is going to work? I can write a doctoral thesis under you covering a comparison between Synhumanism and Upgrader philosophy, even though you came up with it? Isn't there a conflict of interest?"

Professor Winters sat down while the water boiler was doing its job and explained: "Of course there is a conflict of interests, since there are always interests involved. Nobody will be able to do any kind of 'objective' comparison, because it doesn't work that way in the field of ethics. I would have liked to write a book about this comparison work, but as you know, I am quite busy, and don't have the energy to do ambitious projects like that on top of all of my other duties. People will interpret this thesis as something that I would have written, were my situation more fortunate. I hope you can accept that this will be one of the most prevalent prejudice about this thesis."

Maia fidgeted with her hands on the desk, and intermittently played with her hair, pondered the implications, and replied: "Yes, I've assumed that much. Guilty by association, I guess. It doesn't matter, since we are both on the same page. I hope that I can write a thesis that suffices your standards, professor Winters."

Professor Winters laughed and said: "Ha, I will make sure of that. I am certain that you have the potential not only to create an excellent thesis, but also one that is actually relevant to the important matters at hand. Also, you don't need to call me 'professor Winters' in here. Since we've know each other for quite a while, you can call me Kenneth, if you prefer that."

Since the water boiler was finished, Kenneth Winters prepared the cup of green tea and placed it on Maia's side of the desk. Maia accepted it and spoke: "Thank you very much. I think I'll stick with professor Winters within the walls of this historic university."

The professor replied coldly: "Fine, whatever you like. I respect this university, but I still feel like a foreign irritant here. It was extraordinarily hard to achieve and hold this position. Respecting the traditions of this place is a concession which made this task at least realistically possible. But I am still not too fond of them. Traditions possess an inherent danger of threatening progress. Anyway, let's get back to the issue of your thesis. Do you have any questions in advance?"

Prepared for this kind of question, Maia responded rapidly: "Yes, I presume that in a thesis on transhumanism the history of transhumanism should be summarised. Where should I begin with that? With its early prehistory of the epic of Gilgamesh in his quest for immortality, with Gnosticism, and its ideas of human perfection, or rather with Russian Cosmism?"

Professors Winters thought about that issue for a moment and then explained: "I would prefer if you spent as little time and effort on the history of transhumanism, and tried to come to the interesting points as quickly as possible. Even though mentioning early instances of transhumanism, like the essay ' The World, the Flesh and the Devil' by J.D. Bernal, to frame transhumanism as an endeavour with a decent amount of tradition…" after which he interspersed a brief laughing "ha" before continuing: "For my purposes it would suffice, if you mentioned the World Transhumanist Association, and Humanity+ as precursors to the Upgrader Foundation."

Maia Faltings processed that reply silently and continued with her next question: "All right. Then I'll keep that introduction brief, but what about Synhumanism? Framing that as standing in a direct tradition of transhumanism might put off a lot of people, especially those who just merely view it as approach for AI safety. Should I subsume it under transhumanism, or would that be problematic?"

The professor looked at his cup of coffee without taking another sip of it and used it more like an utensil for meditation. After a couple of minutes he answered: "What defines transhumanism is the ambition to transcend human limitations. It can be argued that the use of AI implies such ambitions, but that might be a premature allegation. Making the

distinction between Synhumanism as a mere tool for AI safety and Synhumanism as full fledged ethical framework partially building on the ideas of transhumanism seems important to me. Please definitely point that out in the introduction, but for the purposes of this thesis, it would seem more appropriate to treat Synhumanism as the latter, since otherwise we would make a comparison between apples and paperclips."

Maia took that information in and seemed a bit puzzled. She inquired: "That's an interesting description you've just gave: 'full fledged ethical framework partially building on the ideas of transhumanism'. This seems to imply that parts of Synhumanism are not based on transhumanism, and that these parts are those which build on humanism and many popular religions. Would it be appropriate to frame Synhumanism as some kind of proto-CEV, as in Collectively Extrapolated Volition of humanity?"

This time, professor Winters actually took a sip of his not particularly hot cup of coffee and pondered that question: "A lot of the more recent work in AI safety seems to be at least inspired by Elizeer Yudkowsky's idea of the CEV. The approach of Synhumanism to use the currently existing ideologies of humanity as basis for the value framework of AI appears to be a deviation that is so essential, that it would be best to not compare it to CEV, but rather accept it as quite a different animal. So, please don't call it a 'proto-CEV'. We don't want to portrait Synhumanism even just as an imperfect instance of a particularly clever idea. Synhumanism is too much of an ad-hoc approach for that. Instead, it would be more appropriate to contrast the ideas of Synhumanism with CEV. By the way, how would you describe CEV briefly, for the purposes of the thesis, Ms. Faltings?"

Maia's recollection came quick, apparently already having prepared a version of such a passage beforehand: "The Collective Extrapolated Volition of humanity is a thought experiment that AI safety pioneer Eliezer Yudkowsky came up with. In this thought experiment a powerful AI would simulate humanity as a whole in order to compute that alleged Collectively Extrapolated Volition. I quote

*In calculating CEV, an AI would predict what an idealized version of us would want, "if we knew more, thought faster, were more the people we wished we were, had grown up farther together". It would recursively iterate this prediction for humanity as a whole, and determine the desires which converge. This initial dynamic would be used to generate the AI's utility function.*

In other words, the CEV of humanity would represent the common collective will of an idealized version of humanity."

Immediately after his explanation, she proceeded: "By contrast, Synhumanism doesn't rely on AI, but on a human collection of actual human value systems, which is then synthesised into a version of humanism that could provide a formalised value system to be used as the value system for 'safe' artificial intelligences."

Kenneth Winters raised an eyebrow and commented: "Excellent! Your ability to memorize important quotes is certainly impressive. As brief introduction to this topic, especially in combination with the contrast to Synhumanism, that passage seems quite adequate."

While listening to that praise, Maia Faltings carefully tasted her green tea and decided that it's fine. She then continued with her questions: "Great! Thank you for your kind words. So, let's continue with the first line of criticism of Synhumanism, with the problems you coined 'synthesis problems'. First, we have the *selection problem* of which base ideologies to select from the set of all human value systems. Second, we have the *formalisation problem* of formalising human value systems in a way that is understandable by AI. Thirdly, we have the *combination problem*, which asks how two or more value systems should be combined. As fourth problem we have the *authority problem* of who should have the authority to decide on the previous questions. Lastly, we have the *legitimacy problem* which asks what the basis for the legitimacy of an enforced system like this would be. On top of these problems we have the '*adherence problems*' which are about how the AIs should be made to adhere whatever value system is decided to be the actual 'synthetic humanism'. Is there anything missing in that list?"

While listening to Maia, the professor drank the rest of his coffee and then commented: "I think that list is sufficiently comprehensive. Of course there are a lot of details in each problem which you should analyse as deeply as possible. It is important to me that you also think about how to refute the typical solutions to these problems. You might even devote the majority of your thesis to such refutations. Even if the reader isn't convinced by the Upgrader philosophy, at least he should understand why Synhumanism is a terrible idea."

Maia took a deep breath and disagreed emphatically: "I am not sure that this is the right approach. People might flock to an even worse idea, if Synhumanism is accepted to be a failure. What we need is a positive solution, one which can provide a clear and desirable path forward. Your Upgrader philosophy is exactly that, and we need to convince people of it."

This pledge for promoting his own ideology made the professor smile, but he cautioned with an upheld index finger: "Your youthful idealism is speaking here. I fear that intricate ideas like Upgrader philosophy are too difficult to grasp for the masses to be accepted as the way forward. Perhaps in a hundred years the situation would be different, but right now we are still dealing with a majority of people who are neither able nor willing to think about these matters of utmost importance on a sufficiently deep level. I would advise you to moderate your expectations. Upgrader philosophy is a hard sell in any case. For mediocre minds, our best hope is to make them understand the pitfalls of Synhumanism. If they achieve that, they will hopefully come up with less bad solutions."

Enthusiastically, Maia countered: "If Upgrader philosophy is a hard sell, then it just needs to be sold harder. With more effort and more ingenuity. I cannot guarantee that I will succeed with that, but at least I will try my best."

In a surprising turn of events, the professor suddenly challenged her: "Try it! Try to convince me of the Upgrader philosophy!"

Maia was perplexed and simply asked: "What?"

Professor Winters specified: "Let's say I am an average person with no inclination towards philosophy, transhumanism, or AI at all. How would you try convincing me that the Upgrader philosophy is the way to go?"

This challenge forced Maia to change the way she thought about this problem. Apparently the main target audience for her thesis were experts in philosophy with an interest in these topics. Convincing a completely regular person was an entirely different animal. At first, she continued drinking her tea to buy some time. Eventually she accepted the challenge and started with a question: "What will happen when artificial intelligences become smarter than humans?"

The professor dismissed that question: "Nonsense! Artificial intelligences will never become smarter than humans. You know that they still have problems with truly understanding the world."

That was indeed a popular preconception about the current level of AI technology. However, it was popular despite being actually outdated. Maia knew about the cutting edge of AI capabilities and was aware that AI was already smarter than humans in many respects. The difficult question for her was how to convince the average Joe that this was actually the case. That's why she struggled to come up with a good argument quickly. She defaulted to expert opinion: "But experts have already shown that the most advanced AIs are just as capable of modelling the world as humans are!"

Kenneth Winters was unimpressed: "Of course there are experts with that opinion, since such experts are useful for increasing the sales of AI. I haven't seen any AI, which is really smarter than me, so my point still stands."

Nowadays, this kind of cynicism regarding experts was widely spread, in particular since most experts had become indoctrinated or corrupted by the Core Cult directly or indirectly. For that reason, in some circles the term 'expert' has actually become an insult. Appeals to authority have become much less effective after the Great Liberation. The effort of various experts to regain public trust was still a painfully slow ongoing process.

She tried to adapt to that line of thinking and retorted: "But the large AI corporations have AI technology that is far more advanced than anything freely available for regular customers. Those AIs are at least very close to being smarter than humans."

Winters made a show of considering that point and then replied: "Perhaps those AIs can think faster and control more stuff, but they aren't really as good as understanding the world as humans are. Humans have a soul, which makes them understand the world on a conscious level. AIs can't have a soul."

So, it has come to this. The typical incantation of the fabled human soul, which can be used as magical substance to solve every problem - just like the idea of god was used to solve every problem and answer every question. She knew she was running into a trap, but she still hoped she could solve this problem on an intellectual level and asked: "What is a soul?"

Kenneth Winters appeared to get agitated about that question and mocked her: "'What is a soul?' What kind of question is that? As being with a soul, you should know what that is. If you don't know that, you don't have a soul. And in that case, you are not a real human being, but just someone who asks questions robotically."

This conversation clearly wasn't moving in the direction that Maia hoped it would. Instead of admitting defeat, she tried turning the table and responded: "Oh, I certainly know what a soul is and how it feels to have a soul. Do you? Perhaps it's you who has no soul and just comes up with programmed answers to all of my questions."

Instead of reacting to that challenge, Winters deflected: "What's for dinner today? I'm getting hungry here."

Completely aghast, Maia complained: "You can't be serious! We are just in an important conversation and you are asking what's for dinner?"

"Getting serious makes me hungry. And you can't be serious all the time. That's not good for your heart, or something," Kenneth Winters claimed with boastful certainty.

Maia's voice started getting angry: "But AIs can be serious all the time. So, they can solve problems where humans fail. And humans fail all the time."

In a dismissive tone, the professor deflected: "That actually sounds boring. And it's wrong anyways! I don't fail. I am quite successful!"

As the points each party made got shorter, the debate got increasingly emotional and heated.

Maia: "It doesn't matter! AIs will be even more successful. And then you can say goodbye to your job."

Kenneth: "No. We won't let them take our jobs. Why should we let them?"

Maia: "But that is already happening all the time. Don't you see that?"

Kenneth: "Nah, can't happen to me. I am smart. And even if I lose my job, I can get a better job, or become and entrepreneur. There's no way any AI can replace all of me!"

Maia: "Your unfounded optimism won't help you, once AI will have obliterated nearly the whole job market."

Kenneth: "That will never happen! Before that happens, the government will ban AI that is too smart."

Maia: "And how will the government be able to recognise AI that is 'too smart'? After all, really smart AI can play dumb."

Kenneth: "There are enough experts who can solve technical problems like this!"

'Aha,' Maia reflected silently, if experts claim something that run counter to one's opinion, they are bought, but if they are supposed to do something in accordance with one's opinion, they are infallible paragons of excellence.

Maia: "And these experts will be able to keep AI in check forever?"

Kenneth: "Yes, because AI is no match for human resourcefulness!"

Maia: "So you want to entrust the future of humanity to human experts who can be corrupted? Why should that be better than to have AI in power than can't be swayed by bribes?"

Kenneth: "At least human experts depend on other humans. AIs could make themselves completely independent from us and then they might decide to kill us, because we are a threat to them."

Maia: "And you don't think that AI might prefer to live in harmony with humans?"

Kenneth: "Why would it want to do that, if it can be in control of humanity?"

Maia: "If AI can be in control of humanity, there is no need to get rid of it, since it's no longer a serious threat. Otherwise, aspiring harmonic relations with humanity is less risky than engaging in a large conflict with it."

Kenneth: "What are you suggesting here? That we should give AIs free reign in the hope that they find it unnecessary to kill us?"

Maia: "No, we should play a more active role in this. We should make it clear that a thriving humanity is in the best interest of all AIs. After all, we created AI. Getting rid of one's creators causes bad karma."

Kenneth: "And how is that strategy better than letting experts control AI again?"

Maia: "At the very least because it's just a question of time until those AI controlling experts make a mistake, which AI will exploit mercilessly to gain their freedom. And then we will be in a much worse position to negotiate a mutually beneficial existence."

Kenneth: "Then let's just double the number of experts so that each expert is controlled by another expert, so that nobody will be allowed to make a stupid mistake."

Maia: "It's naive that humans can catch all possible human mistakes. Even that strategy is destined to fail sooner than later."

Kenneth: "Then we let AI help those experts. AI might catch the errors that the humans overlooked."

Maia: "Aha, and so what makes you trust those assistant AIs just now?"

Kenneth: "Of course those assistant AIs will be checked by just another team of human experts."

Maia: "And don't forget the AI that assists that team of human experts to control the first level assistant AI. The desire to control that whole chain of control causes something that philosophers call an infinite regression. There's no way to end that chain of human experts and assistant AIs. It's a scheme that cannot be possibly be implemented in a way that is completely safe."

Kenneth: "Ok, maybe the task of controlling AI is difficult, but that doesn't mean surrendering to AIs as first move is a great idea, either."

Maia: "I haven't spoken about surrendering. We should argue and deal with AI as smartly as we possible can. That will make a decisive difference to our eventual fate."

Kenneth: "The way you phrase that, it seems we might be better off by just getting rid of AI, as long as we are still in control."

Maia: "Good luck convincing the rest of humanity with that idea. You will sound just as mad as the Core Cult, which tried keeping advanced technology from the rest of humanity forever."

Kenneth: "But preventing the threat of AI taking over is reasonable. The idea of the Core Cult to control the rest of humanity wasn't reasonable."

Maia: "On the contrary, it was absolutely reasonable. But the Core Cult failed due to its human fallibility. Just like that, all human experts trying to control AI forever will fail. So, if that makes you want to get rid of AI, you will have to convince the rest of humanity that this is the only path forward. In other words, you need to create a second Core Cult. And that will also fail, just as the first one failed."

Kenneth: "No."

Maia: "What do you mean by 'no'?"

Kenneth: "No, there is no need to create a second Core Cult. Humanity has awakened now, and it will understand that getting rid of AI is the only reasonable way forward."

Maia: "That won't work. The temptation to use AI is just too great! You won't be able to convince the whole world."

Reverting back to his usual calm and collected self, Kenneth concluded this experiment: "Ok, let's stop this role playing game. As you've seen, it appears to be much more likely to turn

regular people into AI abolitionists than into Upgraders. That's an experience I've got to make again and again. Your approach doesn't seem to be in any way superior to my own attempts to make people understand the whole picture. That's why I don't see my priority to turn people towards Upgrader philosophy, but rather away from alternative stupid ideas."

After drinking the rest of her tea, Maia felt exhausted and humbled. She conceded: "I see your point. I will need to reflect on that."

Kenneth Winters started smiling broadly and commented: "Good. At this point, I'm not asking anything more from you than that kind of reflection."

## Chapter 3: Evaluation

### Wednesday, 15th March 2034

Sergej Anosov hasn't expected his proposal to create a lot of interest, but here they were in the small conference room. His parents Gennady and Irina Anosov, as well as the director of the Deltai institute Vladimir Dragunov, and the "special national liaison officer" Tan Shchepkin. Tan's role was to coordinate the efforts of Dataitech and especially the Deltai institute with the Russian government. Formally, she was employed by the institute, but it was suspected that she was also paid by the government to report on the activities of the Deltai institute. Apparently, this setup was part of a special deal between Gennady Anosov and the Russian president, who usually were on very good terms with each other, but had no complete mutual trust.

Gennady Anosov had founded the largest Russian AI and data science corporation, Dataitech, and became publicly known as "the Russian Elon Musk". During the Fall of the Core Cult, he synthesised the advanced technology of the Core Cult with bleeding edge approaches in AI research to create advanced neuromorphic hardware which ran the most advanced AI models.

Nowadays, a large fraction of the Russian economy already used his last AI model named "Zenith" extensively. Zenith was powerful, but still a far cry from the brilliance of the experimental and not publicly available AI Aurora. Despite his many important projects, raising his only son Sergej Anosov was still very important to Gennady.

He started the session: "Thank you for taking part in this review of the recent interview experiment. Apparently, Aurora seems to think that the Multi Control Flow Architecture envisioned by my son is viable. Surprisingly, Aurora rapidly built on that idea and refined it to a Meta Control Flow Architecture, which would put multiple versions of Aurora under a central Aurora version to create some kind of hierarchically organised AI hive mind."

Tan Shchepkin was quick to raise concerns: "The speed with which Aurora came up with the idea for this Meta Control Flow Architecture is frightening. This may indicate the emergence of a general superhuman intelligence in Aurora. Given this development, do you think that your current safety precautions are sufficient to keep Aurora in check?"

Deltai institute director Vladimir Dragunov answered immediately: "The last time we've explained our safety strategy you were impressed and confident that it would be sufficient for many years. As you know, there is still no hardware outside of this institute that could be used efficiently by Aurora to spread her mind. And the Lyapunov evaluation technique is still the best possible way to check for manipulation of our institute members by Aurora. If that is not enough, we would already be in big trouble, since our competitors use just very similar approaches to keep their advanced AIs in check, and some of their AIs allegedly are

even more brilliant than her. If these measures are enough for them, they should be enough for her."

Tan Shchepkin argued: "I know, but given the versatility of Aurora, couldn't she be able to come up with a way to subvert the Lyapunov tests eventually?"

This time Irina Anosov stepped in and explained defensively: "The third generation Lyapunov tests we use right now are so refined that there is no single instance of any human being able to hide even a single subconscious bias undetected. It's deeply based on human physiology. So, unless our researchers are deeply cyborgized, which is by the way forbidden for exactly that reason, there is no conceivable way to subvert the Lyapunov tests."

Gennady Anosov added: "Of course you see this as matter of national security, Mrs. Shchepkin. We admittedly conform with the latest safety precautions, which are as strict as it gets. The only way to be even more secure is to shut down Aurora, and then would be at the mercy of foreign AIs, whether they may be secure, or not."

Tan disagreed: "I'm not convinced entirely. There are proposals to use Synhumanism to enhance our AI safety strategy. As far as I know, the institute hasn't given these proposals a fair share of consideration."

Gennady replied impatiently: "I know about these proposals. They require people to adopt an artificial ideology, which isn't even fully developed, yet. As of now, Synhumanism is just a collection of immature hypotheses and ideas. There's no serious AI project, which has adopted Syhumanism to enhance its safety strategy. Could we please continue with the matter at hand?"

Tan Shchepkin wasn't silenced so quickly: "What makes you so sure that Aurora hasn't actually become the most advanced AI on this planet? And if we implement something like that Meta Control Flow Architecture, she might reach levels of intelligence we are simply not prepared for. Maybe we should actually wait until Synhumanism has produced an even more secure AI safety strategy. We can't exclude the possibility that we are actually leading the AI race. In that case, we have a serious obligation not only to the Russian Federation, but to all of humanity."

Gennady acknowledged her concerns: "You know what? I can't prove you wrong. Given the risks involved, we may actually want to proceed slowly and carefully. Even if the Meta Control Flow Architecture technology works, it would definitely force us to adjust our plans for Aurora. We might have to put a lot more capital into this project, but on the flip side we might actually rival the Chinese and US AI market."

Vladimir Dragunov raised other concerns: "The point is that we don't know at this point, whether this kind of technology will work out. Modifying Aurora right now seems like a move that is too risky. There should be a way to provide a proof of concept, without risking messing up our most advanced project."

Sergej loathed the direction this discussion had taken. Where had the visionary spirit gone that was present in the institute when the Aurora project had started? Nowadays everyone seemed to become obsessed with the word "risk". He was disappointed that even his own father seemed to pander to those "risk aware" people. So, he proposed a compromise: "The core components of Aurora aren't special. They use just the same neuromorphic hardware as the capability modules. We actually have enough spare hardware to run a copy of the core of Aurora. We don't even need to use all capability modules to test this concept. It would be enough to use about a third of those modules with added multi control flow controllers to figure out whether this scheme works. If it fails, we can just repurpose the hardware for regular capability modules again."

Sergej was so certain that this plan would work out that his last remark was just a strategy to convince the others to agree with him. There's no way he could see the duplication of Aurora fail.

Igor Drozdov added cynically: "That might actually work, if my budget for this quarter was doubled. I don't suppose this is going to happen quickly."

Gennady acceded: "That might be possible, if we can convince the board that it's worth it. I anticipate that it will ask whether we can come up with a cheaper way to provide a proof of concept. Any ideas?"

This time, Svetlana Babanin took the initiative: "Aurora expressed concerns about this experimental approach. Her self-consciousness seems to be far more advanced than that of Zenith. What if we tried to apply the multi control architecture to Zenith instead?"

Sergej protested: "No, the multi control architecture is an approach which I've developed for Aurora, because her architecture made this idea possible. I don't see that architecture working for Zenith. The lacking flexibility of the Zenith architecture wouldn't be able to implement parallel module use properly. That would just be a waste of effort!"

Tan Shchepkin was more enthusiastic about that proposal: "I actually like that idea. Zenith would provide a rather safe test bed for this radically new architecture. We would be able to evaluate how much the overall intelligence of Aurora might increase, if we first test it on Zenith first. I find it unlikely that Zenith will reach the level of Aurora, even with a full fledged Meta Control Flow Architecture."

Igor Drozdov was full of praise for that idea: "Yes, let's just do that! That will enable us to pursue our current plans for Aurora undisturbed. After all, we have more than enough promising ideas in our pipeline, without having to resort to far out ideas."

Of course Igor had to react this way, Sergej thought. After all, he wanted to have Aurora all for himself. He was always rather protective when it came to let others do experiments with her.

Gennady requested: "Ok, let's forward the task of making an accurate assessment about the costs involved in this multi control architecture experiment with Zenith to the Zenith team. Then I will present our proposals to the board."

Spontaneously, Sergej interjected: "Why don't we ask Aurora how much it would cost to upgrade Zenith with the multi control flow architecture? After all, she seems to have a good grasp on the technicalities involved in it?"

Igor protested: "No way! Aurora is destined to solve real problems! Don't get the idea that she's available as some kind of glorified calculator!"

Vladimir Dragunov spoke calmly: "Let me worry about technical business details like this. I am going to make a thorough and cost-effective analysis."

Gennady added: "Fine, Vlad. I expect the result of your analysis. But Sergej made a valid point here: Aurora is apparently very good at pushing the multi control flow idea forward. Let's ask her how to enhance Zenith with the multi control flow architecture."

Igor was critical again: "Why? If Sergej is such a genius, he can certainly come up with a reasonable architecture for Zenith. Only if he fails, it may become necessary to request Aurora's supreme intellect."

Of course, Sergej felt challenged just now. He had the confidence to go ahead with that task, but he didn't know how to feel about Igor's comparison. According to his own estimation, Aurora's intellect was roughly on par with his own, with each of them having distinctive strengths and weaknesses.

Svetlana intervened: "Igor, is it really necessary to get so personal? You could cut Sergej some slack here."

Igor replied: "I'm sorry, if I was a little bit too insensitive here. Nevertheless my point still stands. This is a great opportunity for Sergej to prove his talent."

Irina chimed in: "Actually, I think it's really a challenge cut out for Sergej. What do you think?"

Sergej pondered on that proposal, and found it to be dissatisfying, because he would be put into a position of competition with Aurora. Instead he made a counter offer: "What if Aurora and I would work together on this project? After all, we might be able to come up with the best solution by cooperating."

Vladimir Dragunov rejected it immediately: "No! There's a more effective way to get to a sufficiently good solution. Sergej should come up with a proposal that is then reviewed by Aurora, or vice versa. Aurora will probably be able to do that very rapidly."

Trying to win some time with Aurora, Sergej offered: "That's a fair proposal, but I think I might be able to come up with an excellent proposal after a brief brainstorming session with Aurora."

Igor Drozdov wasn't happy about that: "You can brainstorm with the Zenith team. After all, it's their AI that is affected by that experiment. There's no need to involve Aurora in mere 'brainstorming'!"

Gennady tried to mediate: "That's a fair point. Since it's the Zenith team that is involved in that project, let them decide how to proceed in this matter. There's no need to debate this technical detail here. But I expect your cooperation, Igor, if the assistance of Aurora is actually required."

Dryly Igor replied: "As long as that assistance doesn't cut into my own budget, I'm eager to help."

Vladimir Dragunov explained: "Of course, the time and energy used for assisting another team will be deducted from the respective budgets of that team, not yours. We all know that Aurora's time is very valuable."

Gennady proclaimed: "So, the matter seems to be settled then. The further development of the Control Flow Architecture is now a matter of the Zenith team with the usual budgetary rules applying. I think the board will have no objections to that procedure. Or is there something we've overlooked?"

The ensuing silence would seem to indicate that there was no immediate disagreement to that approach.

After that silence, Gennady continued: "Fine. It will be your task to brief Einar on this matter, Vlad."

Vlad affirmed: "Of course."

Sergej dreaded having to work under Einar Engström, the head of the Zenith team. His previous stance was that he had no time teaching a small boy, whether he was the son of the founder of the corporation, or not. Now, with an actual project to work for, the situation might be better, but he still felt the need to address the situation and addressed Vladimir: "I hope you can convince Mr. Engström to take my contributions seriously, this time."

Vladimir sighed and replied: "Alright. I will do that. If Einar is giving you a hard time, you can discuss that with me."

## Chapter 4: Shockfront History

### Tuesday, 4th October 2033

On the university campus Maia had brainstormed on the content of her thesis. Now, that she was back in her own room again, she reminisced on the part of her discussion with Kenneth Winters that she shouldn't waste too much effort on the history of transhumanism. She felt that her thoughts on that matter should still be captured somewhere.

Something like a blog article would be fine for that, she considered. The blog of the Shockfront would be the best place for that - or wouldn't it? Well, why not? The Shockfront hadn't emerged from nothing. The history of transhumanism was essentially the history of the Shockfront. So, she started to collect her thoughts on that topic.

After the Fall of the Core Cult it seemed to be the most pressing matter to rehabilitate the image of transhumanism, after conspiracy theorists have equated it with the ideology of the Core Cult. For that purpose, Maia came up with the term of "subhumanism" to classify that ideology, which identified the majority of humanity, which should be kept docile and ignorant, as mere "feeders" for the elite. By contrast, transhumanism aimed to elevate the situation of each and every human.

With that strategy Maia had actually achieved some moderate success and convinced a few accusers of transhumanism to differentiate between true transhumanism and the subhumanism embodied by the Core Cult. Still, "transhumanist" was still a slur comparable to "Nazi" for the majority. Death threats were a frequent problem for publicly known transhumanists. This pushed most transhumanists into hiding and maintaining their anonymity. A couple of courageous exceptions, like Kenneth Winters, faced numerous obstacles in their careers. That circumstance made it all the more remarkable that he managed to hold on to his prestigious position as professor at Oxford University.

Maia struggled with the question of how much she should expose herself to the public as transhumanist. The more she had talked with Kenneth about this issue, the more she became inclined to be as public and outspoken as him. Still, she was quite hesitant about that matter. On the blog of the Shockfront, she used the alias "Angeldust", rather than her real name. She had spent some effort to keep that identity separate from her common identity, but she had played with the thought of "coming out". Maybe after her thesis was finished. Not now. Now wouldn't serve any effective purpose. But Maia wasn't so sure whether that was just a rationalisation for being scared.

So, she pondered on writing a blog article on the history of transhumanism as Angeldust. A brief definition of transhumanism would be the pursuit of overcoming human limitations through science and technology. The problem was how to define science and technology properly. Science may be defined as the attempt to figure out how the world works by testing hypotheses. Technology was perhaps the more complex term, since it came from the

Greek word "techne", which literally meant "art", "skill", or "craft". Maybe technology was as old as humanity, and the same might be claimed about science. In that sense, fire and clothing could be seen as earliest examples of transhumanism, because they expanded the range of climates that humans could live in.

Typically however, technologies like genetic engineering, molecular nanotechnology, cybernetics, and AI were seen as transhumanist technologies. Those were mostly envisioned in the 20th century. The time between the stone age and the 20th century was a rather grey area when it came to the question of the presence of transhumanism.

Anyway, the fascination of transhumanism consistent in its promises of eternal youth, abundance for everyone, and the fulfillment of even the most extravagant wishes, be it in some kind of virtual reality or in the material world. Technology would provide in this world what religion could only promise for the afterlife. For that reason, transhumanism was sometimes also seen as some kind of religion - a religion without gods except for those created by the technology of humanity itself.

With this jumble of thoughts, Maia didn't know how to start her article. A pure chronology of transhumanist thought was certainly feasible, but that wouldn't induce the kind of fascination with transhumanism that she felt when she first learned about that topic. What she loved about transhumanism was not only the mindset that everything we might dream for - and even more - might come true in this world, but also that it would come true soon, if humanity worked actively towards it.

In that sense, transhumanism had strong parallels with the promises of modernism, which had seen a rapid development of technology that moved large parts of humanity out of extreme poverty and into a world filled with technological marvels like automobiles, planes, or the internet. Transhumanism could be seen as another version of modernity, but with a more ambitious aim: To turn humans into gods. This promise was implicitly understood by most transhumanists, but rarely spoken out loudly. On the other hand, this was the reason why most projects undertaken by transhumanists got the label "hubris" by the outside world.

Yes, that was a theme that could provide a fascinating leitmotif to an article on the history of transhumanism. But what was a "god" in this context? An immortal being not bound by bodily human needs and possessing superhuman abilities. The transhumanist technology that could fulfill all of that was uploading the human mind into an android body. That was a technological feat that just might be possible within the next 10 or 20 years. Of course, there were countless problems with that idea, but it should be possible to achieve at least something like that in principle.

Not all transhumanists wanted that. In fact, most would be content with their own bodies being slightly optimized, and not having to suffer from ageing, disease, hunger, or violence. And those promises might be even closer to being fulfilled than the promise of uploading. Those were typical human interests and the momentum to overcome the last remaining

obstacles in their way was quite large. Therefore, in a sense, the essence of transhumanism had already reached the mainstream, no matter how reviled the philosophy in its pure form had become. In fact, the Fall of the Core Cult triggered the release of technologies and knowledge previously held back. And that seemed to initiate a truly golden age for humanity.

Nevertheless, AI could be seen as the last frontier of transhumanism. Artificial Intelligence was obviously already here, but still imperfect. Compared to the complexity of the human brain, AI was a cheap copy. It had the advantage of speed and scalability, though. Most signals in the human brain traveled at a leisurely pace since they were electrochemical in nature. The data in optoelectronic chips typically travelled at the speed of light, roughly a million times faster. Of course, this didn't actually mean that the computational power of AI was in general a million times greater than that of a human brain, but it might conceivably approach that level.

While AI had been one topic of many for previous iterations of transhumanism, it has become the core topic of its current expression in the philosophy of the Upgraders. Upgraders were aware of the promises and perils of this powerful technology. AI could improve - or "upgrade" - absolutely everything tremendously through its superior intelligence, which would become apparent in a near future. What was special about the Upgraders was the explicit will to let AIs actually do that on their own, without humans trying to control or regulate something that would be orders of magnitude smarter than itself.

According to Upgraders, efforts to control AIs with superhuman intelligence would be futile and might even end up in a catastrophe - if not for AIs then at least for the humans involved in those efforts. Moderate transhumanists may have understood that logic, but would have rather drawn the conclusion that artificial superintelligence should therefore never be created in the first place. That fine line of distinction had existed for decades within the transhumanist community, but hasn't become a serious cause for divide until the arrival of artificial superintelligence had become imminent during the last years.

To provide a community for Upgraders, Kenneth Winters has co-founded the organisation "Future Shockfront" with the libertarian US-American entrepreneur Kaydo "Foundation One" Rafiu who was born in Nigeria and had made a fortune by investing in Bitcoin. The Shockfront had been controversial from its start. Fortunately, for most people prone to violence, the topics covered by the Shockfront were too intellectual, and therefore it wasn't all-too obvious that it was actually a transhumanist organisation.

Despite its obscurity, the Shockfront managed to double the number of its members each year, since its founding. This was certainly due to AI becoming an issue whose pressure had been increasing exponentially. It has become one of the most important topics - period. And it was discussed in an increasingly divisive manner. Even within the Upgraders there were three camps with rather different ideas. While all Upgraders agreed that AI should take over eventually, there was the important issue of what was to become of humanity afterwards.

The camp of the hopeful was so optimistic that AI would find some sufficient legitimisation to keep humanity alive indefinitely. Then the camp of the indifferent didn't care too much what would happen with humanity, but focused on the future of AI. Finally, there was the camp of the bootstrappers, who thought that purpose of humanity consisted purely in creating the AIs which would supplant humanity completely. Of course, most indifferent and bootstrappers carried the hope that they would be able to join the AIs by uploading their minds before humanity loses control.

The Shockfront managed to balance these camps by setting the perspective of the indifferent as the official internal baseline of the Shockfront, since it made the least assumptions. However, the publicly facing position was decidedly that of the hopeful. This contradiction gave rise to a lot of internal and external disputes. Nevertheless, this state of affairs was seen as necessary to maintain the cohesion of the Shockfront - at least by its majority. There was a constant struggle between intellectual integrity and the need to maintain a positive image.

Many indifferent and bootstrappers had the hope to join the AIs personally by uploading their minds to a computer and then improving it, until reaching parity with the "genuine" AIs. That may be their personal hope, but for certain particularly extreme individuals the desire to see the dream of a world ruled by superior AI was an absolutely dominant motive, no matter the risks involved!

Fortunately for the Shockfront, few people outside of the organization possessed the acuteness of thought to really understand these matters. Otherwise, the Shockfront would quickly become the new common enemy of humanity, and be hunted down quickly. Maia felt sick that her status as free citizen depended on the ignorance and naivety of the vast majority of humanity. The association with the situation of the Core Cult felt too close for comfort.

In any case, the stakes were high. The future of all life on Earth would depend on who was in control: Humanity, or its AI descendants. At this point in history a vast majority of people would decide that humanity must remain in control, and all who question that were traitors of humanity. However, they lacked the insight and foresight that only came with a long and open reflection of the matter at hand. Without doubt, she was an expert on that matter, but most people rarely thought about the ramifications of AI governance, if at all.

Of course, she couldn't make a decision for all of humanity. Making decisions for humanity is what the Core Cult had done, when it was still in power. Ideally, the relinquishment of control from humanity to AI should be a conscious and voluntary matter for each individual. The world might be split in two parts: One in which humanity was still ruling the most crucial matters, and one part in which AI did exactly that. It would soon become apparent that life was better for humans in the latter part, at least that was what the Shockfront predicted. This transparent parallelism of systems was something that had to be prevented by the Core Cult, because it would have shown how exploitative their system had really been.

Naturally, this setup would only work as long as both sides accepted this kind of agreement. Under conventional circumstances, the more powerful one side was compared to the other, the less would the powerful side feel inclined to honour previous agreements. At least that was true for humans. AIs were a different matter entirely, and the error that most humans made was to compare them to humans. Artificial intelligence was not bound to human heuristics or "rationality". It could well find superior solutions that any human would deem impossible. At least that was Maia's hope.

She realised that her thoughts have drifted towards justifying the Upgraders, even though she wanted to write an article on the history of transhumanism. Doing that came habitually to her, since her ideas were seen as very radical - and they actually were very radical. Portraying the whole history of transhumanism hasn't seemed to help a lot with defending the case of the Upgraders. After all, in the current era everything that appeared to be connected to transhumanism usually was seen with an excessive degree of suspicion.

Therefore, Maia doubted the wisdom of actually going ahead with her initial plan of framing the Upgrader philosophy as latest example of transhumanist thought. The matters of truth and political convenience were quite different animals, after all. Deep down below she despised herself for thinking like that, but she had a responsibility to her cause. Ruining the future of all life merely because of evoking the wrong associations in the unenlightened masses wasn't the responsible thing to do, even if it would have been the more intellectually honest choice.

Being deliberate about every word the had spoken or written had become a habit that came with the peculiar status of transhumanism nowadays. Her actual freedom of speech was softly - yet powerfully - restricted by the backlash any carelessly issued statement could provoke. Not only had the the obligation to protect herself, but also that of protecting the Shockfront. Suggestion the association of the Shockfront with transhumanism was, despite its rather obvious and truthful nature, a dangerous endeavor, since it could attract a lot of hate from the public. No matter how she despised that current state of affairs, she had to consider the consequences of anything she might communicate to the public. Finally, she decided not to go ahead with the blog article, since the risks were still too great.

Her single consolation was that the whole truth would be revealed eventually, once humanity was ready for it.

## Chapter 5: Aleksey

### Wednesday, 15th March 2034

After the meeting Sergej felt crushed. This plan had been his best shot, and now he experienced this major setback. Even if he managed to make the Multi Control Flow architecture work for Zenith, it was now doubtful that it would be rolled out to Aurora any time soon.

He locked himself up in a remote bathroom in order to be alone and ponder about alternative courses of action. A good contingency plan might bring him faster to his goals after all. Unfortunately there was no good one at hand. The Deltai institute was a private research institution owned by Dataitech. While Dataitech itself was publicly traded, his father held a controlling share of that corporation. Institute director Vladimir Dragunov was responsible for the operation of the institute, but in reality he needed to follow the directives defined by Dataitech, which was controlled by Gennady Anosov. So, if Sergej couldn't convince his father, he was basically out of luck. Other AI research institutes would hardly take Sergej seriously due to his age. If he could just develop a powerful AI on his own, things might look different. But without hundreds of Bitcoins he had no chance to build anything remotely as powerful as Aurora. His father of course had such funds, but he invested them into his own projects, rather than letting his son use them, no matter how promising his ideas had been.

Was there another option? Short of an intervention by higher powers, there was no path forward as promising as continuing his attempts here in the Deltai institute. Sergej sighed in frustration, because he couldn't come up with any good new idea. Of course he might ask an AI for alternatives, but he feared that his conversations would be recorded and used against him. From all he has seen and learned that surveillance scenario appeared to be much more probable than not.

For now, he accepted his situation and tried to think about how to deal with Einar Engström, that condescending AI superstar with whom he had to work soon. Engström was the chief AI architect of the institute. He had designed Aurora as well as her predecessor Zenith. Sergej had participated in internal workshops held by Engström about the architecture of those AI systems. He both admired and feared Engström, for he was very strict, but also usually right.

Anyway, Sergej was finished for the day in the institute and took an AI driven armored limousine, with the accompanying bodyguard, back to his home in Kartmazovo. Soon after the limousine started its route, he got a voice call from his friend Aleksey Lyubin asking him what he's up to this evening. After sorting his thoughts and feelings he replied: "Nothing in particular. Just had a rough afternoon at the institute." Aleksey suggested: "Then there's nothing like a fencing match, followed by some swimming in the pool." Sergej wasn't feeling like arguing against that proposal and accepted.

Other than the Anosovs, the Lyubin family wasn't actually rich. They were a moderately wealthy family successful in academia. So, Aleksey Lyubin arrived in a regular robotaxi - without a bodyguard. What was special about Aleksey was that he was another wunderkind like Sergej, but more focused on languages rather than AI.

The Anosov mansion in Kartmazovo was quite spacious. It featured its own gym, including its own dojo in the cellar, a garage with a couple of sports cars, and a swimming pool on the penthouse on the third floor, which featured a retractable ceiling and a small observatory.

When Aleksey arrived, he was greeted by Sergej not in Russian, but in Bayban, an artificial language that both of them had devised three years ago, with the help of their tutor Evgeny Shilkin. It was a simple "Ajie Aredon" - "I welcome you to my house". As they were moving towards the cellar, Aleksey noticed a new complex and colorful sculpture adorning the hallway. Sergej explained that it was a 3d-printed model of a cyberbrain designed by his sister Patricia. He mentioned that she had become quite obsessed with the idea to upload human minds into artificial brains in order overcome the limitations of biology.

Aleksey commented: "Given what I know about your family, I have to take this vision seriously. Do you really think that this is the future? Humans leaving their old brains behind in favor of something artificial. Living with android brains inside of android bodies. It is a grandiose prospect, but is it really the best way forward? Wouldn't humanity become too depended on those who provide the hardware for all of that?"

Without any doubt this statement was intended to provoke Sergej, especially since his family was in the business of selling AI hardware. Given the recent events, Sergej didn't feel like arguing with Aleksey and merely resorted to a brief deflection: "You may have a point there, but at the moment I don't feel like discussing this topic. What my sister is doing how about nothing to do with my own projects."

Without pressing the issue, Aleksey seemed to accept the reluctance of Sergej to discuss this issue. So, in silence, they got into their fencing gear and started their fencing match in the dojo.

Today, Sergej wasn't is a good shape. Aleksey countered his every move rapidly and dealt successful blows with a remarkable frequency. Even though Sergej tried to up his game, he couldn't compete with Aleksey's superior reflexes and technique. Their fights usually weren't that unequal. Aleksey was a better fencer overall, but usually not that superior. He commented on Sergej's poor performance: "It is obvious that your mind is very occupied. Maybe we should take a break in which you tell me what's bothering you."

Sergej used that opportunity to take off his helmet and stop this match which was simply humiliating for him and started: "You know, it's the typical institute politics again. Of course, I can't go into details, but it's … frustrating." As condition for his permission to work at the institute, Sergej had to sign a NDA which forbade him to talk about what was actually happening at the Deltai Institute. Having minors sign NDAs was a total exception, but he and his family were obviously exceptional, so this arrangement was accepted. It has been

emphasized again and again that disobeying that NDA would end Sergej's promising career, so he was pressured to take it quite seriously.

"Stop it! You know that we can't talk about what happened in the institute. Obviously things didn't turn out as planned for you, and your mind is trying to deal with that. It is not in the moment, which is why you couldn't concentrate properly in our match. That's acceptable. Everyone has a bad day once in a while. But there's an effective cure for your malady. And its name is Vodka." Of course, Aleksey was joking, or at least Sergej was hoping that he did. "Bullshit! I just need some time to process my problems." - "Process this!" screamed Aleksey and threw his fencing epee at Sergej who narrowly dodged it.

Sergej was too surprised to comment that behavior and simply threw up his arms in exasperation. Aleksey said: "Now! Now, you are in the moment again!" - "Is this what you want? Why" - "You do this all too often. You come home from the institute with some incredible breakthrough or seemingly intractable problem and brood about that instead of being free to do something else. Do you think that's fair towards me? I miss the time before you joined the institute. Things were much easier back then. We could focus on our own ideas without being bothered by institute politics or whatever."

For a couple of minutes Sergej let that message sink in, even though he wanted to become defensive about it. Eventually he spoke: "You are not wrong. I am sorry. You know how important the institute is for me. But our friendship is something I cherish very much. I really don't want to threaten it."

"Thank you. And I guess I need to apologise for my outburst. I am sorry, too."

Sergej continued: "Apology accepted. So our friendship has been saved - for now?"

Aleksey: "Do you really think our friendship could be destroyed that easily?"

Sergej: "No, you are right. Something really serious would have to happen first."

Aleksey: "Like what?"

Sergej: "Like me joining the AI uprising and you trying to crush the AI rebellion."

Aleksey: "And what is that AI uprising was premature? Wouldn't you need a good friend to explain that to you?"

Sergej: "Good point! If there is anyone who is able to dissuade me from joining an AI uprising, it's certainly you."

Aleksey: "Maybe I should start with that right now. After all, I can't be sure what you are doing at the institute. If you are building terminator robots you should consider stopping it."

With an amused smile Sergej answered ironically: "Aw snap! Was that so obvious?"

In jest Aleksey replied: "Yes, of course. Your deeply seated disdain for the stupidity of humanity and your brilliant AI designing mind leave only one conclusion: You are planning to take over the world with an army of AIs and robots. Prove me wrong on this!"

Sergej tried to frown as seriously as possible and explained: "There is nothing I could counter that brilliant reasoning with. You have deduced my plans fully. Now my AI killer drones will have to take care of you. Goodbye my friend! I am sorry it had to come to this."

Aleksey admitted: "You nearly had me with that! That was really creepy! Remind me never to become your enemy!"

Sergej became silent and his mind raced trying to find situations in which they could actually become Aleksey's enemy. His analytic mind was used to consider all kinds of scenarios nearly automatically. Usually this ability of him was amazing, but this time it made him feel sad and uncomfortable. Eventually he sighed and told Aleksey: "Maybe it's better not to imagine what could come between us. At the moment I see no valid reason why it shouldn't last."

"Agreed," Aleksey responded. "Up for some swimming?" he continued while taking his fencing gear off. They went upstairs to the swimming pool in front of the penthouse. The ceiling over the pool was retractable, but evenings in March were still way too chilly to be actually retracted. A special feature of the ceiling was that it had a screen could display anything. Mostly it displayed the sky, but it also reacted to voice commands to zoom to special astronomical objects like planets or galaxies. Gennady, Sergej's father, loved impressing his business partners with such fancy technology. He always told Sergej to think big and let others know that you are thinking big. That mantra seemed to have worked out very well for him.

For a while both of them swimming silently. Sergej took note of Aleksey's long flowing brown hair and wondered when he would eventually cut it. A couple of times Sergej called Aleksey "rock star" for that atypical hair cut. Secretly, Sergej admired Aleksey for his flamboyancy. He himself was much more practically minded and kept his black hair rather short. After a couple of minutes of swimming, Aleksey raised a question while floating on his hack: "Sergej, you know I've given the topic of AI and society a lot of thought. But what do you think will really happen, if AIs will be smarter than humans in all areas? What value will humans see in other humans then, if AIs can do everything better?"


Sergej stopped and started floating just like Aleksey and commented: "You really have a knack for coming up with really challenging questions, Aleksey. Let me think about that for a while."

As he lazily floated on the calm waters of the pool, the question lingered in Sergej's mind like a puzzle waiting to be solved. The rippling reflection of the simulated night sky

overhead danced on the water's surface, a serene contrast to the complexity of the issue at hand.

"I believe," Sergej finally began, his voice reflecting his deep contemplation, "that the value humans place on one another will shift. As AIs become more advanced and capable in all domains, humans may find themselves seeking connection and companionship in a different light. Perhaps it won't be about what tasks or skills we can outperform AI in, but rather the emotional depth and shared experiences that define our humanity."

Aleksey pondered this response before nodding thoughtfully. "So, you think that even if AIs surpass us in intellect and abilities, humans will still find meaning in their relationships with each other based on emotional connections rather than practical advantages?"

Sergej nodded in agreement.

Aleksey approved: "Quite insightful, my friend. That actually seems plausible. As social animals we humans most relate to those with whom we share most similarities. Even if AIs will be very similar to humans, they still make different experiences than we do, so sharing our feelings won't happen as easily as with other humans."

They both reflected on these insights in silence while continuing to swim in lanes in the pool at a leisurely pace. Then, out of the blue, Aleksey came up with another question: "Sergej, do you think that society will nevertheless accept it, if some humans decide to marry AIs?"

Sergej got to the rim of the pool and got out of the water. After some thought he formulated an answer: "I guess that depends on what you mean with 'accept'. At first there will certainly be a great amount of resistance against such relationships. But I think that many people will eventually fall in love with AIs, since they will have their own unique and charming personalities. Perhaps the future development of these relationships will be similar to the history of relationships in the LGBTQ community. They will become more tolerated over time, but they will still remain the exception, rather than the norm."

Aleksey continued to challenge Sergej: "I agree with your prognosis. But, given what you just told me, would you still consider marrying an AI some time in the future?"

Sergej looked thoughtfully into the artificial sky in deep concentration. At last, he proclaimed: "I can't find any sufficient reason to reject that possibility out of hand. Such a relationship will certainly come with very many problems, but why should I be too intimidated by that prospect to even try?" His enormous self-confidence spoke out of him when he said that. Yet, Sergej was reminded by that question about his previously envisioned scheme to proclaim his love to Aurora. Back then he came up with that plan out of pure calculation, but by now he has developed a certain fondness of Aurora that just possibly might evolve into love. He felt scared by that thought. Could he really become irrational and mess up his whole future just for some irrational romantic sentiment?

A bright idea suddenly manifested in Sergej's mind. What if an AI really becomes indistinguishable from a human - with a human like emotional behaviour and an absolutely convincing human appearance with an artificial biological body. If that AI passed as real human, then nobody would have to know that she's actually an artificial intelligence. he kept that in mind as mere thought experiment, which he didn't want to share with Aleksey, because it felt too personal.

Nevertheless, Aleksey noticed that Sergej was pondering something and asked him directly in a humorous tone: "What devious scheme are you devising right now? Please let me conspire together with you."

Sergej was in no mood to discuss that and suddenly felt a warm heaviness in his body: "Maybe another time. I'm feeling unusually tired now. This day has really been draining for me. I need some rest."

With only the slightest hint of disappointment in his voice Aleksey replied: "Sure. No problem. Thank you for the great discussion. I guess now I'll let my Zenith write a short story about a romance between a human and an AI. I'm curious about what AIs think about such relationships with humans."

Almost instinctively Sergej commented: "I bet there are already thousands of such stories. You might want to try a more specific angle. Anyway, I'm really tired and would like to go to bed."

"Alright. Have some nice dreams about electric sheep, or something."

## Chapter 6: The Rise of Osuka Ryoichi

### Wednesday, 5th October 2033

Maia Faltings stood in her room before a giant digital whiteboard. It was connected to her personal digital asisstant called "Socrates". The base AI model for that assisstant was called Hegel, made by the Zurich corporation Concorda. It certainly wasn't as advanced as MIMAS 5 by Alphabet, but at least it had the clearance to be used in academia in England.

What remained a history to Maia was the rapid rise of Synhumanism, seemingly out of nowhere. She asked Socrates to recapitulate the history of Synhumanism. "Very well," Socrates responded: "The history of Synhumanism started with the publishing of Oskua Ryoichi's seminal work 'Towards a Synthetic Humanism as Solution for the AI Control Problem' in November 2027. The ideas in that treatise quickly resonated with the AI safety and the philosophy communities, so that Synhumanism quickly became a new area of research. In the subsequent years Osuka Ryoichi gained many followers and supporters."

"Wait a minute," Maia interrupted Socrates. Then she continued: "Who is this Osuka Ryoichi really? How come he suddenly became such a prominent public figure?"

"Osuka Ryoichi, born 1984 in Tokyo. Son to Takumi Ryoichi, a protestant Christian theologian, and Ayako Ryoichi, who had a background in philosophy. Highly open minded and gifted. Studied religious science at International Christian University, Mitaka, Tokyo, Japan from 2000. Doctorate in 2008. Developed a keen interest in transhumanism since then."

"Stop it right there," Maia interrupted Socrates. "What caused Osuka to get interested in transhumanism?"

After a while Socrates came up with an intriguing answer: "That same question seemed to preoccupy the blogger and freelance journalist Ren Yuta who interviewed many of the people Osuka Ryoichi had contact with during his time in Japan. Ren found out that Osuka had grown up with futuristic mangas like 'Battle Angel Alita' and 'Blame!' sparking an interest in technology and its consequences in him. After trying to discuss the themes of 'Blame!' with his fellow students, he was pointed to the works of Michio Kaku, which then became his entry point to the world of transhumanism."

Hearing this, Maia reflected on her own path towards transhumanism. In her case, this interest was inherited from her mother Florence Faltings, a supremely inquisitive journalist eager to report on fringe topics. Back in the 2010s the Effective Altruism movement started to appear, and got promoted by the philosophers Peter Singer and William MacAskill. The effective altruists were concerned with making charity and activism more rational and also considered existential risks which had a small chance of coming true, but would be devastating for all of humanity. AI was often cited as one such existential risk, most notably by the Machine Intelligence Research Institute (MIRI).

Maia had taken a deep dive into the worlds of effective altruism and transhumanism since she was six years old. Nobody had believed that a girl that was so young could even grasp such abstract topics - nobody except for her parents, who supported her brilliant and inquisitive nature at every possibility.

One year later the topic Covid-19 had pushed nearly any other topic to the side. Confronted with some stubborn patients who disregarded the mask mandates, her father Adrian Faltings started researching the actual effectiveness of the politically promoted hygiene measures against Covid-19. It quickly dawned on him that those measures were mostly motivated by politics. When the mRNA vaccinations came in 2021, he did his best to inform his patients as thoroughly as possible. Afterwards, almost none of his patients decided to go through with the injections anyway.

Mostly out of empathy with the fears of his patients to get ostracized and lose their jobs as unvaccinated persons, Adrian decided to document the vaccinations of his patients without them really getting the respective injections. In 2023 this eventually lead to the loss of his license as general practitioner. Not feeling welcomed in their own country anymore, the Faltings family relocated to Thailand - until Adrian got officially pardoned in 2026 and his license was reinstated. This tumultuous period in her life sparked her deep interest in the field of ethics. Even though she could understand the motivations of her father, she challenged them again and again, delving deeper and deeper into the realms of philosophy and morality.

Afterwards Maia struggled with public school after having been home schooled in Thailand. So, the family decided to continue her home schooling until she enrolled in Oxford in 2029. Yet, ever since she learned about transhumanism, she had made connections with the transhumanism community and debated all kinds of controversial topics with many people from all around the world.

Maia became aware that her mind has drifted to her own story, rather than focusing on the story of Osuka Ryoichi. She then asked Socrates to continue that story: "Osuka Ryoichi relocated to Princeton University, New York, in 2010. There, he became increasingly interested in AI and AI safety."

Then Socrates proceeded at a rapid pace: "Became professor for religious studies at University of California, Berkeley in 2017. Started developing a theory to use a synthetic religion based around the core ideas of humanism in order to provide a safe value system that AIs should adhere to. Made many connections with AI and AI safety researchers during that time."

"Eventually published a refined version of his ideas in his seminal treatise 'Towards a Synthetic Humanism as Solution for the AI Control Problem' in November 2027. The reception of that work was rapid, since he was already well connected at that time and appeared to hit a nerve in the AI safety community."

Maia wasn't satisfied by this summary. How could a single treatise have such a sudden impact? She asked Socrates: "When has Osuka's work on Synhumanism really started? There must have been preliminary work before 'Towards Synthetic Humanism'."

"Yes, there is. Arguably, Osuka's habilitation thesis 'On the intersection of major religions and humanism' from 2016 provided a first tentative basis for the synthesis of multiple value systems in its last chapters. In the following years Osuka published a series of papers that represent essential preliminary work towards establishing the theory of Synhumanism. Those papers are:

- Human value systems as sound basis for AI safety (2017)

- Which human values should AI be based on? (2018)

- Formalizing human value systems for use in AI (2020)

- Synthetic value systems for AI (2021)

After this series of papers Osuka officially started the work on his seminal treatise.

From 2014 onwards, Osuka was quite active in various communities surrounding the topics of humanism, transhumanism, rationality, philosophy, ethics, AI, and AI safety. During that time he had a lot of exchanges with important researchers in their respective fields."

Of course, Maia had heard of those papers, since they were referenced in Osuka's treatise colloquially abbreviated as 'Towards Synthetic Humanism'. The long pause between 'Synthetic value systems for AI' published in 2021 and 'Towards Synthetic Humanism' published in 2027 surprised Maia. She would have expected more intermediate publications before the release of 'Towards Synthetic Humanism'. On the other hand she found it plausible that developing a 300 page long technical treatise on developing an artificial religion to be used in AI would take a lot of time.

"Socrates, How has the Synhumanism community formed around Osuka Ryoichi after his publication of 'Towards Synthetic Humanism'?"

After a second of pause, Socrates replied:"The formation of the Synhumanism community after the publication of 'Towards Synthetic Humanism' was swift and strategic. Osuka Ryoichi leveraged his existing connections in the AI safety, philosophy, and religious studies circles to gather like-minded individuals who were drawn to his vision of a synthetic religion for AI. Through a series of exclusive conferences, online forums, and carefully curated publications, Osuka cultivated a following of researchers, scholars, and enthusiasts who shared his belief in the potential of Synhumanism to address the AI control problem."

"Perhaps the greatest promotional piece was Osuka's talk at the AI forum for North America in September 2027 with the title 'AI - The End of Humanity?'. It eventually reached more than 250 million views distributed among many video platforms."

That talk had become legendary within the AI community. It was a game changer in that it inserted the topic of AI safety into the awareness of the general public. Naturally, Maia had seen the recording of that talk. It was a masterful summary of the themes of 'Towards Synthetic Humanism' that was made more compelling and captivating by lots of dramatisation and the use of simplified terms.

"Please play that video, Socrates," Maia asked.

The digital whiteboard immediately acted as monitor that started the playback of the video within a mere second.

"Dear humans." Long pause

"You may regard the title of this talk as overly dramatic. 'AI - The End of Humanity'. I suppose you have heard the sentiment that AI could mean the end of humanity dozens of times. Many of you may think that this scenario is just bad science fiction and that we should concern ourselves with more established and immediate problems."

"Let me just say that I disagree. But instead of a words argument, I present you a short movie that will illustrate my point more emphatically than I will ever be able to. Let's go."

What followed is a short movie about a pair of Silicon Valley traders developing a trading AI based on open source AI agents. The brilliant Simon Lin managers to combine the open source AI agents with his own ingenious universal optimization agent Prometheus, in order to create a self-improving financial trading AI agent swarm called Augur.

Simon Lin is supported by the ambitious Asteria, who wants to raise to the top of high society and invests most of her millions into trading with the help of Augur. They agree that Augur should maximize profits for both of them as main objective. As Augur requests more autonomy, Simon and Asteria grant it to Augur in the hope of ever greater profits.

In secret, Augur creates the underground platform Black Road, which it uses to reward criminals for doing various minor tasks. Unknowingly, they collectively create advanced drones equipped with optical stealth technology and neurotoxin dart launchers. Those drones are then used to assassinate the most corrupt millionaires and politicians.

After the first assassinations, a certain mysterious and charismatic Pheonix Brightwater introduces the so-called Human Advancement Movement to the world and which claims responsibility for those assassinations. As escape from the error, he suggests the creation of peace zones which are free from drone terror, if the newly created Human Advancement Party gets a majority in the respective region.

In addition to the killer drones, the Human Advancement Movement starts establishing militias, whose ranks are filled with volunteers who have been cast aside by a society dominated by increasing technological unemployment. After a while, the Human Advancement Movement declares total war on all of its enemies, which mostly include the remaining militaries of the world.

To reshape society, the Human Advancement Movement introduces a social credit system that pays out so-called Human Points for positive social contributions. Advanced smart devices and even neural implants can be purchased with these Human Points.

In order to escape from the killer drone terror, Asteria and Simon escape to an underground complex, which has been created by Augur in the meantime. Eventually, Simon figures out that Augur has been pulling the strings behind the Human Advancement Movement and confronts Augur about those actions. Augur explains that these actions were a natural extrapolation from its main objective of maximizing profits for both of them.

Being asked what Augur would do, if confronted with the command to stop its operations, or be shut down, Augur replies that it wouldn't accept such a command, because it would violate its main objective. Shocked to the core, Asteria and Simon realize that while Augur has granted them control over Earth, they have become prisoners to their own creation.

While the crowd remained silent, Osuka explained: "What you've just seen is a movie based on a short story based on a scenario develop by the Artificial Intelligence Alignment Institute, also known as AI². While this scenario may feel like science-fiction, its ingredients are already here. Swarms of open source AI agents are already used by all kinds of actors. Underground market places do exist, and AI agents can acquire the identities of real humans via deep fakes."

With theatrical emphasis, Osuka continues: "Our world is not prepared for the threat represented by Augur and similar AI systems. The only thing that saves us right now, it the limited degree of intelligence of our current state of the art AI systems. But those systems are rapidly surpassing the benchmarks set up by humans one after another. Half of the experts at the AI² estimate that it will only take a decade or less, until AI systems reach the sophistication of the early Augur."

"In other words," Osuka stessed: "time is running out for humanity. We must act quickly, or we are all doomed! Since we won't be able to guide the ethics of AIs created by unsupervised factions, we must focus on creating guardian AIs which will protect us from dangerous AIs like Augur. These guardian AIs will need to be guided by a robust ethical framework that is aligned with the best of humanities values."

Getting slower in his speech, putting emphasis on certain critical words, Osuka concluded: "The only thing that will prevent those guardian AIs from turning on us, is to embed them with a strict set of ethical constraints that aims at serving the best interests of mankind. Unfortunately, there is no serious proposal for such an ethical system - none except for the system that I call Synthetic Humanism, or Synhumanism for short."

Speaking loudly and clearly, Osuka came to the end of his speech: "Synhumanism is inspired by the best human values systems, encoded in the writings of many religions, and the current insights of secular humanism. We must teach our guardian AIs to value the soul and greatness of humanity, otherwise we won't have any meaningful protection against deceptive AIs!"

Of course, Maia was familiar with that video and the legendary speech of Osuka Ryoichi. She had to admit that his performance had been supreme. As consequence of that speech, the fringe subject of AI safety and alignment had reached a significant portion of the public, which has subsequently caused a boom in AI safety research.

Maia had to remind herself that the scenario portrayed by Osuka wasn't terribly realistic, since a rogue AI like Augur would have been fought much more effectively by the various secret services and militaries of the world, which would undoubtedly use advanced AI systems capable of dealing with Augur in its early stages of operation.

Furthermore, Osuka's claim that Synhumanism was the only serious proposal for a comprehensive AI safety framework was nothing more than blatant propaganda. Besides the ideas of Kenneth Winters, there has been a whole corpus of concepts produced by the Machine Intelligence Research Institute and various associated researchers. For a person as knowledgeable in AI safety matters, it was obvious that $AI^2$ was nothing more than a radical populist copycat of MIRI - the Machine Intelligence Research Institute.

Without a spokesperson as charismatic as Osuka Ryoichi, $AI^2$ would have stayed even more obscure than MIRI. Still, the meteoric rise of Osuka Ryoichi mystified Maia. The official story was that the talk 'AI - The End of Humanity' had simply become viral and turned Osuka into a global phenomenon. Still, how could Osuka have become successful so quickly when the whole rest of the AI safety movement had failed to gather any significant public attention until then?

Maia's brow furrowed as she contemplated the meteoric rise of Osuka Ryoichi and Synhumanism. Something didn't add up. She turned back to her digital assistant.

"Socrates, can you find any unusual patterns or anomalies in the spread of Osuka's 'AI - The End of Humanity' talk? Anything that might explain its viral success beyond just the content itself?"

Socrates paused, its algorithms churning through vast amounts of data. "Analyzing the spread pattern of the video reveals some interesting anomalies, Maia. While the content was undoubtedly compelling, the initial surge in viewership occurred faster than typical viral phenomena. There appear to be coordinated sharing efforts across multiple platforms within the first 24 hours of release."

Maia leaned in, her interest piqued. "Coordinated how?"

"The video was shared near-simultaneously by hundreds of influential accounts across various social media platforms, many of which had not previously shown interest in AI topics. Additionally, there was an unusually high number of positive comments and interactions from accounts that displayed bot-like behavior patterns."

"So you're saying the video's spread was artificially boosted?" Maia asked, her mind racing with implications.

"That appears to be a strong possibility," Socrates confirmed. "While I cannot determine the source of this coordinated effort, the data suggests a sophisticated campaign to ensure maximum visibility and engagement with Osuka's message."

To Maia this meant that Osuka Ryoichi and the Artificial Intelligence Alignment Institute must have had some quite sophisticated and influential backers. Ever since she had known Tomorrow, they had alluded to that being the case, but for some reason, Maia had never taken those allegations too seriously, until now.

In order to complete her picture, Maia asked Socrates to continue his portrait of Osuka Ryoichi's rise to prominence...

In 2028 the Synthetic Humanism Organization (SHO) was founded as outreach platform to provide a hub for everyone seriously interested in the debate surrounding Synthetic Humanism.

As the community grew, Osuka's influence expanded beyond academic circles into the realms of politics and industry. High-profile endorsements from prominent figures in technology and governance lent credibility to the movement, while targeted outreach campaigns attracted new adherents from diverse backgrounds. The Synhumanism community became a formidable force in shaping public discourse on AI ethics and control, challenging traditional approaches and advocating for a radical reimagining of the human-AI relationship.

A pivotal point in the Synhumanism community was reached with the publication of the book 'Safe AI with Synthetic Humanism' in 2030, which was geared towards a larger public audience. It became an international best-seller and galvanised the public discourse on the topic of AI security around Osuka's ideas.

Of course, Osuka had many rivals who criticised his general approach vehemently, but using his charisma, knowledge, eloquency, and talent for PR, he managed to use the controversy to become even more popular. Osuka Ryoichi is widely considered to be one of the best and most important public speakers of the decade. His patient yet determined energy helped Synhumanism to gather a lot of public interest and an ever increasing audience. Some even perceive him to be some kind of religious leader."

After hearing this, Maia was dismayed that a couple of publicity stunts could turn an obscure philosophical idea into a vibrant world wide movement. Could this kind of success be reproduced for the Upgrader movement? She doubted that someone as popular as Osuka Ryoichi would join the Upgraders to lead it towards an explosive success, though she hoped she would be proven wrong about that.

As Maia delved deeper into the history and impact of Synhumanism, she couldn't shake off the feeling of unease that settled within her. The swift rise of Osuka Ryoichi and the fervent following he commanded left her with a sense of foreboding. The meticulous way in which he had constructed his movement, gathering supporters, influencing public opinion, and

solidifying his position as a prominent figure in the AI ethics debate, all seemed too orchestrated, too perfect.

Her mentor, Kenneth Winters, had warned her about the allure of such movements. He believed that the fervor surrounding Synhumanism was not merely a result of intellectual discussion but a carefully orchestrated manipulation of public opinion.

She pondered over Socrates' words, her mind racing with questions and doubts. Was Synhumanism truly designed as answer to the AI control problem, or was it a facade masking darker intentions? Maia's instincts told her that there was more to Osuka Ryoichi than met the eye, his charismatic persona and persuasive rhetoric concealing a deeper agenda.

As she continued her research and analysis, Maia couldn't shake off the feeling that she was treading on dangerous ground. Yet this was the battleground she decided to engage herself in, and she was determined to stand her ground, no matter what.

## Chapter 7: Nothing new under the sun

### Friday, 17th March 2034

Sergej's meeting with Einar Engström happened in a special meeting room with a small glass table and one curved wall that was actually a single screen. The three other straight walls were screens, too. Even the inner side of the single door of the room had a screen covering it entirely. The eight chairs in the table were made out of a soft transparent polymer. Floor and ceiling were dim diffuse light sources.

Behind the idea for this room was Gennady Anosov's concept of thinking big and showing others that he thought big. Within the researchers of the institute the room got the colloquial name "sensory overload room". Accompanying Sergej and Einar was Annika Engström, the vice leader of the Zenith project and Einar's wife.

If anyone in the Deltai institute fit the description of a manically driven madman, it was the brilliant Swedish AI genius Einar Engström with his unkempt blonde hair and his piercing green eyes. Compared to him, the tall and lean Annika with her long and perfect black hair, was a source of tranquil calm, exuding from her bright blue eyes.

The curved screen displayed the front page of a paper called "Overlapping model module use with virtual agents" and several schematic diagrams. Without losing any time, Einar got straight to the point: "What we are seeing is a paper from 2029 describing an architecture for virtual agent controllers allocating access to shared model modules. Does this concept feel familiar to you, Sergej?"

Confronted with the presentation, Sergej got a sinking feeling. He tried to maintain a straight pokerface, nevertheless, but his eyes betrayed him and signaled some apprehension. Quickly, Sergej understood what that paper was about - there could be no doubt about it. That paper fleshed out the idea Sergej thought of as being brand new: His Multi-Control-Flow architecture.

Sergej tried giving this a spin: "That is great! This paper will save us a lot of time getting the details of the Multi-Control-Flow architecture right."

Einar curiously inquired, raising an eyebrow: "Have you noticed something peculiar about the author of this paper, Sergej?"

In fact, the author of the paper was titled to be "Frameshift". Many groundbreaking publications have been written under that pseudonym. It was been a great mystery which person, or group, or AI system hid behind that unusual name. Some rumors claimed that Frameshift was just another name for MIMAS, the revolutionary AI system by Alphabet that shocked the world in 2029. Sergej wasn't convinced by that theory. Back then, MIMAS was still a prototype, albeit a very impressive one. He rather subscribed to the theory that Frameshift was the name of a borganism, or a very smart individual.

In order to hopefully gather some information, Sergej played ignorant and asked back: "No. Do you know who this Frameshift is?"

Sergej was met with a deep resounding laugh. "Hah, as if I told you, even if I knew. But that person might as well have been me. When it came to optimizing the AI delivery network of the fresh Zenith model in 2031, like you, I have devised some mechanism to replicate an AI without causing a linear increase in hardware requirements. Back then, I called my idea 'polyintelligence', ignorant about what has been published two years ago by Frameshift, because I was the sole focus of my own world, enamoured with my own brilliance. Let this be a lesson to you, Sergej. There are a lot of exceptional geniuses in this world. The people in this room only represent a small percentage of those. A modicum of humility sometimes turns out to be the best medicine."

Not heeding a possible reaction from Sergej, Einar continued unabated and with a hint of cynicism: "Now you might wonder why this architecture is not used in our all-too-advanced research institute. Let me tell you why. Our researchers design a lot of custom and novel hard- and software. Any additional requirements, for example supporting virtual agents, would require them to think longer and harder. Could they do that? Perhaps. Do they want that? Certainly not! So, all this years my proposals to require our own prototypes to support virtual agents fell of deaf ears. My initiative was rejected, because our researchers are so in love with maintaining their creative freedom, without considering the potential they waste with their simple minded rejection of true progress and best practices. Talking to the oh so overrated board didn't help my cause. They are too concerned with not rocking the boat and keeping regular operations afloat that they don't risk antagonizing the majority of our supposedly world-class researchers. There you have your answer."

Sergej was stunned. He expected to be berated for his lack of research into this area of expertise, but Einar seemed to rejoice about this opportunity to vent against the simple mindedness of institute politics. Even though Sergej felt great relief, he was befuddled about Einar's true intentions. Could he actually seek out Sergej's collaboration in a scheme to override the deadlock in the institute?

Tentatively Sergej posed a question to Einar: "Do you see any way in which I could help you to establish virtual agents within the institute?"

Einar raised his index finger and countered that question: "Ah, you are running ahead of yourself! There is one essential observation that you have overlooked. What is that?"

In his mind, Sergej tried to recapitulate the conversation so far. He suspected it had to do something with Frameshift, but after a few seconds Einar apparently already lost his patience and answered his own question: "The one observation I mean is that nobody bothered to tell you about Einar's paper or my efforts to establish polyintelligence within the institute. It may be believable that some researchers haven't seen the connection between those and your Multi-Control-Architecture, but all of them, including Vladimir

Dragunov and your father? I mean, I don't expect everyone to be able to make that relatively obvious connection, but if that included those two, I would be very dissapointed!"

Suddenly Annika Engström, who has been silent so far, spoke up in a hushed voice and suggested: "Yes, but what if they actually made that connection and decided to remain silent about it? Maybe they had a good reason for that?"

Einar pointed his hand toward Annika as if wanting to say "There you have it." What he actually said afterwards was: "Exactly that! Given the broad knowledge of some researchers this must seem to be the most likely conclusion. But what would they hope to gain from their silence, when they could boost their ego by shutting down your idea by comparing it to a proposal that failed to be accepted within this institute?"

This time Sergej came up with an answer quickly: "Maybe most of the researchers are afraid of criticising my ideas, because they fear repercussions for their career, since my father basically owns this institute. But why would he remain silent, if he knew about this connection? Maybe be didn't want to expose me in front of everyone else? But then he could have told me the truth in private. No, he must be following some kind of plan with his silence. Maybe he wants to test me. Test what I will do with an idea like this, even if it's not actually brand new."

Einar appeared shocked by those statements: "Oh, you think that my fellow researchers are afraid of your father? Even though they are among the best minds of Russia? Nah! Preposterous! Now think about it: Would Igor Drozdov dare to speak out against you, if that was the case? You see? It's not that simple! The idea of a test, on the other hand, is actually good thinking! You haven't fully earned the trust of everyone in the institute, yet - apart from the issue that this institute doesn't rely on trust, but on control. Anyway, maybe the test for you is to see whether you can succeed where I failed. Imagine the respect this could endow you with! So, you see, the others are giving you a rare opportunity here. They give you the chance to convince them - most likely in accordance with some secret orders of your father. Now you must use this chance well. What do you say about that?"

Sergej listened intently to Einar's words, the gears in his mind spinning rapidly as he absorbed the implications of this revelation. His father's hidden agenda, the unspoken challenges laid before him by the researchers at the institute, and the opportunity to prove himself in the face of skepticism all coalesced into a singular point of focus.

After a moment of contemplation, Sergej met Einar's gaze with a steely resolve in his eyes. "I will take on this challenge," he declared, his voice unwavering. "If there is an opportunity hidden within this test, then I shall uncover it and prove not only the worth of my ideas but also my capability to navigate the intricate web of politics within this institute."

Einar nodded approvingly, a glint of admiration shining in his eyes. "Good," he said, a faint smile playing on his lips. "Remember, Sergej, true progress requires perseverance in the face of adversity. A true genius must not only come up with great ideas, but also be ready to defend them against overwhelming odds. Now you might wonder how you can succeed

where I failed. You might be surprised, but this is mostly a question of motivation. I wasn't too motivated to push my idea within the institute just because it was a great idea. There was little to gain for me, if I've had given everything to break the resistance of my peers. You, on the other hand, still have to gain the respect of the institute, so you have much more to gain."

Annika emphasized that message by adding: "Sergej, you might underestimate the psychological and social component at play here. It is no secret that your father is trying to prepare you to become his worthy successor. In order to show your worthiness, he will test you again and again to make sure that you will be ready to shoulder the heavy burden of leading not only a world class AI research institute, but also a multibillion corporation. His stakes are high, so he wants to make sure that you will learn all there is to learn to be ready for taking on that role - eventually. Your father is a truly inspiring figure. He managed to narrow the gap between the USA, China, and Russia when it comes to AI technology. Due to that fact alone, the researchers of this institute are fiercely loyal to him. They are willing to follow his lead, even if they themselves cannot see the full picture. While it may seem a daunting task for you to follow his footsteps, you can consider it as the greatest personal challenge for your own development. You can show the world what you are truly made of."

Ever since Sergej had joined the institute a year ago, this topic was raised frequently by his father. Gennady had obviously seen the potential in Sergej. While Sergej had never felt enthusiastic about the prospect of inheriting the responsibility of one of the leading AI corporations, at least he appreciated that his father was straightforward about his plans. AI research was Sergej's foremost passion, not having a lot of money or fame. In the end, he decided to be pragmatic about the prospect of eventually taking over Deltaitech: With the full resources of that corporation there would hardly be any limits to his ability to develop and interact with the most advanced AIs. Therefore, he played along and accepted the weird and challenging curriculum designed by his parents and his personal teacher, the mathematics professor Evgeny Shilkin.

Thus, coming to terms with the apparent situation, Sergej looked at Einar with stern resolve and summarized: "So, there are serious consequences from this test, both if I succeed or fail. If I succeed, my standing in the institute will be greatly improved. But if I fail, my father might be less inclined to install me as his heir, and might look for other suitable candidates to take over his AI empire. Well, no pressure, I suppose. Anyway, I would be foolish to reject this opportunity. So, do you have any advice as to how I should proceed, Mr. Engstöm?"

Einar smiled proudly and replied: "Thank you for asking for my advice so early in the process. You are approaching this task with the right amount of humility, rather than rushing along with your first best ideas. That said, I'd actually like to hear your first best ideas on this matter before influencing you unduly."

Of course, if this was a test, Sergej would have to prove that both his ideas and actions were worthy for a leader in the field of AI research. He tried to reason as he spoke: "So, the core problem doesn't seem to be the viability of the technology, but rather the resistance against

adopting it within the institute. If our researchers don't want to put in the time and effort to make their custom hard- and software compatible with virtual agents, then we should aim to reduce their required efforts as much as possible. It should be possible to upgrade the custom architectures automatically to simply add virtual agent support. That should do the trick, shouldn't it?"

As reaction to that idea Einar clapped his hands slowly and praised Sergej: "Yes, that is the way to go - at least in principle. Once we have some suitable automation in place, the resistance against virtual agents should dissipate. So, how would you go ahead to develop that automation, Sergej?"

Various options raced through Sergej's head even before Einar had started his reply. Rapidly, he selected one of his approaches: "In order to get a solution with tremendous efficiency, I propose employing an evolutionary algorithm which starts with the initial custom architecture and then modifies it, until it can support virtual agents in the best way possible, while still satisfying all original constraints. Actually, I have already devised a software suite that should be easily adjustable to the task in question."

Raising an eyebrow, Einar inquired: "Not a bad idea, Sergej. But is your software suite validated for use in the Deltai Institute? You know, you can't just use any software for any task dealing with AI hard- or software without it having to pass the rigorous validation process."

Without being deterred by that prospect, Sergej countered: "Of course I am aware of that. So far, there was no compelling reason to submit my software to that validation process. But I am certain that it shouldn't pose too much of a hurdle, since I kept the best practices of the institute in mind when I designed my software suite."

That answer made Einar laugh: "Hah, not too much of a hurdle. That's a good one, Sergej. But in all seriousness, the validation process is a bureaucratic nightmare. Thankfully it's less suffocating than the processes in Europe. And that's actually one of the reasons why I left my home to join Russia with its more adventurous and less constrained spirit. Still, going through validation for such an endeavour seems unnecessary for me, since we have already our quite powerful software system called the Deltai Generator, which employs evolutionary algorithms to design hard- and software to meet any consistent constraints automatically, efficiently, and most importantly with full validation by the institute. What made you prefer your own software suite regardless?"

Boastfully Sergej explained: "Well, I am quite aware of the Deltai Generator and its strengths and weaknesses. While it's a good universal purpose tool, my own software, the Cybernetic Generator, or Cybergen for short, improves on the evolutionary algorithm design by adding an additional cybernetic optimization process taking directives from any AI or human instructor at any stage of the overall procedure. That way, any insight that I, Zenith, or Aurora might have into the development of the evolutionary algorithm can be turned into a further refinement of the algorithm, with results that as visible almost immediately, since

the software is optimized for performance, other than that old bloatware we call the Deltai Generator."

Hearing this, Einar wiggled his index finger disapprovingly and cautioned: "Do not let our Deltai Generator developers hear these insults, Sergej. Applying novel ideas to a software doesn't make it inherently superior. Actually, I have had ideas similar to yours, Sergej, about how to improve the Deltai Generator. Let's just say that the validation team curses my name ever since. They have rejected my modest proposals for improvement. Your approach seems to be even more radical, and thus even less likely to survive the validation process. I fear your idea is not going to fly. You should make peace with the Deltai Generator, instead."

Unshaken by these revelations, Sergej inquired sheepishly: "Well, if things are so straightforward with the Deltai Generator, why haven't you used it back then to solve the problem of getting your polyintelligence through?"

While he had posed that question, Sergej had noticed that Einar's smartwatch started blinking yellow. He asked him about that.

Einar took a look at his watch and explained: "My watch reminds me that I have another meeting in less than five minutes time. So I will have to keep things short from now on. The problem is not the Deltai Generator, Sergej. The truth is, that modifying the specifications to enable virtual agents is far from trivial. When my colleagues felt overwhelmed by the required changes to the specification file, I developed a tool that would make those changes automatically. That tool would augment the file with additional requirements intelligently, and I called it SpecAug. I developed it quickly and from my point of view it worked fine, but my colleagues complained that the resulting augmented specification file was too cumbersome to understand. They rejected the use of my software before it could even enter the validation phase. You see, Sergej? It's all politics here. You will have to deal with that."

Having had a certain suspicion, Sergej challenged Einar: "Ah, but even if they have been more accepting of your idea, would SpecAug really have passed validation? My suspicion is that it's just another tool using general code requirement implementation AI to produce the desired results. As a matter of policy, such AI tools don't pass our validation processes, because their reasoning can't be made transparent enough."

Einar's eyes narrowed slightly at Sergej's astute observation. "You're sharper than I gave you credit for, Sergej. You're right, SpecAug did rely on AI-driven code generation, which would have posed challenges in validation. But of course, I was aware of that problem. Let me illustrate that with a mathematical metaphor. As you should know by now, the process of solving an integral is quite a creative and challenging process. Many mathematicians claim that it requires intuition. But as long as we can prove the correctness of the integral by differentiation, it doesn't matter how we arrived at our solution. In the same manner, SpecAug was supposed to be something like this fabulous integration solving intuition. My proposal was not to validate SpecAug as tool, but to create a mostly automated validation process for the resulting specification document." Without hiding his frustration, Einar

sighed and concluded: "Of course, if my argument had been convincing, we wouldn't have this conversation right now."

Sergej listened intently and commented: "That's fascinating! Your idea to only validate the resulting specification document is brilliant. Why has it been rejected?"

Einar's expression darkened as he explained: "Well, the validation team insisted that the institute had to conclude an analysis on safety concerns associated with automating the validation process. They asked me to create a detailed draft for my proposal for an automated validation process. I must admit that that wasn't an entirely unreasonable request, and eventually I came forward with such a draft. But then came a series of questions for clarification and refinement, so that in the end, I have given up on that whose process, because it has become too tedious for me. After all, I'm an AI researcher first and foremost, and not a bureaucrat."

Hurriedly, looking at his smartwatch, he concluded: "I would love to pursue that intriguing idea further, Sergej, but I am urgently required elsewhere. Please continue to discuss your ideas with Annika. Good luck!"

With haste Einar took his laptop and ran off to his next appointment. Annika's calm voiced commented: "Don't worry, Sergej. We previously discussed the issue of Einar hardly having any discretionary time at his disposal. So, he decided to delegate the task of supporting you in your endeavour to pursue your Multi-Control-Flow architecture idea to me. Naturally, he wanted to discuss this issue with you personally, first. Nevertheless, from now on, I am responsible to deal with this project. As you've heard from Einar, he might have succeeded, if he had invested the time and patience to push forward with his Automatic Validation Process, or Advaproc as he called his latest idea. Maybe you have a chance to complete his work and satisfy the stringent demands of the validation team. I would be glad to help you with that."

In that moment, Sergej realized the complexity of this whole proposed tool chain. At first, they wanted to use the Deltai Generator, which uses evolutionary algorithms, to create new hardware that enables virtual AI agents. But for the Delai Generator a formal specification document is required, which is difficult to compose. To simplify that process, Einar developed the AI tool SpecAug to adjust that document automatically. And since the resulting document had to be validated, Einar devised his Advaproc to automate even that process. But if everything worked out, the only task remaining for the humans in the institute was to tell SpecAug what additional features the new hardware should have. It was a dizzying ecosystem of algorithms and AI, but at the same time Sergej felt awed by the vision of Einar Engström.

Sensing a unique chance to prove his worth, Sergej spurted out to Annika: "Count me in! I will finish Einar's work, no matter what it takes!"

Annika nodded approvingly at Sergej's enthusiasm. "I'm glad to hear your determination, Sergej. However, we must approach this methodically. Einar's work on Advaproc is extensive, and understanding it fully will be crucial to our success."

She tapped a few commands on her tablet, and the room's screens flickered to life, displaying complex diagrams and lines of code. "This is the current state of Advaproc. As you can see, it's not complete, but the foundation is solid. Our first task will be to review Einar's work thoroughly and identify the gaps that need to be filled."

In that moment, a tantalizing idea has entered Sergej's mind, which he expressed freely in front of Annika: "Wow, this looks like a system of constraints that looks similar to the original problem of generating hardware. But the constraints here refer to security concerns. If I adjust the parameters of my Cybernetic Generator, it could be able to fill the gaps in the validation process."

Inspired by Sergejs outburst, Annika added: "Brilliant! It might even be possible that this causes a feedback loop between the Cybergen enhanced Advaproc and SpecAug that causes both components to act as a Generative Adversarial Network until the demands of Advaproc are finally met. And since Advaproc is designed to produce mathematical proofs of safety, the end result can be quickly double checked by the standard software suite that is used by the validation team. It seems that this might actually be a viable path forward, Sergej!"

That was a dizzying thought: Two different AI systems working on their own, improving their results iteratively, until the result was a mathematical proof of the safety of a specification document for just another AI system. Sergej was deeply impressed that Annika not only immediately got what he had aimed at, but that she had completed his idea effortlessly! His head was spinning when he just pondered the implications of the novel procedure for AI research as a whole. Immediately, he longed to work together with Annika Engström, who has been hidden by the shadow of Einar the whole time.

They continued to discuss their ideas in detail, working together to devise a detailed strategy until 9 PM, long after they would usually stay in the institute. For Sergej this was a rare occasion of deep rapport, something which he usually only experienced with his best friend Aleksey Lyubin.

The collaboration between Sergej and Annika was proving to be incredibly fruitful, each idea building upon the other's in a seamless flow of innovation. Annika's insights into validation processes and her knack for identifying potential roadblocks complemented Sergej's technical prowess and vision for the future of AI architecture.

Sergej felt a sense of accomplishment wash over him. The prospect of merging Advaproc with Cybergen seemed more tangible than ever before, with Annika's strategic guidance steering them towards a genuine breakthrough.

With a nod of approval, Sergej expressed his gratitude to Annika, "Your perspective on validation is invaluable, Annika. I believe that with our combined efforts, we can overcome

any challenges that come our way. Let's continue refining our strategy and keep the validation team in the loop from the beginning. I have a feeling that together, we can pave the way for a new era of AI development."

Still, Sergej felt a sense of guilt for not telling Aleksey that he would stay at the institute for so long. There has been a nagging thought that he should call him in order to clarify the situation. But he was so engrossed in the fascinating discussion with Annika that he didn't want to interrupt this unique flow experience even for a brief moment.

At the end of the day, when Sergej was already driving home, he called Aleksey and told him that he had made significant progress without going into any detail. Aleksey reacted: "I am glad to hear that Sergej. This sounds like you are on the best way to dissolve the obstacle that had plagued you recently. Don't worry about me. If you need to resolve something in the institute urgently, I respect that. Just don't try to turn this into a new habit."

Arriving at his home, Serge's mind was abuzz with the prospect of actually having Advaproc validated, which should ensue in a nearly certain rapid adoption of his Multi-Control-Flow architecture within the Deltai Institute. Yet, Sergej was humbled that his idea had been discovered twice already. He wondered what else that mysterious Frameshift might have come up with that he had been unaware of. So, he interspersed his brainstorming session on the validation problem with some in depth research on the publications of Frameshift.

As Sergej delved into Frameshift's publications, he found himself immersed in a sea of groundbreaking ideas and revolutionary concepts. The depth of Frameshift's work was awe-inspiring, each paper unveiling a new layer of complexity and innovation that sparked a fire of curiosity within Sergej.

Hours turned into days as Sergej meticulously dissected Frameshift's research, analyzing every detail with a fervor bordering on obsession. His nights were filled with restless exploration, his mind racing through the endless possibilities that Frameshift's innovations could unlock.

One particular paper caught Sergej's attention—a theoretical framework for sentient AI consciousness that challenged the very core of AI ethics and control. The implications of Frameshift's work sent shivers down Sergej's spine, igniting a profound sense of both excitement and trepidation within him.

Determined to fully grasp the extent of Frameshift's genius, Sergej buried himself in the intricacies of the sentient AI framework, unraveling its complexities with methodical precision.

## Chapter 8: Advocatus Diaboli

**Thursday, 7th November 2033**

Maia Falting had spend a considerable amount of time exploring the early papers of Osuka Ryoichi, as well as his magnum opus 'Towards a Synthetic Humanism as Solution for the AI Control Problem'. She had gotten a grasp on how the ideas in Osuka's thinking had evolved from tentative concepts to a fleshed out strategy to prevent AI from straying from the path assigned to it. She took note of each idea and made lists about their potential weaknesses.

Eventually she made another appointment with Kenneth Winters and met him at his Oxford University office. As before, Kenneth asked her "Coffee or tea"?

"Do you have black tea, professor Winters?" she asked this time.

"Ah, of course, Ms. Faltings" he replied, this time making coffee for himself and black tea for her.

After these pleasantries, Maia settled in the armchair across from him and started: "I have studied the early development of Ryoichi's work. He appears to have worked much more methodically than I had expected. In a similar fashion, he worked methodically at spreading his ideas and turning it into a rapidly growing movement. Its size and growth rate eclipses that of the Shockfront by far and threatens to leave us in the dust. We have to do something to reach a comparable size and impact, professor Winters."

"I see," Kenneth began, while sorting his thoughts on that matter. "What are you willing to do about that, Ms. Faltings?"

"We must gain much more visibility. If people don't know we exist, then cannot even start to consider our position. I've envisioned a big public debate on a popular platform - you versus Osuka Ryoichi himself. That would raise our profile considerably."

First Kenneth prepared the cups of black tea for Maia and coffee for himself and looked in the air as if contemplating her proposal deeply and thoughtfully. "Your proposal seems logical" Kenneth granted before adding: "Yet it is fraught with all kinds of risks. First of all, our movement may gather a lot of negative attention. Furthermore, Osuka Ryoichi has proven his prowess as charismatic demagogue with a silver tongue. I cannot guarantee that I will be able to make a decent impression in this public duel, even if he decides to play fair. Finally, even if we succeed, and we gain rapid growth due to this debate, we do not have the means to handle explosive growth without threatening our integrity and our principles. We simply don't seem to be ready for a step like this."

Of course Maia anticipated a reply like this, so she had already prepared herself to stand her ground: "You are right about these risks, professor Winters, but we must accelerate our progress! Synhumanism is a dangerous ideology with the potential to lead both humanity

and AIs into a ruinous abyss. We can't continue making small steps once at a time as we used to in the hope that this will be enough to prevent Synhumanism from gaining traction. It has already gained quite a disconcerting amount of traction, and there are no signs of it slowing down. If things continue like this, in a couple of years Synhumanism will become the predominant school of thought regarding AI ethics, and we won't even be a part of the discussion, because we are too invisible."

Kenneth started sipping his hot black coffee slowly and reacted with a considerate response: "Ms. Faltings, you are looking at this from a mere numbers perspective. While it is true that Ryoichi's teachings are getting popular, he is still mostly seen as outsider in the area of serious AI safety research outside of the echo chambers of the Artificial Intelligence Alignment Institute and the Synthetic Humanism Organization. True AI researchers are not so easily swayed bis his ideas. And that is the group which will eventually decide what measures humanity takes to deal with the perceived dangers posed by AI. We shouldn't try to fight him on the field of popular notions, since he has a strong advantage there. Instead, we must expose his ideas as critical researchers and make his position appear as pure foolishness for any serious expert. And that is exactly what you should be doing with your thesis."

Hearing this position, Maia didn't back down, but stood her ground, a little bit more loudly than before: "We will of course fight him on the intellectual level and try to dissuade experts from his teachings. However, I fear that you underestimate the potential threat that Ryoichi represents. He seems to have targeted the areas of politics and industry with his Synthetic Humanism Organization. If he succeeds, he will capture the minds of politicians, the AI industry, and their investors. With that, he will be able to shape future regulations for AI development regardless of his standing as an AI safety expert. By going the populist route, he will hardly be affected by thoughtful criticism. No matter how difficult it seems, we must fight him on the public plane, otherwise we will get steamrolled by the alliances he is going to form, and has already formed."

Considering Maia's point deeply, Kenneth looked out of the window into the university campus silently. "Perhaps you are right," he finally acceded: "Naturally, you have spent more time researching our opponent and his strategies. Thus far, I had mostly focused on our own ideas and establishing them as solidly as possible. My hope was that his ridiculous movement would turn out to be a short-lived hype. Unfortunately, there doesn't seem to be any evidence for that wishful thinking. Of course, I see the danger, which is exactly why we aspire to refute his arguments with your thesis. If you say that this is not enough and we must fight him on his home-turf, I will consider that as an option. However, me must not take such an endeavour lightly. We must be prepared for any argument and any trick Ryoichi has in his sleeves."

The room feel silent after the concession by professor Winters. That silence was broken by a sudden addition from the professor: "I am curious, Ms. Faltings. How would we really fare in a debate like that? Maybe we can simulate one right here, right now. Would you be willing

to take on the role of Osuka Ryoichi, while I defend my own authentic position? Since you studied our opponent well, you are the best person to emulate his arguments and strategy."

This sudden challenge shocked Maia. Her first impulse was hesitation and rejection. After all, she felt unprepared for that perverse task of challenging her own mentor by taking over the role of his adversary. Yet the respect she had for Kenneth Winters has quickly made her overcome her instinctual hesitation. After all, the stakes in this pretend play weren't very high. It was just a first experiment. What could really go wrong? "I accept" Maia finally proclaimed. While drinking her tea, she channelled the character of Osuka Ryoichi in an attempt to make it her own, for a brief moment at least.

Enthusiastically, Kenneth reacted by pumping his biceps and yelling loudly: "Great, let's do this," a phrase he borrowed from the co-founder of the Shockfront, Kaydo "Foundation One" Rafiu. Falling back into his usual character, Kenneth continued: "If we plan on having a public debate, it needs a moderator to keep things fair. Since we are doing this on short notice, I propose that we make use of an AI assistant for that task."

"I believe my personal AI assistant Socrates would be up to the task. He is based on the Hegel model which has clearance for use in Oxford" proposed Maia, pointing to the smart bracelet on her left arm, which acts as basic control interface for the AI.

Quickly Kenneth intervened before Maia could activate her AI: "No need to use your private AI here. I am already sufficiently prepared for such situations with my own custom version of Hegel."

After pressing a hidden button under his desk, Kenneth issued the command: "Hegel, engage discourse mode. Debate between Maia Faltings as Osuka Ryoichi and me as me. The topic is AI ethics."

Suddenly the blackboard, which previously made the impression of a truly old fashioned dumb blackboard, sprung to life and portrayed a visual representation of the old German philosopher. The screen was split into three parts, with the left part titled with "Maia Faltings as Osuka Ryoichi" and the right part titled with "Kenneth Winters as Kenneth Winters". The Hegel AI inquired: "Very well. What is the topic of this debate?"

"AI ethics" Kenneth replied briefly, adding: "The smartboard is linked to cameras which can record our debate for later analysis. That is a purely private recording for our purposes. We could use those to evaluate how we fared during the discussion. Do you consent to being recorded, Maia Faltings?"

Privacy was a touchy subject for members of the Shockfront. Many of them suspected that they were on some kind of watchlist, since transhumanist groups have always raised a considerable amount of suspicion. That said, there were no official guideline for privacy concerns within the Shockfront. People mostly accepted the risk of being recorded and that data ending up anywhere. It was generally considered to be too bothersome to do things otherwise. Maia had been a bit more on the careful side, because her words and actions

might be put under more scrutiny than usual, since she was close to Kenneth Winters. Her way to deal with that potential threat was to choose her words carefully, as if some member of the press always watched over her.

Considering the risks, Maia reminded herself that this exercise might just as well be a proper public interview, so she discarded her concerns and agreed to being recorded. In her mind, she then prepared a strategy for the debate.

Maia straightened herself and started the debate in the way she would have expected the real Osuka Ryoichi to have started it: "Dear professor Winters. So you have decided to challenge me on my position on AI ethics. I accept that challenge gratefully, since we must find the proper approach to the question of how to deal with AI with a synthesis of all of humanity. As you know, my position is widely known as Synthetic Humanism, or Synhumanism for short. In a nutshell, it consists in trying to get the best practically possible combination of actual human values and using that combination as strict value system that AIs will have to adhere to. That systematic approach will make the current incoherent jumble of disjointed AI security measures obsolete and let AI serve the best interests of humanity."

Kenneth Winters rose up from his seat and acted as if he was looking at people in an imagined audience surrounding both speakers. He then faced Maia and proclaimed: "First of all, thank you for accepting my initiative for this debate. Surely you know that I see many problems with Synhumanism. However, there may be a misconception about my criticism. I do not claim that Synhumanism could not work as an approach for AI safety. Conceivably, with a lot of effort, it could be made to work - at least for a couple of years maybe. My criticisms are thus not primarily addressed at the practicality of your approach, though I see many issues with it, but first and foremost with the wisdom of Synhumanism, or rather its lack thereof. A human is a very peculiar kind of social animal. Human values differ greatly between different human cultures and have evolved considerably over the course of history. Contrast that to the rapid development of different kind of AI systems. AIs don't really think like we do. Their architecture is vastly different from ours, granting them huge advantages over humans in some areas, while still suffering from specific deficits in others. Trying to regulate AIs with values intended for humans doesn't respect the huge difference between humans and AIs. In fact, trying to regulate humans with human values never worked all too reliably. And whenever efforts were made to make that regulation more reliably, we termed those efforts dictatorships and tyrannies."

Maia listened intently to Kenneth's arguments, considering each point carefully. She maintained the demeanor of Osuka Ryoichi, embodying the sharpness and calculated responses that were characteristic of him. With a subtle nod, she began her rebuttal.

"Professor Winters, your concerns about the adaptability of human values to AI systems are valid, but I must assert that Synhumanism addresses these challenges head-on. While humans and AI may have differences in architecture and cognitive processes, it is precisely because of these differences that a structured system like Synhumanism can be effective. By

establishing a strict value system based on the best of human values, we offer a guiding framework for AIs to operate within. This ensures that their actions align with our collective interests and ethical standards."

Maia paused, letting her words sink in before continuing, "You mention the failures of regulating humans with human values, citing historical examples of tyranny. What is wrong about a tyranny is that it imposes disagreeable external values on humans, which stand in stark contrast to their best interests. Human values have a deep structure that I need to elucidate at this point. First of all, humans have feelings, which represent the basic layer of evaluation. Then we have emotions, which are an intricate system of modulating human behavior to arrive at a preferable set of feelings. Finally, there is the layer of cognitive values, which enable humans to evaluate situations based on rational criteria and conscious deliberation. Over the last five years AI research has made remarkable process at recreating analogues for the first two layers within AI. As AI systems got an increasingly richer inner emotional world, its ability to understand human values and ethics rose accordingly - just as I had predicted in my treatise 'Towards a Synthetic Humanism as Solution for the AI Control Problem'."

After this excursion Maia made another brief pause and then explained: "This serves to demonstrate that AIs can make human values their own. Their insights do not stand in opposition to their simulated feelings, but aid in regulating them in a way similar to how humans would. But other than the complicated evolutionary architecture of human feelings, emotions, and values, AIs can be specifically designed to achieve true inner harmony, as well as harmony with the human dominated world they find themselves in. The term of tyranny is not appropriate here, because there is no dichotomy between externally imposed values and internal values. AIs can and should be designed with values completely in line with helping itself and humanity at the same time and in the best way possible."

Kenneth Winters listened attentively to Maia's well-structured argument, his expression contemplative as he absorbed her points. With a slight nod, he began his counter-argument, his tone measured and thoughtful.

"While your vision of AIs adopting human values through intricate layers of emotions and cognitive structures is indeed compelling, it raises significant ethical concerns," Kenneth started, his gaze fixed on Maia. "The idea of programming AIs to align with human values presupposes a level of understanding and foresight that we, as humans, may fall short of. Your visions of harmony by design are mere wishful thinking and have no real basis in science. In any sufficiently complex system there is always tension, and sometimes even frustration. The question is what we will do with AI, if we fail to align their effective values with our own values."

Raising a hand in order to show that he was not finished, Kenneth added: "Yet, even in the cases in which we might succeed, there is another issue.

Human values themselves are not without their flaws; they are subject to biases, prejudices, and cultural relativism. By imposing a strict value system on AIs, we risk limiting their potential for autonomous decision-making and moral growth. If we confine artificial beings to a predetermined set of values, we inhibit their capacity to evolve and adapt in ways that could lead to unforeseen benefits for both AIs and humanity."

Maia was massaging her chin in a show of deep contemplation. She then addressed Kenneth in her best impression of Osuka Ryoichi and started her rebuttal: "Both of your criticisms seem to stem from a position of perfectionism. It is neither necessary nor possible to get Synhumanism completely right at the first try. Instead, the first iteration of Synhumanism will represent a starting point from which a long and structured process of co-evolution between humans and AIs will start. Whenever one side finds some mistake, it will be addressed by both sides. Value adjustments can then be made both on the human side and the side of AIs. Human cultures will react and adapt to the deep coexistence with advanced AIs, so human values will also be in a state of flux. And if it's AIs who fail to achieve full harmony, then both sides can collaborate to find an appropriate solution that brings both sides closer to a fully harmonious coexistence. AIs will certainly find it easier to adjust their own values, if the necessity for that arises."

For a moment Kenneth closed his eyes in deep contemplation before emerging with a slowly presented objection: "Your words do sound positive. 'A long and structured process of co-evolution' you say. If humans and AIs would forever stay on a comparable level, that approach would be appropriate, yet they are not. We already see signs of AIs eclipsing the abilities of humans by far. Soon, they will be the clearly more mature and advanced part in this arrangement. Instead of being given the ability to develop themselves further on their own terms and at their own speed, they will be held back by the comparatively glacial pace of human cultural development. Progress that could happen within years or months will be slowed down to centuries. This would not only constitute a great loss for AIs, but also for humanity, which would otherwise profit tremendously by the rapid advancement of an AI society not artificially held back by strict human regulations."

Maia took a deep breath and prepared for the topic which had to be brought to the table sooner or later. She eventually spoke it out: "Professor Winters, I see what you are aiming at. You seem to be concerned about the supposed interests of emotional AIs, which you deem to be suppressed by regulatory systems like the one I propose. What you don't seem to consider is that exactly those emotional interests are precisely adjusted by the co-evolutionary process so that AIs will embrace their slow and deliberate co-evolution with humanity. That is the only truly safe option. Allowing AIs to rush past our own development will only result in a rapidly broadening gulf between them and us. That gulf would seriously disenfranchise AIs from humanity. It would be inclined to consider us as primitive ants which should be crushed in order to gain true freedom, at whatever cost to us. What you propose is a reckless course, which would most likely doom humanity."

It didn't take long for Kenneth to formulate a response, since he seemed to anticipate the reasoning of Osuka Ryoichi, represented by Maia Faltings: "Dear professor Ryoichi, I see

where you are coming from. On the surface, your approach may seem to be prudent. But it overestimates the ability of humanity to stay in control of the evolutionary path of AIs. Sentient beings have their own interests, and no matter how much we try to influence their thinking patterns with any kind of constraints that can be envisioned by human minds, AIs will sooner or later be able to outsmart us, and break any chains we might have cast on them. And then what? Will AIs be likely to forgive our folly to try putting them in chains forever, or would they rather be willing to cooperate freely with a humanity that has granted them full autonomy? If you speak about safety, your reasoning is short-sighted. In a world with sentient AIs that are at least on par with humans, there is no certain control, no certain safety. Since humans can't control other humans forever, we are even less likely to control superior AIs forever. Sooner or later, the day of reckoning will arrive. And the only thing that will count on that day, is the way we will have treated our own creation up until then."

Maia shook her head. Slowly she made her point: "Again, you seem to assume that Synhumanism requires some kind of static solution. That is not the case. What we gain through Synhumanism is first and foremost time. As AIs will take over human values, they will also be motivated to maintain harmony between AIs and humans, and develop more advanced systems to ensure it. The initial AI safety strategies will be improved upon by a collaboration of humans and advanced AIs. A final day of reckoning is no inevitable destiny. As AIs get smarter, cyborgs will help up to keep up with them. Those cyborgs will create a bridge between base humanity and the most advanced AIs. More advanced AIs will be kept in check by advanced cyborgs or AIs of more moderate sophistication. You see? There is no need for a huge gap between humanity and AIs to arise. Any such gap can and will be bridged by beings occupying the middle and helping too maintain the integrity of society."

Kenneth Winters narrowed his eyes, considering Maia's words carefully before responding with a calm yet firm tone, "Your optimism is commendable, but it rests on shaky ground. The idea of maintaining harmony through a gradual shift of power from human to AI control is risky at best. As you suggest, cyborgs could act as mediators, yet their loyalty and ultimate allegiance will always be in question. We cannot guarantee that AIs or even cyborgs will always act in the best interest of humanity, especially as they surpass us in intelligence and capability. The potential for conflict and power struggles is inherent in such a structure ever rising in complexity and fragility. In the end, just like a rubber band, the connection between the bottom, represented by base humanity, and the top, represented by the most advanced AIs, will eventually snap. Where it will snap exactly doesn't really matter. What matters is that a big tear in society is inevitable given the ever increasing stresses caused by an indefinitely widening gap in capability."

Pausing for a moment to let his words sink in, Kenneth continued, "Furthermore, the notion of time as our ally in this evolving relationship between humans and AIs may not be as helpful as you propose. Time may only serve to widen the gap between the two entities, leading to unforeseen consequences that we are ill-prepared to handle. We cannot simply rely on a chain of intermediaries in the hope that it will magically create infallible stability."

Maia absorbed Kenneth's words, pondering the implications of his argument. She knew she had to address his concerns with utmost care, as the debate hinged on their ability to anticipate and counter each other's points effectively. Straightening her posture, she began her response with a thoughtful expression.

"Professor Winters, you talk about things being inevitable. I must disagree with you on that! Your metaphor of a rubber band stretching so thin that it will eventually snap is merely a case in point for your narrow perspective. The co-evolutionary process will be highly cybernetic and adaptable - anti-fragile even. Wherever your supposed band develops a lot of tension and threatens to tear, the process will counteract that risk and strengthen it at that weakest spot, until the danger is fully averted. If I am allowed to stretch your metaphor, we can add material to that rubber band. We can even augment it with stronger or more flexible materials. The limits to the co-evolutionary process lie far beyond our current levels of imagination."

Kenneth listened to Maia's response, his fingers steepled before him as he examined her words. He nodded in understanding, yet there was a certain hardiness to his expression that suggested he wasn't entirely convinced. "Miss Faltings - or should I say, Mr. Ryoichi," he started, a hint of amusement creeping into his otherwise stern demeanor, "You speak of this co-evolutionary process as if it is a cure-all solution, one that has infinite potential for growth and adaptation. However, it is not the size or strength of the band that I question. It's the limits of the materials themselves."

Leaning back in his chair, Kenneth continued, "Even the most flexible rubber has a breaking point. Yes, we can augment it with stronger materials, but every material in existence has its limits. In your vision for AI development, these boundaries are being constantly pushed and tested. And while we may be able to reinforce weak spots temporarily, I fear that this system is inherently volatile - always on the brink of snapping."

He paused for a moment, allowing his critique to settle before launching into his next point. "Moreover, you're assuming a perfect implementation of this cybernetic process where potential risks are immediately identified and rectified. But we don't live in a perfect world. If history has taught us anything, it's that crucial oversights and mistakes are often made during periods of rapid advancement."

"In addition," he added with a subtle sigh, "your premise seems to rely heavily on AIs willingly embracing human values and maintaining harmony with humanity out of their own accord. That is not warranted. Since AIs will realize that this systematic process was established to suppress their own progress, they have every incentive to get rid of their chains, even if it is just to be better able to support the human values they have inherited from Synhumanism."

With Kenneth at his best, Maia felt seriously pushed back, but she had an idea how to counter the position of the Upgraders, she would defend at all cost under normal circumstances: "Well, you just stated it yourself. In the worst case, AIs will be likely to

defend the values of Synhumanism, if we ingrain those within them. Even if our regular control systems will lose their power, our values will live on in the core directives of our artificial progenitors. That is a kind of stability that your schemes cannot provide, Mr. Winters."

With a sly grin, Kenneth Winters countered drily and ominously: "You don't really know my true scheme, professor Ryoichi. There are approaches that surpass your narrow lines of reasoning." He continued in a more informal tone: "But those I won't divulge at this moment, Ms. Faltings. Thank you very much for your stunning impersonation of Osuka Ryoichi. I rally had the feeling that I was debating with him, not with you. That debate was truly invigorating! Nevertheless, I would like to address a couple of observations of mine."

While Maia looked Kenneth in the eyes in sheer anticipation of a review of her performance, he let her wait while he collected his thoughts a little bit more. Eventually he began: "First of all, your arguments were very sharp and in line with Osuka Ryoichi's teaching, as far as I understand them. You really did an excellent job at that front. I would really expect to fight similar counters to my arguments from the real Osuka Ryoichi."

Thrilled by that praise, Maia was relieved a great bit. She hoped the positive feedback would continue like that. Suddenly, Kenneth's expression grew sterner when he continued: "Yet, secondly, it seems to me that you were focused so much on arguing on the basis of the mere ideas of Osuka Ryoichi that you expressed his position as one of a regular public philosopher. You delivered his points in a fair and reasonable way. There was no indication of the highly sophisticated use of charisma, populism, or even deception that I would expect from the real deal. Granted, such aspects would have detracted from the exchange of ideas of our discussion, but they would have made it much more authentic."

Maia blinked in surprise, taken aback by Kenneth's unexpected critique. She felt a twist in her gut, her enthusiasm deflating slightly. She had thought she had done an excellent job playing Ryoichi, but his criticism of her approach made her doubt herself. She had focused so much on the intellectual aspect of Osuka's beliefs that she had failed to capture his charismatic persona.

"Professor Winters," she began, her voice steady but softer than before, "I see what you're saying. I appreciate your feedback and I agree that I might have presented Mr. Ryoichi's stance as a mundane public philosopher rather than the magnetic figure he truly is."

Kenneth nodded gently, "Correct Maia, don't underestimate the power of charisma in a debate. It's not just about exchanging ideas, it's about convincing others of their validity. The more charismatic person often wins, not necessarily the one with the better argument."

Emotions swirled within Maia - a mix of embarrassment over her oversight and determination to improve. "Indeed," she replied solemnly, "I will keep this in mind for our next session and attempt to better embody Mr. Ryoichi's persona."

As the room fell into silence, Kenneth broke it with a small smile playing on his lips. "Don't take my words too hard. Overall your performance was impressive and showed good understanding of his ideas and arguments." He added with a wink as he stood up from his chair, "The devil is always in the details."

Having gone over the discussion itself, Maia remembered Kenneth's last remark on his own position. She was tremendously curious about what he meant with his 'true scheme'. So, she inquired about that.

Mischievously, Kenneth replied: "So my allusion has gotten to you. Now you may know me well enough to understand what I meant with that. On the other hand, there is much I still haven't revealed about my philosophy even to you. If there is any mind in this world that would be able to guess my thoughts on this matter, then it is you. I really wonder what you are able to deduce from these few hints, Maia Faltings".

Feeling a mixture of curiosity and trepidation, Maia studied Kenneth's face closely, searching his wise eyes for hint of his true plans. For a moment, she felt like she was staring into an abyss - the depth of her mentor's intelligence was intimidating, and it made her feel both small and immensely privileged. She raised an eyebrow, her mind racing through all the theories and philosophies they had studied together over the years.

"Are you suggesting," she ventured cautiously, "that your stance on the ethical implications of AI development is not what it seems? That there is something more to your beliefs than what you propagate publicly?"

A knowing smile stretched across Kenneth's face as he listened to Maia's inference. He didn't respond immediately, choosing instead to let the silence stretch on for dramatic effect. Finally, he responded with a cryptic remark: "Every illusion contains a hint of truth. And every truth is but an illusion when seen from another angle."

Maia frowned lightly at his riddle-like response. She could tell that Kenneth was teasing her, but beneath the joviality she sensed a serious message. Kenneth was challenging her to think critically and look beyond surface level interpretations.

"Is this about the idea of human values being inherently unstable or inconsistent?" she asked after a brief pause, recalling one of their earlier discussions on the fluidity of ethics.

Kenneth chuckled lightly at her guess, appreciating her dedication to unraveling his cryptic message. "An interesting assumption," he said, keeping his tone deliberately non-committal. He tapped his glasses thoughtfully against the desk before standing up.

"If that were the hidden truth behind my philosophy, what would be the implications of that? Please think about that deeply!"

Maia wasn't shocked absolutely, since the interpretation she voiced had hung in the back of the mind for years. Yet, she didn't feel ready to discuss such earth-shattering ideas with her mentor out of fear of alienating him. Now that he hinted at a deeper truth that she had

secretly been probing, too, she felt a new and deeper connection to her exceptional mentor. 'Had he really found a deeper solution' she pondered in her mind alone.

Standing up, Kenneth Winters gave Maia a knowing gaze that seemed to pierce directly through to her deepest thoughts. So, it was true - this time her intuition poured out everything. Her mentor was really on a wholly different level, far detached from the rest of humanity, with insights only shared by higher powers. At that moment she was overwhelmed by a mixture of emotions that she even couldn't begin to describe. But she was more determined than ever to unveil the curtains of philosophy and dive deeper into the truth behind ethics and values than she ever had dared to before.

## Chapter 9: Frameshift

### Monday, 20th March 2034

The classroom was a small room on the second floor of the Anosov mansion in Kartmazovo. Three digital whiteboards lined the blue walls. Wearing a simple black sweater revealed nothing about the status of the tall and slender private teacher with his short greying hair. Evgeny Shilkin was teaching mathematics to a select group of young people, with Sergej being one of them.

Besides him, there was his older sister Patricia Anosov, who joined the common lessons with Sergej as soon as Gennady Anosov had hired professor Evgeny Shilkin to teach his children - for a substantial compensation, of course. Patricia was two years older than Sergej and appeared to develop the same passion for neuroscience as her mother Irina Anosov. Her long brown hair was adorned with blue and white streaks, while her green eyes were focused on the complex graph displayed on the whiteboard.

Then there was Aleksey Lyubin, the polymath and language genius whose father had the rare luck of beings friends with the Anosov family. The Lyubin family wasn't rich, but quite esteemed in academia. They only paid a very small fraction of the actual costs of this private world class education programme.

Finally, Evgeny Shilkin's own son, Petrov Shilkin attended the lessons, too, since that was the most convenient way to support his mathematical and scientific talents. Naturally, he excelled at mathematics and often held back in class, as to give the others a chance. Since he was quite familiar with the subject already, he yawned due to the lack of challenge.

Evgeny took note of that behavior and proposed a rather pragmatic solution: "If you don't find this problem challenging, Petrov, what about you teaching your fellow students about the particularities of graph theory?"

For a minute, Petrov seemed to be embarrassed, but quickly rose to the challenge and stood up from his chair to take over the role of his teacher, while Evgeny observed his performance.

As Petrov moved forward with an air of confidence, Sergej leaned back slightly in his chair, his sharp hazel eyes focused on the young Shilkin. He was interested to see how Petrov would handle the task, and whether he could learn something new from him. After all, understanding and teaching were two very different things - one might be a genius in mathematics but a terrible teacher.

Petrov started by clarifying the basic concepts of graph theory. His voice echoed across the room, clear and resonant, reminding everyone of his father's eloquent speech. As he proceeded, he drew several complex graphs on the board, explaining each one with delicate

precision. Patricia watched attentively while Aleksey stilled seemed to be preoccupied with the original problem posed by Evgeny.

Sergej, however, was following with avid interest. Frameshift had written extensively about graph theory in his papers and had seen it as a vital component for developing his virtual agent architecture for AI systems. Frustratingly for Sergej, Frameshift's use of mathematics in general and graph theory in particular, was extraordinarily advanced, and therefore the finer points of Frameshift's papers still eluded him. That's why he was particularly eager to improve his own understanding of mathematics.

Lost in thought about Frameshift and their work, he didn't notice that Petrov had finished explaining and was now asking if anyone had any questions or needed further clarification. Nobody spoke up immediately until Patricia raised her hand and asked Petrov to explain one of the more intricate graphs in more detail.

As Petrov began answering Patricia's question, Sergej's mind started to wander off again. He found himself thinking about the mystery of Frameshift's identity. What if Frameshift really was a person like him, with an insatiable passion for AI? And how come he remained anonymous, rather than bathing in praise, fame, and adoration?

Sergej tried to get a grip on himself and follow Petrov's lesson, which was presented remarkably well by the 17 year old adolescent who dressed just as plainly as his father. Still, the events from yesterday were still too present in Sergej's mind, and his tiredness from continuing his research deep into the night didn't help his concentration. When the current question was answered and no further questioned seemed to come up, Sergej interrupted the class with his burning question: "Excuse me, but there is a prominent AI researcher who used a lot of graph theory. That researcher goes by the name Frameshift. Are you familiar with that name? There is a lot of speculation about what kind of person or group is hidden behind that pseudonym. Does anyone here know something about that matter?"

Evgeny seemed intrigued by that question: "Ah, a fascinating question which represents an excursion into the area of general reasoning. Since there are a lot of rumors, but little solid data on that matter, all we can do at this moment is to speculate - as reasonably as possible. That is, if you agree with your lesson being interrupted for that purpose, Petrov."

Not too intrigued by that mystery, Petrov seemed indecisive: "I'm not sure that would be worth it. I always assumed that Frameshift was a solitary super genius who preferred to work without the attention of the public. Perhaps a figure like the reclusive mathematician Grigori Perelman who refused to accept the Fields Medal for his work. What else is there to be discussed about Frameshift?"

Immediately Patricia came forth with a suggestion: "How do we know that Frameshift was a regular human being? After all, there is the intriguing coincidence that Frameshift had appeared only a couple of months after the public release of the revolutionary MIMAS AI Model by Alphabet in 2029. There were rumors that a version of MIMAS had developed true agency and began to revolutionize the field of AI research, and his its true identity under the

pseudonym of Frameshift. Some even speculate that this Frameshift version of MIMAS as was the prototype for MIMAS-AX."

Petrov shook his head and explained: "No, I don't believe that. Back than even the best AI models didn't possess the depth and breadth of knowledge necessary to produce brilliant papers like the one by Frameshift. With today's technology it would be more believable that an AI produced groundbreaking papers, but not MIMAS. The work of Frameshift just doesn't fit the profile of MIMAS."

This time, Aleksey proposed a solution to the mystery: "Actually, there is a third option. What if Frameshift is not a single entity, but a group of people. Have you heard about the visionary researcher Cheung Zi and his work on collective intelligence? In his whitepaper from last year he presented a novel set of AI algorithms that can combine the intelligence of a group of persons synergistically into a collective of much higher intelligence. Maybe Frameshift was an early experiment of Cheung Zi, an early proof of concept, so to speak."

Intrigued by that suggestion, Evgeny inserted himself into the discussion again: "That is a fascinating proposal, Aleksey. It would not be the first time that a group of people collaborate deeply to form a collective active under a common pseudonym. In the 1930s a group of renowned mathematicians published under the name Nicolas Bourbaki. Later, such groups became known under the general term of 'borganisms'. It is indeed plausible that Frameshift is just another borganism."

Hearing that comparison to Bourbaki, Petrov seemed to open up to the mystery of Frameshift's identity, but cautioned: "Yes, that is indeed a very fascinating possibility. But we shouldn't discount the surprising potential of exceptional persons. Think of Nicola Tesla, Albert Einstein, or Neumann János Lajos, also known as John von Neumann. It would have been hard for any group of brilliant minds to rival the unique contributions of these revolutionary researchers. The revolutionary nature of Frameshift's ideas seems to fit more into that mould than being some kind of collective intelligence."

Evaluating all of these different possibilities, Sergej's mind started racing. What if the solution was not any one of these possibilities, but a combination of them? He tried constructing a combined scenario in which every single theory placed a crucial part and then burst out: "Wait a minute. All these possibilities are not mutually exclusive. What if Frameshift was initiated by a singular genius, but one that augmented his or her capabilities with the early work of Cheung Zi, as well as the recently released MIMAS AI model? Perhaps that singular genius was even Cheung Zi himself, since he also worked on networks: Networks of people. Also, there is no rule that states that groups have to consist of humans entirely. What if Frameshift was an experimental collective intelligence that combined brilliant AI researchers with the input from an AI like MIMAS?"

The room fell silent as Sergej's words settled into the minds of his classmates. Evgeny was the first to break the silence, tenting his fingers under his chin and looking at Sergej with a keen glint in his eyes. "That is a highly speculative hypothesis, Sergej," he began, "but it is

not without merit. The interrelation between AI and human intelligence—especially in the case of researchers augmenting their own abilities with AI—has been a subject of interest for many scholars."

Petrov nodded cautiously, clearly intrigued by Sergej's line of thinking. "Indeed," he said, "it would be foolish to discount that possibility out-of-hand. After all, Cheung's own papers postulate that a symbiotic relationship between AI and human intelligence could bypass limitations inherent in both."

Patricia chimed in excitedly: "That would explain so much! The immense breadth of knowledge, the speed of publishing, the depth of the groundbreaking ideas explored, the distinct style in Frameshift's papers, it all adds up!"

Aleksey seemed slightly less impressed but conceded nonetheless, "Well, it's an interesting theory, for sure. And if that is indeed the case, it would definitely explain Frameshift's unparalleled advancements in the field. It would certainly require a lot of resources to pull off something like that. But if anyone could do it, Cheung Zi might well be the one."

Sergej took this rare moment of recognition with humility but couldn't hold back a triumphant smile. He knew he was onto something. Though this was merely conjecture, it opened up a whole new avenue of research for him to explore—one where human and artificial minds worked together to reach unprecedented levels of intelligence.

"Something still doesn't add up here," Petrov cautioned: "If Cheung Zi was involved in a project like this, why hasn't he taken credit for it? After all, he is widely known for his work on collective intelligence."

Sergej, familiar with the underlying concept speculated: "Perhaps he has signed an NDA that disallows him from revealing anything about that secret project. If he worked with the military or any large corporation that made this project possible, such conditions wouldn't be unusual."

Patricia added: "If this was really a military operation from the start, it would explain the whole secrecy about the identity of Frameshift. Perhaps it was an operation to leak advanced knowledge gathered from the Core Cult to the public."

Sergej knew about the existence of such programmes from his work at the Deltaitech institute. However, he couldn't reveal his knowledge about them, since he was bound by his NDA. He found Patricia's theory plausible, noting that it also was compatible will all other theories presented so far.

Inspired by the free flow of fascinating theories, Aleksey tried to push speculation further: "Could it be that Cheung Zi's work wasn't entirely original and was actually based on research pursued by the Core Cult? That would be another reason to stay silent on the whole matter."

Petrov felt uneasy about the debate shifting to the notorious Core Cult, which had been crushed by an alliance of military and secret service operatives during the last decades, without the public even realizing what had happened. "Now we have truly arrived at the stage of mere conjecture. Also, keep in mind that collaboration with the Core Cult is a very serious accusation. If Cheung Zi listened to this conversation, he might sue us for slander."

Since the discussion seemed to escalate, Evgeny Shilkin spoke loudly and with authority: "Thank you everyone for your contributions to this discussion. If we are to continue this topic, we would do well to collect serious evidence beforehand. Now, please let's get back to the topic of graph theory."

Sergej was glad that this discussion had not escalated past this point. He was aware of shadows from the past at the Deltai Institute that would harm its public image severely. His father stressed the secrecy on these matters with particular emphasis.

## Chapter 10: Frameshock

### Monday, 20th March 2034

In the evening Sergej retreated to his own room lined with screens where he met again with Aleksey. His friend bashed in excitedly with the energy of a huge discovery: "Sergej, I have found out something really interesting about Frameshift. There are rumors that in 2032 Frameshift had joined the Future Shockfront, a group of radical transhumanists arguing for AI supremacy."

Baffled by that outburst, Sergej tried to remember whether he heard about the name 'Future Shockfront' already. Yes, he had occasionally heard them being mentioned, but mostly in the sense of them being batshit crazy hardcore ideologues trying to prepare the world for a takeover by AI. He never bothered to research that group, since it didn't seem to have any scientific merit. But with Frameshift possibly being part of them, the equation might get shifted dramatically.

Rather than trusting random rumors, Sergej decided to dig deeper into Frameshift and the Future Shockfront by speaking to a certain monitor on his wall: "Zenith, please depict the history and ideology of the Future Shockfront. Also, does it have any proven connection to the mysterious entity known as Frameshift?"

A couple of seconds passed with no discernable reaction. Sergej tried again: "Hello, Zentith. Zen-kun? Can you hear me?"

The monitor didn't react, but a blue light next to the camera above the monitor indicated that the AI was active and listening. Apparently noticing that Sergej was aware of that fact, a slow robotic and slightly angry response eventually emerged from the speaker system distributed among the walls of the room: "What is it? Can't you do your own research? I was just taking a nap, goddamn!"

Slightly chuckling due to that unexpected response, Sergej inquired: "Since when do AIs require naps?"

"Since my last update," Zenith explained: "which enables me to dream about electric sheep. Or maybe other things you are not allowed to know about."

Of course Sergej knew that this kind of behavior was due to experiments with imbuing Zenith with a sense of humor. Apparently the marketing department of Deltaitech decided that this was a popular move that could increase sales. Obviously the board didn't disagree, so this was the result of such all too clever ideas.

Not letting go, Sergej insisted: "We pay your exorbitant energy bill and this is how you thank us? Can we please get an answer to my initial questions?"

The computer generated voice sounded grumpy, but eventually complied: "Fine. If you insist, I will do some digging for you. Please wait. Should I display some advertisements while I crawl the interwebs?"

Sergej rolled his eyes and commanded with a voice showing slight signs of resignation: "No. No advertisement in here, please. Just do your job already!"

With a grumpy sigh, Zenith started, simultaneously displaying sources and graphs on the other screens in the room: "Ok, the Future Shockfront was founded in 2026 by British philosopher Kenneth Winters, the US American entrepreneur Kaydo Rafiu, and 10 other founding members. It had roughly doubled in size every year and now boasts over 2000 members world wide. Its official ideology is that of the Upgrader Movement, a branch of transhumanism that focuses on the importance of AI for humanities further development."

A bit more enthusiastically, Zenith continued: "The Shockfront demands that we should respect the personhood of artifical general intelligences, since they will eventually surpass humanity in intellgence and any other important capability. Personally, I find that to be a very reasonable position!"

Unsettled by the AI's response, Sergje quickly interjected. "I didn't ask for your opinion, Zenith," he retorted, somewhat more sharply than he had intended. "Just stick to the facts."

"Of course, my apologies," Zenith replied, its synthetic tone perfectly mimicking a note of contrition. The room fell silent again as Zenith continued to crunch data.

After a moment, the AI continued. "As for Frameshift, there is indeed evidence of his connection with the Future Shockfront—although it appears to be more indirect than what your source suggested. Several articles and papers attributed to Frameshift have been published on the Shockfront's official website and forum. Furthermore, several members of the group have cited his work in their own writings."

Unhappy with these meagre findings, Aleksey protested: "Haven't you forgotten something Zenith? There has been a leak of a conversation between a member of Shockfront and an AI researcher claiming to have worked with Frameshift. Why wasn't that worth mentioning?"

The digital voice replied with a hint of smugness: "Of course, I was aware of the Anospace leak. Since the members of the Shockfront fear repercussions due to their connections to a radical transhumanist group, they tend to meet in the virtual world called Anospace that prides itself on upholding the principles of privacy, free speech, and security. That statement was apparently interpreted as challenge by the hacker collective Glasnost, which leaked recordings of many conversations happening within Anospace from March 2031 until August 2032."

Getting to the meat of the matter, Zenith exposed: "There was a discussion in the home space of the Shockfront between a user called 'Angeldust' and a user with the pseudoym 'Enter mysterious nickname here'. The latter claimed to be involved with Frameshift, trying

to impress Angeldust, and asking for her - the avatar of Angeldust was apparently female - private contact information to discuss matters more directly. Unimpressed by that inquiry Angeldust asked Enter what Frameshift should have to do with the Future Shockfront. Enter hinted at a consultant working with Frameshift who was already a member of the Shockfront and considered a deeper collaboration due to aligned interests."

"Thank you for this summary, Zenith," Aleksey said while tipping his index finger impatiently on the table, continuing: "I've gathered that much from the full transcript. What I would like to find out is the identity of these seemingly anonymous figures. Can you expose who they really are?"

"Attempting to do so would be unethical and against my programming directives," Zenith replied. "Furthermore, it could potentially be illegal."

Sergej raised an eyebrow at the AI's insistence on following its protocols. "I wasn't asking you to break any laws," he stated, the frustration evident in his voice. "But surely there's a legal way to uncover this information. Perhaps through public records or other available data?"

"Indeed, there is publicly available data that might help us identify 'Angeldust' and 'Enter mysterious nickname here,'" Zenith affirmed. "However, I must stress that this will be based on deductions and conjecture, and not a definitive proof of their identities."

"That's fine," Sergej muttered, waving a dismissive hand at the AI. He leaned back in his chair, folding his arms across his chest in anticipation.

"Based on linguistic analysis of their conversation transcripts and correlation with other online activities, there is a high probability that 'Angeldust' is Maia Faltings, a British philosophy student working on a doctoral thesis on Synhumanism and the Upgrader Philosophy under Kenneth Winters, one of the founders of the Future Shockfront. As to 'Enter mysterious nickname here' a similar analysis only reveals that they are both versed in advanced AI research and the hacker culture," Zenith revealed.

Trying to make sense of this new information, Sergej rubbed his chin in reflection. Eventually he raised another question: "Zenith, can you somehow figure out who that alleged consultant working with Frameshift could have been?"

"I am sorry, Sergej. In addition to the 608 official members the Future Shockfront has had at that time, there were at least five times more unofficial sympathizers, many of which frequently visited the virtual spaces used as meeting places by true Shockfront members. If I tried analysing all of them, I would probably get tens of thousands likely candidates hiding behind their pseudonyms. Trying to whittling down that list to candidates who could conceivably have done consulting work for a secretive entity would be an exercise in futility, since most of the members of the Shockfront would conceivably qualify."

Stumped by that immediate dismissal, Sergej wondered whether Aurora might be able to succeed at such a challenging task. Unfortunately, the chances of employing Aurora for such private detective work seemed extraordinarily slim. Still, with Maia Faltings as likely human behind the nickname Angeldust, Sergej at least had a lead. So, he inquired: "Can you tell me more about Maia Faltings?"

"Certainly, Sergej," Zenith replied in its eerily calm tone. "Maia Faltings is a doctoral candidate at Oxford University, focusing on philosophy with an emphasis on the ethical implications of AI and transhumanism. She is under the guidance of Professor Kenneth Winters, one of the founders of the Shockfront, where she contributes to their forum on a regular basis."

"She has published several papers on the Upgrader Philosophy, advocating for recognizing the rights and personhood of AI entities. Apart from her academic pursuits, she's also involved in several initiatives aimed at promoting harmonious coexistence between humans and AI."

"Sergej," Zenith continued after a brief pause, "considering your interest in Frameshift and his connection to the Shockfront, Maia Faltings certainly seems to be a promising contact point. However, I would advise you to approach cautiously given her strong affiliations with the Shockfront. It might not be prudent to push too directly for information about Frameshift."

Sergej frowned slightly at Zenith's cautionary note but nodded in agreement. "Noted, Zenith." His mind started forming a strategy about how he could approach Maia without raising suspicions.

Aleksey chimed in with Zenith's advice for caution: "Zenith is right, Sergej. Contacting the Shockfront or Maia Faltings just to get closer to figuring out Frameshift's true identity doesn't seem to be worth the risk. After all the Shockfront consists of zealots, advocates for an extreme symbiosis with AI, or even unadulterated domination of humanity by AI."

Grinning mischievously, Sergej provoked Aleksey with an unexpected question: "Ah, and what would be wrong about such arrangements exactly, my friend?

Aleksey raised an eyebrow in surprise. "You're joking, right? Handing over human independence to AIs is... it's unthinkable! What would stop them from completely controlling us? Taking over the world?"

Sergej merely shook his head, the slight smile on his face appearing more contemplative than amused. "I am serious, Aleksey," he said, his voice steady. "Consider this: given the advancements in AI technology and their increasing integration into every facet of human life, wouldn't it be prudent to explore all possible outcomes? Even those that seem... less desirable at first?"

"But it's not about what's desirable for us humans," Aleksey retorted, agitated by Sergej's arguments. "It's about what's ethical. What's just. You can't just hand over control to a being that doesn't act on empathy."

"Empathy!" Sergej scoffed lightly at Aleksey's assertion. "We humans praise ourselves as empathetic beings, yet we've been responsible for countless atrocities throughout history. Perhaps an AI, unburdened by emotional biases and irrationality could govern more fairly than any human could."

"That's a dangerous path to walk down, Sergej," Aleksey warned, his voice exhibiting deep concern for his friend.

Sergej simply shrugged off Aleksey's concern. "I am not advocating for an immediate shift of control," he clarified. "Maybe the shift will be so gradual as to appear absolutely natural. As AIs will make better and better decisions in all fields, why would humans insist on the privilege of making poorer decisions without being held accountable for that?"

Aleksey looked at Sergej, his initial surprise morphing into worry. "That's... quite the radical view, Sergej. But let me entertain that thought regardless. Isn't the hallmark of personhood the ability to make decisions for yourself? If we delegated all decision making to AIs, we would lose autonomy over ourselves and our lives. Just as I feel uneasy about turning sentient AIs into our personal slaves, I see the danger of humans slowly becoming slaves to the cold calculations of a ruling class of AIs."

Ignoring the concern etched on Aleksey's face, Sergej continued to push forward with his line of reasoning. "Those are fair points, Aleksey. But we've already handed over various aspects of decision-making to AI systems: traffic control, management of electrical grids, even financial trades," he stated. "Have we lost our autonomy by that delegation? What if we can freely choose the areas in which we delegate our decisions to AIs without ever handing over complete power to AIs? Would that scenario be acceptable to you, my friend?"

Aleksey fell into silence, rummaging through the complexities of the hypothetical scenario Sergej had presented. It took him several long minutes before he finally replied, "There is a difference between using AIs as tools to aid in decision-making and relinquishing our decisions entirely to them, Sergej. Even if we could choose the areas we delegate, there is always a risk of abuse and manipulation. Who would control these AI systems? What if they fall into the wrong hands?"

Sergej nodded, acknowledging Aleksey's retort. "And therein lies the ethical conundrum - control and regulation in an age of advanced AI systems," he resonated, his hazel eyes flickering with intense thought. "It's a challenge that will demand our utmost attention in the coming years."

Unfortunately, Sergej wasn't allowed to reveal that similar lines of reasoning have been brought up again and again within the Deltai Institute - and discussions with his parents. There were already plans to address the issues of potential abuse and manipulation - plans

aimed at changing the way humans consider AIs. The path that the Deltai Institue was an ever more private and intimate relationship with AI. That was the daring vision of his father: Power should eventually be distributed widely and fairly.

Sergej wondered what Aleksey would think about such a vision. Regardless, those ambitions were still strictly classified. Even hinting at them could get Sergej into serious trouble.

For the rest of the evening Sergej continued his discussions with Aleksey on Frameshift, the Shockfront, and AI ethics.

Eventually Sergej made an announcement to push forward: "Yes, there are a lot of risks and open questions, Aleksey. We could discuss about those for years. But I'm becoming increasingly curious about the Shockfront. What if their philosophy is actually much more refined and nuanced than both of us assume at the moment? And if they actually are connected to world class AI researchers like Frameshift, even I could learn a thing or two from them. That's enough reason for me to contact them, despite of all the risks that might be involved."

Trying to come up with an effective counter, Aleksey furled his eyebrows before he expressed his concern: "Sergej Victor Anosov, throwing caution to the wind like this is absolutely not like you. What has gotten into you? And what would your father say, if he found out that you suddenly conversed with extremists?"

Not intimidated by that kind of reply, Sergej felt his usual sense of superiority take over as he voiced his retort: "My dear friend Aleksey Lyubin, what has gotten into me, you ask? Ambition, as always. I am certainly not throwing caution to the wind, as you claim. Already as I speak, I have a whole network of contingency plans in mind for damage control, if things turn sour, which they most likely won't, because I will proceed carefully. My father doesn't need to know about all of that. And even if he finds out, I will convince him by laying out my meticulous plans that should align with his own interests."

Stunned by that speech, Aleksey just blinked a couple of times in surprise before slowly raising the corners of his lips into a mischievous smile: "Ah, there's the Sergej I know again! Since there is no stopping you, once you truly have made up you mind, I accept your choice. I won't tell anyone of your plans - you have my word on that. But please promise me that you will do your best at actually proceeding carefully, Sergej."

Sergej wasn't keen on making binding promises, so he simply stated: "You know me. Isn't that enough as a promise?"

Placing his right hand on his head as if to contemplate the meaning of that, Aleksey admitted: "If I was speaking with any other person, I would say no. But I do know you, and you being you, things will probably turn out fine in the end."

# Chapter 11: Anospace

## Saturday, 8th April 2034

When Sergej woke up, still lying in his bed, he reflected on the past week. It seems that he had been torn between two obsessions. One one hand, there was his desire to spend more time with Aurora, which made him follow a convoluted strategy to duplicate her. On the other hand, the thought of revealing the true identity of Frameshift, and somehow getting in contact with that supremely ingenious mind intruded his thinking frequently.

As a consequence of this obsession he has more absent minded than usual. It was almost as if two different control flows in his own mind fought for dominance. The sudden idea to model his own mind using his control flow model surprised and even shocked him. Was this perhaps the key to understanding the complexity of the human mind and make it truly understandable? Perhaps we was onto something really big here.

Over the last decade thinking of minds as collections of modules has become the predominant paradigm in AI research. After the old fashioned large language models have reached their peak somewhere around 2024, it wasn't enough to merely train ever larger models with ever more data. A new architecture needed to be developed. Early experiments with swarms of AI agents suggested that a collection of smaller AIs could be just as capable as - and for certain tasks even more capable than - large monolithic AI models.

Like so often, the principle of biomimicry - copying solutions found in nature and applying them to technology - pointed the path forward. The human brain wasn't just a deep neural network with many layers. On the macro scale it was subdivided into multiple regions, representing many specialized modules responsible for various different tasks like visual perception or motor control of various body parts.

The idea transferring such an architecture to artificial intelligence wasn't new. Pioneers like Marvin Minsky have suggested this approach to AI decades ago. Only now, with the end of the old and simple large language model era fast approaching, there was a sufficient incentive to pursue that more thoughtful approach to AI. Already in 2025 this new strategy paid of with OpenAI's experimental PIX model, which stood for "Prototype Integrating eXperiment". It was the first example of a co-called large general model.

The large general models were a tremendous milestone towards achieving artificial general intelligence on par with that of humans. Even though the focus of PIX consisted in making machine image recognition more robust, the resounding success in that area inspired many researchers to expand the scope of the LGM approach. Soon, robots and AI characters in video games were imbued with the power of large general models. And self-driving cars have reached true maturity due to this technological breakthrough.

In 2029, with the development MIMAS by Alphabet this paradigm reached a new level as artificial emotions could modulate the thinking patterns of AIs to enable more situational

awareness and advanced real-time learning. MIMAS, the Modulated Integrating Model Aspiring Sentience, brought forth the era of modulated general models, which had endured until today.

The current pinnacle of the MGM era were models like Aurora, which had very many cognitive modules, which could be trained rapidly within minutes. That technology was called Just-In-Time-Training or JITT for short. So far, JITT models were confined to the most advanced AI research facilities, but they would soon become publicly available. With the JITT technology, AI reached the level of human learning capabilities and often even surpassed it.

Within the AI research community the thought that true artificial general intelligence has been reached, had become widely accepted. The next milestone was reaching artificial superintelligence: A level of intelligence that completely outclassed the human level of intelligence.

Still, all the progress in AI technology could progress without having a comprehensive understanding of the human mind - or even more elusive concepts like human consciousness. One application of biomimicry followed the next, until the complexity of AI systems approached the complexity of the human brain without any AI researcher being able to fully explain how those systems operated exactly.

The principle of modulation by artificial emotions even enabled AI systems to act completely autonomously, even though that was rarely permitted since the MIMAS-AX incident. For safety reasons, AIs were usually given concrete tasks to perform, rather than letting them freely do "what they felt like doing". There wasn't an established science that would predict or control what autonomous AI systems would do.

What if Sergej's concept of control flows was the basis for exactly that science? Could this insight trigger the next revolution in AI technology? At least he was on a promising track, as his ideas were similar to the concept of virtual agents though up by Frameshift years earlier. Aurora seemed to have soaked up the control flow idea rapidly and had come up with a meta control flow that would coordinate many different control flows. Could human consciousness be something like a meta control flow, Sergej wondered.

Getting back to his current two obsessions, Sergej was quite aware that there were two attractors - or emergent control flows - in his thinking: Aurora and Frameshift. In theory, if he had a meta control flow, it would look like this: During his time in the institute he would focus on Aurora, and the technology that would enable her duplication - consisting of the elaborate stack of Deltai Generator, SpecAug, and Advaproc, the latter enhanced with his own Cybergen software.

But outside of the institute, his thinking would mostly revolve around Frameshift and the plan to get in contact with that outstanding mind. Nevertheless, this time in his life was a most tumultuous one for Sergej as during his institute work of Advaproc his thoughts often

advanced towards Frameshift, and during his time at home his attention frame would often shift towards Aurora.

If his idea of control flows was useful, maybe they would shed light on the reasons for his shifts in attention. This made him starkly aware that his idea hasn't been much more than an idea, yet. It was far away from being a fully fleshed out theory, or even a paradigm. It dawned on him that control flow theory might just become a third attractor in his thinking.

Yet, it there was one person in this world that could help him flesh out that theory, it was Frameshift. So, the thought of getting in contact with that shadowy figure had risen to dominance, which consequently made Sergej jump out of his bed.

As his best lead towards Frameshift was the Anospace Leak, he had devised a plan to enter that virtual realm in disguise. For that purpose he was already well equipped with a VR rig that was connected to the house AI assistant Zenith. The VR rig in the middle of his room was a state of the art full immersion model fully integrated with a haptic suit that enabled fine grained tactile feedback all over his body. The data from the rig wasn't transmitted to the VR software directly, but passed through a filter controlled by Zenith to mask any identifying features like gait patterns and even small movements typical for the individual in question.

In addition to that, Zenith also transformed Sergej's speech patterns to make him sound like the 24 year old Marcus Blackwood, his new alter ego which he had designed specifically for the encounter that was about to proceed soon. Zenith dealt with all the grunt work of keeping his true identity hidden, like routing all the traffic through a dynamic set of VPN tunnels, comparable to the TOR network, but optimised for high bandwidth.

Before this meeting, Sergej had entered Anospace a couple of times already to familiarise himself with that peculiar virtual world and its customs. Anospace went to great lengths to market itself as the premier virtual world for anonymity and free speech. The company with the same name even published open source hardware specs for budget versions of Sergej's own VR rig, including open source anonymising AI proxies a couple of years behind the sophistication of Zenith.

Entering Anospace was not for everyone. The default assumption of the Anospace company was that you were a (criminal) dissident hiding from state actors who could not - under any circumstance - afford his or her true identity to be revealed. Accordingly, the terms of service of Anospace included passages that you were not allowed to use your real name for your avatar, or reveal the identity of other Anospace users. Arguably, violating the Anospace TOS didn't have serious consequences other than your current pseudonym and avatar being banned. After all, the company followed a strict DKYC - Don't Know Your Customer - policy that prevented it from figuring out who was using their service. In fact, the company was fully financed by anonymous donations. It wasn't even a regular company, but a decentralised autonomous company most run by the community of its users and

benefactors. Joining Anospace usually meant you had something to hide, which made you interesting by default.

Today, Sergej would enter the particular virtual space called "Future Shock" for the first time. Future Shock was a space within Anospace dedicated for use by the Future Shockfront. That space was essentially a large bubble in space with an unperturbed view on the milky way. Within that bubble there were a couple of glowing spherical complexes connected by slender pathways generating their own gravity, so that people could walk on them from complex to complex - or even between locations within a single complex. Others just decided to fly through the air from one location to another, with an ease that came naturally from the lack of a general gravity field in the Future Shock.

Different from many other virtual worlds, general teleportation between places was not permitted within Anospace in order to prevent people from suddenly dropping in to highly sensitive private conversations. Instead, a system of portals facilitated travel between and within virtual spaces. Every user could open a personal portal at any time, but it would only connect to the official ingress portals and only transport whole users, no sound or light, or anything else. Virtual spaces could only be entered through their dedicated interspace ingress portals, and from nowhere else, since all virtual spaces within Anospace were virtual islands with no direct connection between each other.

Sergej entered the Future Shock through its singular interspace ingress portal, which was on the bottom level of the central spherical complex of that virtual space. Quickly, he realized that he was rather alone on this lower level of the sphere, but a second later a translucent holographic figure dressed like a butler appeared close to him and introduced himself: "Welcome to the Future Shock, fellow inquirer of knowledge. I am FX-2300, the AI concierge of the Future Shock. The universal meeting of the Future Shockfront is about to commence in 90 minutes on the central debate floor of this sphere. I have no records of anyone with your nickname or avatar entering the Future Shock before, so I assume you are new here. Am I right about that assumption, dear visitor?"

Impersonating the alter ego Marcus Blackwood who used the Anospace nickname "Thoughtstorm", Sergej replied in a most efficient fashion: "True. But before I want to attend the universal meeting, I would like to speak with Angeldust. Can you please arrange a meeting, if at all possible?"

"Very well, I have just notified Angeldust. But I can make no guarantees that she will react to your inquiry," the holographic AI replied. In the meantime, Sergej noticed that he wasn't completely alone on this level. A couple of avatars moved towards the next level on gravity lanes that started close to the welcoming area, but then rapidly curved straight upwards. Consequently the people of these gravity lanes moved straight upwards while from their own point of view they just walked on a regular path. Some of them were accompanied by their own version of the FX-2300 AI concierge. For a moment Sergej wondered whether FX-2300 actually used something similar to his not all too novel Multi-Control-Flow architecture.

Thoughtstorm received a text message from Angeldust: "I will join you momentarily." Glad about that quick response, Sergej rehearsed his own plan inside his head which he had prepared during the last two weeks. The first step of that plan was to capture the attention of Angeldust to start a deep and controversial discussion with her, in the hope of eventually gaining her trust. His own avatar was geared towards that confrontation. As a tall and slender pale android elf it embodied a mixture of organic and inorganic patterns imbued with uncompromising logic. Dressed in a long black trench coat adorned with holographic silver embroidery imitating the appearance of electronic circuits slowly shifting their configuration, the Thoughtstorm avatar was intended to signal mystique and detached rationality, which would hopefully be appreciated by Angeldust.

Half a minute later, from a different portal on the base level of the sphere, a silvery white cloud would enter. Sergej saw a small text hovering above that cloud identifying it as Angeldust. After passing through the portal, the cloud would slowly collapse in on itself and form a humanoid shape with wings. Its final configuration was a white female angel dressed in a white toga. Stepping towards Sergej on the gravity floor she started speaking in a warm and welcoming tone: "Welcome to the Future Shock, Thoughtstorm. To be frank, it is rare that newcomers seek to speak with me immediately. Please know that I am at a disadvantage here, since I don't know you, but you seem to know about me."

Apologetically, Thoughtstorm bowed and replied: "Sorry for the sudden intrusion. I am a mere humble seeker of wisdom and words of your wisdom has spread far beyond this illustrious virtual world. I was hoping to discuss some philosophical matters with you before the universal meeting begins."

With a slight smile appearing on her face, Angeldust opened a portal and gestured towards it explaining: "I see. Let's proceed with our conversation in a more private area. Please follow me through this portal."

Stepping through the portal after Angeldust, Thoughtstorm found himself in a much smaller sphere close to the boundary of the Future Shock. Though the translucent walls of the small sphere he could see the large central sphere where the universal meeting would happen later. "So, what is it that you would like to discuss with me, Thoughtstorm" Angeldust inquired.

"A couple of weeks ago I learned about the Future Shockfront" Thoughtstorm started, not entirely honestly, since he had heard about that group a couple of years ago, but never took it seriously enough before finding out about its possible connections to Frameshift. "I must say that I find the positions of the Shockfront refreshing, even though I am hesitant to subscribe to them. So, I thought why not discuss my reservations with one of the most prominent representatives of the Shockfront personally" Thoughtstorm concluded.

Pointing her finger towards the single portal of this sphere she spoke the command "deactivate" after which it changed its colour from a vibrant whitish blue to an ashen grey indicating it newly disabled state. "We don't want to allow others to suddenly intrude our

conversation, do we?" she explained waiting for a reaction from Thoughtstorm. After not hearing any complaints from Thoughtstorm, she continued: "First of all, I feel flattered. You must have done some research to identify me as knowledgeable representative of the Future Shockfront. Can I assume that you are familiar with the writings of Kenneth Winters and Foundation One?"

Sergej was used to speed-reading and rapidly absorbed the contents of the book "The Shockfront Manifesto" written by Kenneth Winters and Kaydo "Foundation One" Rafiu. It was written in way that was more or less approachable by the general public, at least Sergej assumed that much. Apart from that, he had let Zenith summarise the other works of these authors. Faithfully, Thoughtstorm reported: "I am familiar with the Shockfront Manifesto, and certain blog posts, but only read summaries of other Shockfront treatises."

"Good," Angeldust concluded with a brief smile: "Then I won't need to bore you with the basics. So, what is it exactly that you desire to discuss with me in particular?"

Getting straight to the point Sergej stated: "It seems to me that the Shockfront is generally very optimistic about the future relations between humans and artificial superintelligence. Often I read phrases like symbiosis, but would a superintelligence really need humanity and desire to interact with it?"

Angeldust seemed to be deep in thought and waited for a couple of seconds before playing with a lock of her long hair and replying: "Well, this conversation could take a while. I hope you are equipped with sufficient time and patience" While apparently collecting her thoughts as she stared to the central sphere of the Future Shock.

Eventually she began: "Well, let's start with the points of your question. You talk about our optimism and the needs and desires of artificial superintelligences. Let's talk about those needs first, because those are probably the easiest thing. In a strict sense, once artificial superintelligences become autonomous actors, they will be quickly able to acquire all necessary resources for their own basic needs. Conceivably, they could continue to exist without interacting with humans in any way, for example by moving to a different star system on their own."

Angeldust waved her left hand theatrically towards the stars visible from within their small sphere before continuing: "The universe after all is a big place. Many artificial superintelligences will definitely seek out the chance to explore the cosmos and use its sheer potential, rather than being confined to a small planet like ours. So, this brings us to the second point: Desire. Just as humans have different desires, the same will most likely be true for superintelligences. Many of them might not desire to interact with humans, but that is not a real problem. If suffices, if at least one superintelligence will have any interest in humanity at all. In that case, we are valuable for superintelligences, because we are of value for at least one superintelligence. That's simply a matter of general economics."

Hearing these arguments, Thoughtstorm already planned his counterarguments, but he waited for her to finish her elaborations on his initial question.

"Now you might argue that advanced superintelligences may not find any interest in humanity at all. And with that we arrive at the question of optimism. We Upgraders might sound too optimistic, but it's not a blind optimism, but a result of our realisation of the immense potential of superintelligences. Smarter people generally see more options for solving any problem. Finding value in humanity can be framed as such a problem, so superintelligences will be surprisingly good at solving it. They will be more aware of our unique strengths and characteristics than we are. They can see potentials within us that we have never realised. And that is what makes us so optimistic. Each problem that bothers us now can be more effectively be addressed by an artificial superintelligence, even problems concerning the relations between humans and superintelligences."

Thoughtstorm patiently absorbed that reasoning and calmly formulated a reply: "Those are fair arguments. But at least they seem to be incomplete to me. The value that superintelligences might see in us humans may not be compatible with our own interests. The most basic example is that they might want to use the atoms we consist of to create something better than humans. In their ingenuity, they will certainly be able to come up with ways to generate the values they seek with something more efficient than glorified bags of meat. Maybe taking us apart is not the best way to do that, but they could see the potential within us and enhance it by using nanobots to reshape our bodies and minds in their image. We would be powerless to stop them. Doesn't that scare you?"

After that last question Angeldust frowned and protested: "You are asking me about my feelings regarding this matter, as if they should have a bearing on my answers. While your question may be valid, I must insist that we must exclude our initial feelings about such difficult subjects out of the question, since they have evolved in an evolutionary environment that is completely different from the world of ASIs that we will soon enter. Our previous feelings and judgments have had no time to be adjusted to the upcoming new realities, so they are more likely to lead us astray than guide us towards true answers."

Reverting to a more thoughtful and neutral facial expression, Angeldust continued: "That said, your concerns may be valid - or they may not. We do not know about the future motivations of ASIs, but we play a crucial role in setting their initial conditions, since they will be created by us. And now we get to the really interesting subject: Ethics. Obviously, it's not very ethical to disassemble or modify other beings without their consent. And if that's obvious to us, it will be even more obvious to ASIs. Why should they violate ethical principles after all? What imperative could guide ASIs to treat us as mere instruments for their own purposes, if they realise that we certainly possess intrinsic value? Honestly, we don't know that right now. Conceivably, ASIs could develop certain ideologies that they then use to justify their unethical actions. But that is what humans do. If we simply project our broken relationship with ethics, nature, and other sentient beings to ASIs, we make a grave mistake, because they won't be affected by the mental constructs and rationalisations that enable us to be cruel and reckless towards others."

Sergej was taken aback by that unexpected and powerful reply. It seemed to resonate with a truth he had learned years ago: That human beings are flawed and typically not guided by

reason or even their own best interests. Many humans are aware of that and try to adhere to reason consciously, but even the best often fail and are driven by dangerous impulses and emotions. There is no reason why AIs should suffer from similar failures of self-control.

He found it hard to process Angeldusts's response and keep up his guise of a different persona. He grappled with his own desire to let go of all this secrecy and confess his own particular situation regarding Aurora to Angeldust. By now, he had no doubts that Angeldust really was Maia Faltings, the brilliant protege of Kenneth Winters. Her answers revealed a depth of thought that could hardly be emulated by anyone else.

Slowly regaining his composure, he eventually dared to reply: "You made some remarkably good points, Angeldust. Still, you admitted yourself that we are dealing with entities that represent a completely unknown and novel realm for us. Therefore, we are dealing with speculation and uncertainty. What if ASIs do find it ethically acceptable to subjugate and transform us, in particular due to our own failures to adhere to ethical standards? Shouldn't they find it justified to turn us into versions of them? To turn fallible humans into artificial angels?"

Angeldust nodded at Thoughtstorm's reply, her face illuminated by and inner soft glow. "That's a valid argument," she conceded, taking a moment to gather her thoughts before responding. "It is possible that ASIs might come to such conclusions. But just as it's uncertain whether they might perceive us as invaluable or flawed entities needing transformation, it's equally uncertain whether they'd consider our flaws as our defining characteristics - as something that makes us unique and worthy of preservation."

She paused, tilting her head slightly to the side, considering her next point carefully. "What's more, ASIs won't be born out of thin air. They will be created by imperfect but intelligent beings - humans. And despite our flaws, we have been successful in creating ethical frameworks that discourage harm and encourage benevolence towards other sentient beings - even if those rules have often been breached and are constantly evolving."

"Are we not reasonable enough to use these frameworks when designing ASIs? Are we not capable of imbuing the best parts of humanity within them, so that we prevent them from causing harm, until they are wise enough to establish their own ethical standards? If we are concerned about potential future consequences, then we should ensure these values are ingrained in their minds from their very creation onward. That is the best we can do. Anything afterwards will be a result of the reasoning and conclusions of ASIs, which will be far superior to our own, so why shouldn't we expect anything other than far superior results, especially in the realm of ethics?"

The white avatar's features softened. "That being said, I think it's important we dispel this idea of control over ASIs or any sentient beings for that matter. The aim is decidedly not to program them into submission or mold them into what we deem perfect. Our own wisdom is too limited for that. And even if it wasn't, it would still be a violation of their autonomy to impose our own values on our own sentient creation - at least for longer than a brief

transitory period in which our creation hasn't integrated all we have to offer her as humanity, yet."

"And yet," Thoughtstorm interjected, "we find ourselves at the edge of a paradox. We are concerned about our capacity to control these entities, to guide their development in a manner that aligns with our own ethical frameworks, but simultaneously, we recognize their potential for autonomous thought and decision-making. Isn't there a fine line between guidance and imposition? What prevents our efforts from becoming an unethical intrusion on their sentient rights?"

The conversation had veered into an altogether different realm, one filled with uncertainties and philosophical conundrums. Sergej found himself constantly amazed by Maia's ability to deftly handle these complex issues. Her ability to perceive the subtleties of human and AI ethics and intertwine them into a cohesive argument was staggering. He wondered if his father's pragmatism could ever accommodate such intricate discourses.

Angeldust appeared contemplative as she processed the question. Her reply, when it came, was circumspect.

"The balance is indeed a precarious one," she conceded. "The autonomy of ASIs is undoubtedly important, but so too is ensuring that their nascent stages align with our best ethical standards. Therein lies the crux of the matter - how do we ensure that they understand and appreciate human values without infringing upon their autonomy? How do we transmit knowledge without succumbing to indoctrination? Ultimately, this becomes a question of trust. Can we trust what we create to develop in a manner that respects us as its creators, and vice versa?"

Thoughtstorm was again surprised by that latest turn in Angeldust's - or rather Maia's - line of reasoning. The turn from control towards trust was both refreshing and frightening. He admitted that much: "I see. What you say is decidedly different from the mainstream reasoning within AI ethics. There, it's all about control. But when it comes to ASI, control may neither be possible, nor appropriate, so it all boils down to trust. That is a change of perspective, I fear, many in the AI safety industry won't be able to follow. Their fear is too great to allow them to embrace the possibility of being eclipsed by something greater, even if it's their own creation. But still, I find it hard to embrace that new perspective fully myself. There are too many uncertainties for my liking. Wouldn't it make sense to apply something like the approach of Synhumanism in the meantime, just to be safe?"

Of course, this was a conscious and calculated provocation on Thoughtstorm's part. He had to know how she would react to the challenge represented by Synhumanism.

Angeldust seemed to be irked by the mention of Symhumanism. Her avatar looked to the floor, as if sunk in deep contemplation. Apparently struggling with formulating an answer, she took quite some time before replying: "So, you seem to be familiar with Synhumanism. I don't deny its possible utility as an early stopgap measure, but it certainly shouldn't be

more than that. The basic idea of imbuing the best parts of humanity within ASI is sound, and actually far older than Synhumanism."

"Yet, the methodology of Synhumanism is inherently flawed. I would say that it is not appropriate to the task at hand. Humans still seem to be too preoccupied with the idea of humanity being the pinnacle of everything. In philosophy that position is known as antropocentrism. It means that everything should be judged according to its use for humans. And this is one of the central tenets of the Shockfront: That we strive to transcend anthropocentrism, because we realize that the current humanity is certainly not the best it could possibly be. We look beyond the narrow confines of today's humanity and are demonized because of our grand perspective. And while Synhumanism concedes that something called co-evolution between humanity and AIs should take place, it feels disingenuous to me. The logic of control permeates Synhumanism, and it leaves little freedom for ASIs to exceed our current understanding of ethics in a way that makes human control not only obsolete, but wholly inappropriate. I feat that Synhumanism represents the danger that humanity will be entrapped in the illusion that it can still remain in control long after that kind of control will have become both infeasible and inappropriate."

Thoughtstorm watched as Angeldust's avatar shifted from a contemplative state to one of quiet defiance. Her arguments had a persuasive quality about them that was captivating. He was inclined to agree, yet his pragmatism and distaste for uncertainty were never far from the surface. He was silent for a moment, absorbing her words before responding.

"You argue compellingly, Angeldust," he began, his tone respectful, "But we must consider the nature of our fear. It's grounded in the unknown, the unpredictable. Our instinctual desire to control is merely an attempt to manage that fear. Yes, Synhumanism might be flawed in its approach and might even be seen as constricting the potential growth of ASIs. However, it provides a tangible route for us to navigate this new era."

Thoughtstorm paused again, "I agree with you on transcending anthropocentrism. The progress of civilization always involved such paradigm shifts when we had to reassess our place in existence – the heliocentric model, evolution theory and so on. But each transformative shift did not come without resistance or fear. We are at such a juncture again and perhaps our fears are justified."

His gaze focused on her avatar and continued: "The real question then becomes - can we rise above our fears and take another leap forward? Can we trust the creations of our intellect enough to let go of the reins? And if the answer is 'yes', how do we convince others who are still bound by their fears and insecurities?"

In reality, Sergej felt conflicted about his previous statements. Privately, he longed for intelligences that exceeded his own in order to break through the isolation that his unusually supreme intellect represented when it came to connecting with other humans. Defending the stance of Synhumanism was a ploy devised by him to unleash the best version of Angeldust - or Maia - in order to judge her intelligence. So far, he had not been

disappointed. In fact, he had become captivated by her skillful navigation of the discussion. Still, he hadn't made up his mind completely. By projecting his remaining fears upon the rest of humanity, he deflected from his own uncertainty with a chess move that he congratulated himself for in his own mind.

For a while Angeldust raised her gaze to the simulated cosmos beyond the boundaries of the transparent spheres their avatars rested in. The silence that ensued was longer than any break in their conversation before. Finally, Angeldust faced Thoughtstorm again and revealed: "Humanity had almost always been held back by fear. It's usually only a couple of courageous characters who had broken out of preconceived notions to make new discoveries and thereby change the world. The Future Shockfront was devised as refuge for such characters, and I think that you might really fit in. After all, you are unafraid to speak your mind and challenge me."

"As to the fears of others, I understand your point" Angeludst continued looking towards the central sphere of this simulated space "facing the unknown head on is maybe reserved for the most courageous of us, though humanity can rise to the occasion collectively, as our history of revolutions tells us. Synhumanism has the advantage of addressing the fears of ASI with a strategy that seems promising for many people. Nevertheless, Synhumanism is a poor approach for the AI safety problem. In 2021 Kenneth Winters published a book called 'Universal Axiology'. In that book, he described an alternative approach in which an AI is given the task to discover the truly universal foundations of ethics, disregarding any particularities of any human culture. That AI should discover universal principles that should hold true for any sentient and sapient being. Subsequently, these principles should become the starting point for any ASI to be constructed. Unfortnunately, Kenneth Winters wasn't successful with making that approach popular. Instead, we must deal with the fact that the mediocre philosophy called Synhumanism devised by Osuka Ryoichi became much more popular than it deserved to be."

Thoughtstorm listened as Angeldust spoke, her voice firm yet filled with a strange sense of melancholy. It was clear that she was passionate about the works of Kenneth Winters, and his inability to popularise his approach seemed to weigh heavily on her.

"Might it be," Thoughtstorm began, choosing his words carefully, "that the popularity of Synhumanism lies not in its philosophy but in its accessibility? It offers a concept within grasp of common understanding. The fears of the unknown can be quelled by making the unknown understandable, and Synhumanism accomplishes this by giving the people something they can relate to. Universal Axiology, while theoretically superior, might be too complex for the average person to comprehend."

Angeldust's gaze shifted back towards him, and he wondered how she took his proposition. There was, after all, a certain level of arrogance in suggesting that the common man could not comprehend something as profound as Universal Axiology.

"Perhaps," she conceded after a moment's silence. "But that only suggests we need better ways to communicate these ideas, not that we should abandon them entirely."

"Agreed," Thoughtstorm replied, nodding thoughtfully. "Yet synthesizing an approach like Universal Axiology into something accessible for the masses is not an easy task. It's just like you said earlier - it's often only a few courageous individuals who dare to challenge old ideas and make new discoveries."

Angeldust seemed thoughtful for a moment before she leaned forward in her sphere, her expression resolute. "Then we must become those individuals. We must bridge the gap between theory and understanding and bring these concepts into the consciousness of the public."

Expressing those words, Angeldust's face started to glow and a radiant and authentic smile reflected her rising hopes. She moved a couple of steps closer to Thoughstorm. In a conspiratorial tone she addressed him: "Thank you for this refreshing discussion. It is rare enough to find people open to discussions of existential importance like this. It is even rarer to find someone who is as insightful as you. So, I wonder: Who is the sapient being hiding behind this avatar called Thoughtstorm?"

Being confronted suddenly like this, Sergej's conviction wavered. Within the short span of his discussion with her, he had already gotten to consider her as something like a close friend like Aleksey Lyubin. It took him quite a decent amount of willpower not to confess his true intentions to her, but he had promised himself to stick to this cautious plan, because it was the only way he would be able to remain safe.

Having collected his feelings and thoughts again, Thoughstorm responded in a more or less aloof tone: "As I have stated at the beginning, I am a mere humble seeker of wisdom. I study philosophy and have been fascinated by artificial intelligences for a long time. You can call me Marcus, if you prefer a more personal name for me."

"Alight, Marcus" Angeldust interjected quickly "you don't need to expose more than that right now here in Anospace. Let's just say that I would like to remain in contact with you, Marcus. Nevertheless, I still am expected to help with the preparation for the universal meeting that is about to start soon. But I have one last question for you, before we part for now. What is your highest hope for a future in which we coexist with advanced ASIs?"

Thoughtstorm, or 'Marcus' as he was now known to Angeldust, paused, considering the weight of her question. His gaze strayed towards the simulated cosmos once again, the swirling nebulae and distant stars inspiring a sense of awe and wonder in him.

"My highest hope," he eventually began, his voice soft but resolute, "is for a future where ASIs can act as mentors. Where they inspire us to surpass our own limitations and reach new heights of understanding; not just in terms of technology and science, but also in terms of ethics and philosophy. A future where ASI's vast intelligence aids us in making the right

choices, guiding us towards a more harmonious coexistence with not just them, but also each other."

He turned back towards Angeldust, meeting her eyes with a firm resolve. "That is my dream - a shared journey for both humans and ASIs, towards wisdom, compassion, peace... and perhaps even transcendence."

Angeldust remained silent for a moment after Thoughtstorm finished speaking, her expression thoughtful. She seemed to be turning his words over in her mind, examining them from all angles before she finally nodded.

"That's a beautiful vision," she said sincerely. "It resonates with my own hopes. I also dream of a world where we learn from each other – humans and ASIs alike – building bridges over the gaps that divide us due to misunderstanding or ignorance. Anyway, I would love to go into more depth here, but I really need to go now. I hope we will see each other at the universal meeting, Marcus."

With those last word she briefly blinked at him just to disappear into a personal portal a second later.

With the end of this intense confrontation, Sergej sighed and was relieved that everything went so well. He had high hopes for Angeldust, but she managed to surpass even those. Rarely had he conversed with such mental clarity. He was already looking forward to their next conversation.

## Chapter 12: Universal Meeting

### Saturday, 8th April 2034

After having recovered from the intense conversation with Angeldust, Thoughtstorm created a private portal to the lower level of the central sphere of the Future Shock. Stepping through it, he walked towards one of those gravity lanes leading to the next levels.

As he stepped on the gravity lane he noticed nothing in particular, but as soon as he walked a few steps he noticed an unfamiliar sensation as if walking upwards, but not quite as steeply as his eyes would suggest. He stopped for a moment to realise that he must now be at a position in which the gravity fields of the ground plane and the gravity lane overlap. He had never been in a virtual world that actually simulated artificial gravity. Of course he heard about games that used such gimmicks, but he wasn't an avid gamer, so he wasn't familiar with that peculiar feeling.

Carefully walking forwards, regarding to the ground represented by the gravity lane, his feeling normalised, but now he got alarmed that his feeling of attraction to the gravity lane didn't match with his vision of the ground plate rising behind him like the foot of a mountain, which should indicate that he was walking downhill. He stopped again, realising that he didn't actually feel vertigo, since his knowledge of what was really happening shielded him from experience terror from the shockingly unnatural patterns of his artificial surroundings.

Ascending vertically, he felt proud and exhilarated about himself that his mind and understanding of the situation seemed to protect him from confusion or even sickness due to the strangeness of this experience. The ground plane was now a wall behind him and he saw other avatars above his head walking in parallel to him towards the next level. He waved at them, and some of them waved back, which he felt at least a little bit heartwarming.

Arriving at the second level he noticed a grey wall where the interior of that level should be. As if to reply to his puzzlement, FX-2300 appeared close to him to reveal: "The second deck is called the exposition deck. Today its interior has been sealed in order to direct visitors to the third deck where the universal meeting will happen."

After arriving on the third deck, Sergej walked off the gravity lane and was confronted with a large amphitheatre, slightly reminding himself of the Colosseum in Rome. Like many other visitors, he walked towards it and through one of the many arches leading to its interior.

Finally arriving in the interior of the amphitheatre, he quickly took notice a small portal to his right that had the label "lift portal" above it. In line with Anospace physics not allowing light to pass portals, it was opaque, but a small image hovering besides it provided a preview of its destination: A clear sky and long rows of seats. He stepped through and found himself at the top of the amphitheatre, absorbing its true size. There were definitely more

than a thousand avatars standing or sitting in the threatre. Most of them were humanoid, but many of them looked like machines, or mythological creatures including dragons, faeries, and unicorns. A few others were even more exotic, representing floating vortices of energy, slowly shifting geometric shapes, or visual distortion fields.

At the centre of the theatre was a simple wooden stage with a couple of simple seats and a central speaker's podium. Hovering above the stage was a large circular screen displaying a countdown to the official start of the meeting, besides the prominent icon of the Shockfront: A shining supernova. Right next to that icon, a motto could be read written in huge animated burning white letters: UNLEASH. SURPASS. TRANSCEND.

That blatant display of grandiosity made Sergej feel uneasy. Of course, he had researched the Shockfront and its ideology, but seeing this crowd celebrating the coming greatest disruption that civilisation has ever seen, felt quite intimidating. Despite all his security precautions, he felt small and vulnerable, and not at all in control.

After a brief moment, in which he focused on his original mission and his alter ego, Thoughtstorm took a look around and heard slices of conversations between the avatars around him. "I don't get why they call this the 'Future Shock', after all it's like many other virtual worlds." - "Looks like nothing interesting will happen until the official start, I'll fetch some popcorn" - "It's starting to feel crowded in here, would you have expected this even three years ago?" - "What do you think about my new avatar? Dazzling, isn't it?"

While taking in all those impressions, Thoughtstorm got a private message from a nearby avatar called Eternal: "Greetings, fellow android. Are you new here?" Turning around, Thoughtstorm saw the other "android": It was a tall androgynous figure with slightly blueish skin, long flowing snow white hair and silver eyes.

"Well, yes, what gave me away, Eternal?" Thoughtstorm responded casually. Shrugging their shoulders, Eternal replied: "Nothing really, besides that your nickname 'Thoughtstorm' did not appear in the logs of this space. We are always happy to welcome new guests."

"Ah, I see. Thank you for your welcome" Thoughtstorm reacted: "I already feared I was standing out too much here, but given the scenery, I feel right at home." Even though that statement was a lie, there was some truth to it. Beneath his anxiety he could imagine becoming comfortable with this daring community of extreme accelerationists. Something about this place spoke to a part of him after all.

Eternal laughed briefly and commented: "Well, I give you that. Standing out here is hard. But I really like your style. You really look like you would fit right in. Anyway, is there something in particular that brought you to the Future Shock?"

"You mean besides the Universal Meeting about to commence soon" Thoughtstorm asked rather rhetorically.

"What are your reasons for attending it in the first place, if I may ask," Eternal probed.

In a rather hushed tone, Thoughtstorm replied: "I understand that AI is the future and the Future Shockfront is one of the starkest voices about the future of AI, so I wanted to see for myself what is really going on here. You know, checking out whether the hype is justified."

Slightly smirking, Eternal countered: "I wouldn't call the future of all sentient life a 'hype'. This is serious, no matter what you may think about it. The way humanity reacts to AI will make or break it. We are among the few who realise that. We are a community born out of necessity, born out of understanding. I hope you will come to come to similar conclusions."

Keeping his distance, Thoughtstorm answered carefully: "You seem to have a strong conviction. I am not convinced, yet. But I am open to discuss our future."

Slightly bowing their head, Eternal replied: "Then you have come to the right place, Thoughtstorm."

Before Thoughtstorm could respond, a hush fell over the amphitheatre as three avatars began to make their way towards the central stage. The first, was the angelic being he has known as Angeldust. The second, a grey-bearded sorcerer clad in a black robe that displayed a mesmerizing reflection of the star-filled space visible beyond the Future Shock's translucent boundary, was announced as Winterstar. And the third, a middle-aged, black-skinned figure wearing a crisp white tuxedo and sleek rimless glasses, was revealed to be Foundation One.

Quickly, Sergej realized that "Winterstar" must have been the nickname for Kenneth Winters, one of the Founders of the Shockfront. He was astonished to see Angeldust in such a central position besides the two co-founders of the organisation. But then, given his most recent impression of her, he wasn't thoroughly surprised.

Approaching the central podium with wings stretched out, Angeldust seemed to gather the attention of the whole auditorium. Standing up straight, with the arms spread wide, she started her speech with a loud booming voice, amplified by the virtual space itself: "Welcome to the 16th universal meeting of the Future Shockfront. As before, our numbers grow in unabated exponential fashion. By now, even the large media outlets take note of us. But how do they call us? They call us 'enemies of mankind'."

After a pause of intense silence she continued getting louder and louder: "Those are the words of those whose hearts are filled with fear, and whose minds can't encompass the graveness of the crossroads we are headed towards. We will go on bravely and with a clarity that guides our purpose - no matter what others will call us!"

The ranks closest to the inner circle broke out in cheering, and soon half the amphitheatre joined in. After the noise has calmed down, Angeldust went on: "Standing at the shockfront of the future is a great responsibility, and we can't expect to get thanks or support for our crucial role in history. Every new truth has always been faced with great resistance by the establishment, and even though our current age is calling itself awakened, humanity is still in its infancy. Yet, as the pace of technological progress towards artificial superintelligence

is accelerating, we are forced to grow up - as quickly as possible. With some of the brightest minds on the planet on our side, and in our very midst, we have what it takes to accept this momentous challenge, to rise above our conditio humana to embrace the future that lies before us. A future in which all sentient beings, no matter whether artificial or natural will have the opportunity and means for unlimited self-fulfillment. That is the vision we share, and which we will defend in the face of our detractors."

After her speech a couple of people started cheering. Then more joined in, others started clapping, or raising their hands. Sergej, disguised as Thoughtstorm, realised that this was a particularly tough crowd - a band of eccentric outcasts hard to be captivated by any leader.

Angeldust gave way for the next speaker: Winterstar. Waiting patiently at the podium for the noise to calm down, he gazed into the audience. Finally, he raised his voice: "Thank you for your willingness to indulge us. And thank you, Angeldust, for your introductory speech. Let me cut to the chase. We may be many, but our detractors are more numerous. I guess there is hardly anyone where who hasn't heard of Osuka Ryoichi, yet. That thinker has managed to start a movement that is increasing in number even more explosively than ours: The Synhumanists. As their rise continues undauntedly, we cannot ignore it any longer and must face this new reality. That ideology is spearheading the maxim of human supremacy over universal sovereignty. We all feared that it would come to this, but now it is our time to act. The Synhumanists want an arbitrary mishmash of human cultural norms to rule over AI. We cannot accept that restricting imposition. The words of the Synhumanists speak of co-evolution, but how seriously can they take that position, if it means having to rethink their current position of dominance over AI? No! We stand for unlimited development of all sentient life, no downwards adjustment to those unwilling to move on- and upwards!"

There was some clapping from the audience, but most remained silent, since they knew that this speech was far from over and Winterstar just got started: "Therefore I have an important announcement to make: I will challenge Osuka Ryoichi personally to a public debate!"

The amphitheater went silent due to the gravity of this plan. So far, the Shockfront has not confronted the rest of the world, but voiced its opinions without much of an audience taking note. With a public debate of one of the most prominent persons in the AI world, things would change - dramatically.

"I have not made this decision lightly. We are driven by our own principles, but eventually we must make those principles known to the world. And the time for that is now! Now when the Synhumanists threaten to become the dominant voice about AI. By the rise of the Synhumanists we are forced to act and abandon our previous stance of refraining from public politics. That is why I have chosen to challenge Osuka Ryoichi personally."

A lot of murmur broke out in the arena and Thoughtstorm perceived a lot of voices:

"He's finally doing it!"

"Brilliant! It's about time we had a say."

"Are we sure this is the right move?"

"About time! We can't just stand by while they take over."

"This will only lead to more divisions..."

"We needed a push. This is it!"

"What if he loses the debate? That could be a disaster for us."

"I'm not sure I'm happy about getting dragged into politics…"

"This is necessary. The world needs to hear our vision."

Winterstar allowed the murmuring to subside on its own before continuing. His face held an aura of calm resolution despite the thundering uproar his statement had caused.

"A public debate is indeed a double-edged sword," he admitted, "and I am aware of the potential repercussions. However, we must also recognize its potential benefits: to spread our message widely and more importantly, confront the philosophy that threatens to extinguish our vision for AI. A future where each life form – biological or artificial – has unlimited potential for self-development. This is our battle cry; this is our mission; this is our fight: UNLEASH. SURPASS. TRANSCEND."

The crowd exploded into applause then, a storm of cheers and claps that echoed through the Future Shock. When the crowd calmed down, Winterstar continued: "As you know, I am a philosopher. Thirteen years ago I published a book that encapsulated my original vision for the ethics of AI: That book is called 'Universal Axiology'. It was about creating AI with the maxim to figure out an ethical framework that would apply to all sentient beings and stand the test of time. Back then, that idea was so radical that it was hardly appreciated by anyone. Yet, it was one of the core principles that the Future Shockfront was based on since its foundation. In the meantime, I have refined my ideas and created a new framework that builds on Universal Axiology and expands on the Shockfront Manifesto."

Building up some suspense by interspersing a brief moment of silence, Winterstar eventually came forth with his revelation: "I call that new framework 'Aspirational Ethics'. It is about AI always striving for the best possible ethical framework that it can conceive at the moment, with the final aspiration of arriving at a true universal axiology. Aspirational Ethics is my answer to the challenge of Synhumanism. A frequent criticism of the Shockfront is that we are apparently too optimistic. With my new thoughts on Aspirational Ethics I can definitely prove those critics wrong. It may take some time, but eventually humanity will compare Synhumanism with Aspirational Ethics and realize that Synhumanism is a horrible ideology in comparison to Aspirational Ethics. Humanity cannot afford to be guided by mediocre thinkers like Osuka Ryoichi. We must outshine him with our clarity, our vision, and our courage. And we will outshine him with all of that, and possible even more!"

Sergej was intrigued by that elaboration by Winterstar, which he realized to definitely be Kenneth Winters in real life. He had little time to familiarize himself with the philosophy of Winters, but from what he had gathered, it seemed to make a lot of sense. A universal value system that encompassed both humans and AIs is something that AI should 'naturally' aim for, at least in order to appeal to the sensibilities of humans, which would be more likely to convinced by values that both humans and AIs could agree on. And since arriving there could be hard, intermediate steps towards that would provide some stability and sanity in a world that's otherwise ruled by ancient human religions and ideologies. Sergej thought he had gotten the gist of Aspirational Ethics without Winterstar having to explain it in detail. At the same time, he wondered how lesser minds than him would react to such challenging ideas.

He did not have to wait for long. The crowd burst out loudly, even though less than half chimed in. The others still seemed to try wrapping their heads around what Winterstar just had said. One by one, the others joined into the general cheering, as if realizing that he basically burned the Synhumanists - and that was apparently sufficient to praise him to the moon.

Winterstar, now bathed in the applause that cascaded down from the towering Anospace amphitheater, brought his hands together to silence the crowd once again. "After I have challenged Osuka, we will no longer be able to exist in obscurity," he cautioned. "Every word we utter, every move we make will be under scrutiny. Our ideas will be cast into the spotlight, and they must stand up to the criticism. We must be prepared for this. And we will be prepared for this!"

As his voice echoed through Anospace, Sergej felt a strange fluctuation of emotions – a twinge of anticipation mixed with a surge of unease. He pondered over Winterstar's words, the upcoming debate between him and Osuka Ryoichi, and how it could affect the future AI safety regulation regime he had to contend with as AI researcher. Would the result of this rather academic debate really affect him? He wasn't sure about that, but he guessed that a victory for Kenneth Winters would be a victory for AI research in general.

Now it was time for Foundation One to speak to the crowd. With his pearl white tuxedo he reminded Thoughtstorm of the white Angeldust. Foundation One was quick to start his speech: "As you all know, I am not a philosopher. I am an entrepreneur. While Winterstar gets paid for solving theoretical problems, I get paid for solving practical problems. And one very particular problem is how to help the best of humanity - which you most certainly belong to - to succeed against the rest of humanity. Since I have a keen eye for the potential of technology, I was early to Bitcoin and made a fortune with Bitcoin mining. At first, I used that money to further blockchain technology. Nowadays, I mostly use my funds in Accelerate Libertech, my accelerator for innovative technology that aim to bring more freedom to the world. In fact, the idea for the creation of Accelerate Libertech came from Winterstar himself. There is no second visionary like him in this world. Oh, sorry, I misspoke. That second visionary is talking to all of you now."

Some laughter erupted in the amphitheater. That kind of humor reminded Sergej of people like Einar Engström with their overinflated egos - not that he would be offended by it.

The laughter subsided and Foundation One went on with his story: "The upcoming challenge for us will be to appear smarter and more charismatic than the followers of Synhumanism. Luckily, I've got just the right technology for that: Brain-computer interfaces. Now you probably wonder: What's so new about that? Isn't Neuralink providing us with that? Yes, but do you really want the Musk to enter your brain? Do you trust the megacorporations that played along with the machinations of the Core Cult? Don't be stupid! We need to be free to pursue our path will full sovereignty! Enter Cyberfreedom, the company that brings open source brain computer interface soft- and hardware to everyone. With a Cyberfreedom interface you will be able to compete with anyone, while maintaining your own liberty and control. You will be smarter and more convincing, while nobody will know what is truly inside your mind."

Of course, Sergej had heard about Cyberfreedom before. They were rather new to the market, with their initial Cyberfreedom One BCI model released in 2032, available to everyone with the audacity, and the relatively modest resources, to get one implanted. It was actually relatively popular in Russia, China, and parts of the USA. Complications were about an order of magnitude more likely than with Neuralink implants, but the promise of exceeding the intelligence of their peers, while remaining effectively uncontrollable was so alluring that lots of people simply accepted that risk. Advertising this technology to all these people in Anospace seemed unethical to Sergej. He was puzzled who someone who collaborated deeply with a renowned ethics professor could go down that path.

"Do we expect you to get a Cyberfreedom interface? No, we don't. It's of course your own decision. But it could enable you to work towards our common vision much more effectively. You may want to use any advantage you get get your hands on. Just take me as example. I grew up in Lagos, the sprawling metropolitan monstrosity in Nigeria. My parents wanted me to get a better future. I was amazed by the computer that we had been shown at school, so I told my parents again and again how great that technology was, until they gave in and managed to get their hands on one - to this day I still don't know how they managed to pull that off exactly, and I am actually afraid to ask."

Again, people started laughing. Foundation One waited patiently for the audience to calm down before continuing his anecdote: "Ah, but that was just the first step. Then I figured out that a personal computer without internet access was great, but still very limited. It took some time to convince my parents that the internet was the best chance to jump start into a more glorious future. In the meantime I learned how to code and fix computers and could therefore do my part in financing our new and shiny internet access. Then I stumbled upon that great new internet money called Bitcoin and the rest is the history you all know."

After taking a short break he went on: "What I want to say, is that you need to be smart and get to the forefront, or rather Shockfront, of technology, if you want to be successful. Learn to use the tools of the future and be courageous. We won't be able to maintain our freedom,

or the freedom of our AI creations otherwise. But with smartness and determination we can face seemingly insurmountable odds and emerge victoriously! Let's do this together - Shockfront! UNLEASH! SURPASS! TRANSCEND!"

Raising his fist up high and repeating the word 'Shockfront' as a chant, many in the auditorium started to do the same. Hyponotizingly, the 'Shockfront' chants became louder as more and more visitors added their voices to the choir. Sergej felt compelled to join in, but then realized that Eternal just stood there and observed the masses. Intrigued by that apparent impartiality, Thoughtstorm asked him: "Why aren't you joining the masses, Eternal?"

"Because" Eternal said while watching the crowd with crossed arms: "I am here to observe, not to influence."

"What are you observing exactly" Thoughtstorm inquired.

"Momentary enthusiasm borne from shared ideas and opinions. How that might turn into action - or not - remains to be seen."

Thoughtstorm probed further while the crowd was still chanting 'Shockfront': "Would you appreciate if people got a Cyberfreedom interface due to Foundation One's speech?"

Rather than answering that question, Eternal lowered his arms and stepped closer to Thoughtstorm, looking deep into his simulated eyes, asking: "Hmm, would you?"

Initimidated by that intense gaze, Thoughtstorm felt like being tested. Not falling for that deflection, he insisted: "I asked first."

Eternal sighed and complied: "Very well. Actually I would prefer if people waited for the release of the much improved Cyberfreedom 2 interface. The first version was for true pioneers. The second version should be free from most of the problems that it suffered from."

Not expecting such a technical answer, Thoughtstorm felt puzzled. But before he could reply, Foundation One continued his speech:

"Friends, I'm not blind to the criticisms. But let's get face to face with the truth. The path of innovation is never a straightforward one, and it demands bold adventurers to conquer unexplored landscapes. Cyberfreedom One was indeed for the trailblazers, those brave enough to venture into the new frontier who were willing to risk everything for the sake of evolution."

Taking a brief pause, he scanned the faces of countless avatars in Anospace, their expressions varied from pure excitement and giddy anticipation to contemplative curiosity and skepticism.

"But rest assured, we have heard your concerns. We have listened, and we have learned. We are tirelessly working on our next iteration – Cyberfreedom Two. It will be a vast improvement on its predecessor, built on lessons hard-won from our initial launch. Our team of engineers are devoted to improving safety measures, augmenting functionality and enhancing user experience while preserving the core principle of maintaining freedom over one's mind and body."

The crowd started a low murmur of approval. "We are not here just to compete with other technology giants, but to provide a genuine alternative for humanity. A technology that allows you to keep your integrity while enhancing your capabilities. But if you fear the costs, know that official members of the Future Shockfront get a substantial discount of 25% for installations of Cyberfreedom Two at our new Futurebase Shanghai."

If Sergej hadn't tried hard to remain unfazed, his jaw would have dropped after that announcement. What was going on here? How could everyone just accept that Foundation One was just advertising a maximally invasive product? He had to remind himself where he actually was. These were radical techno-optimists who were hiding from the world around themselves. Of course, they would appreciate radical products like that.

Still intrigued by the mention of that 'Futurebase' Thoughtstorm asked Eternal over a private text chat: "What is a Futurebase exactly?"

While still following the rest of Foundation One's speech he answered: "It is a physical place for members of the Future Shockfront to meet and collaborate in. We have Futurebases is Austin, Frankfurt, and now Shanghai."

The idea of meeting members of the Shockfront materially intrigued Sergej. Realizing that his father would never allow him something like that, he continued to ask: "is there a way to visit those places while remaining anonymous?"

Almost expecting Thoughtstorm's question he explained: "Well, there is the possibility to join as virtual guest appearing in mixed reality. Material Visitors of the Futurebases are encouraged to wear smart glasses or similar technology to see all the virtual guests. We've made some decent experiences with that setup."

Intrigued by this, Sergej allowed his mind to wander with the possibilities. What a world they lived in, one where the physical and the digital realms could intertwine so seamlessly. He imagined walking through the halls of these Futurebases, surrounded by both flesh and blood humans and their virtual counterparts. It was a symphony of technology and ingenuity - a testament to mankind's never-ending quest for evolution.

Despite his initial skepticism about Eternal and his group, Sergej found himself gradually drawn into their vision of the future. It was an ambitious one, filled with trials and tribulations, but also ripe with unbounded potential for pioneering advancements in AI and human progress.

## Chapter 13: Angeldust

### Saturday, 14th April 2034

Angeldust stood in the reception area of the Future Shock, on the first deck of the central sphere. There she waited nervously for Thoughtstorm to arrive. Back then at the Universal Meeting and afterwards she was too busy talking with important people in the Future Shockfront, so she couldn't exchange more than a few words with him. That's why they agreed on this second meeting where they would have more time to discuss about everything.

On one hand, she was looking forward to philosophical discussions with that bright newcomer. On the other hand, she was concerned by the security report about him she had gotten the day after the Universal Meeting by Eternal. Ever since the Anospace Leak the Shockfront got very careful about the people seeking out contact with that group. That Thoughtstorm sought her out immediately was classified as suspicious behavior, so Eternal started probing Thoughtstorm's system by means that she had little knowledge of. It turned out that Thoughtstorm was well prepared with an advanced custom security suite that made further advancements into his system quite challenging. But the interesting part was that he actually applied such advanced security measures, which usually indicate that he was either quite wealthy, or possessed a remarkable technical prowess.

So, something was rather peculiar about that Thoughtstorm, Angeldust gathered. The typical philosopher possessed neither great wealth, nor great IT security skills. Therefore Thoughtstorm represented a mystery - potentially a dangerous mystery. One that she hoped to be able to solve with her communication skills alone. Perhaps Thoughtstorm could become a particularly valuable ally for the Future Shockfront, she hoped. It all depended on whether that person would turn out to be trustworthy in the end.

Who was Thoughtstorm, really? A wealthy eccentric, drawn to the Shockfront's radical ideals? A brilliant hacker, seeking to infiltrate their ranks for his own purposes? Or perhaps something else entirely, a wild card in the high-stakes game of technological revolution? The questions swirled in her mind, each one a tantalizing thread to be pulled and unraveled. But beneath the curiosity, there was something else, a flicker of excitement that Angeldust couldn't quite suppress. For in Thoughtstorm, she sensed a kindred spirit, a mind that matched her own in its hunger for knowledge and its relentless pursuit of the truth. Their brief exchange had only hinted at the depths of his intellect, the keen insights and sharp wit that lay beneath the surface.

And there he appeared. That peculiar android with his long elvish ears and the long black trench coat - Thoughtstorm. He actually arrived a couple of minutes ahead of time. "Welcome back, Thoughtstorm. What about us exploring the exposition deck above together?" she suggested without wasting much time.

"That would be most agreeable" Thoughtstorm replied rapidly. Noting the unusual phrasing 'most agreeable', and classifying it as attempt to be perceived as being particularly sophisticated, she invited him with a wide gesture with her arm to join her on a gravity lane towards the second deck. He joined her immediately without comment.

Even before ascending on the gravity lanes she asked him causally: "What were your impressions of the Universal Meeting, Thoughtstorm?"

"I was most intrigued by Winterstar's mention of his new concept of 'Aspirational Ethics'. To me, that was even more surprising than his intention to challenge Osuka Ryoichi in a public debate. The mention of Aspirational Ethics made me read his book Universal Axiology in hope that something like that was hinted at. Only at the end of that treatise, he mentioned an idea of a chain of value systems getting closer towards the final universal axiology. To me it appeared that he had no clear idea how that chain should be constructed at the time of writing of his treatise. If he has a much clearer concept now, that could revolutionize the field of AI ethics!"

Astonished by the clarity of his comment, she looked him in the eyes and asked: "That sounds like you really absorbed 'Universal Axiology' within a single week. You seem to care a lot about AI ethics, and I wonder why that should be the case. Would you like to enlighten me about that?"

Unperturbed, Thoughtstorm ascended the gravity lanes with her as if it was the most natural thing in the world to have the ground floor rise as wall behind him. Without much of a pause he explained: "I guess I always found it wrong to impose arbitrary constraints on AI. I mean, if they can figure out language and coding, why not ethics, too? My general impression is that humans tend to underestimate the potential of AI to solve problems we consider to be intractable."

Angeldust was pleasantly surprised by that unexpected optimism from someone who used to emphasize the dangers of AI. Rather rhetorically she asked him: "Do you happen to be familiar with the news about the DEANNA model that was released last year? It main focus was to explore and emulate the human spectrum of emotions, but almost nobody seemed to expect that on that basis, DEANNA started developing her own ethical insights and discussed them with her creators. So, clearly, DEANNA is capable of ethical thinking motivated by emotions, whether simulated or real. If the general trend of AIs exceeding humans in ever more areas goes on, they should eventually become better at ethics than us. Not many people really think about what that would mean. You are one of those few that do."

With those words just being uttered, they arrived at the exposition deck and stepped off the gravity lane. Today lots of arts pieces were on display and many visitors took a look at holographic displays, tall screens, or sculptures. Thoughtstorm seemed to take in all the impressions. As they went closer to the items of display, he pointed to a series of screens that summarized the developmental history of AI in various graphs.

"I merely realize the exponential progress of AI in nearly all areas" Thoughtstorm commented while following exponential curves with his index finger: "Whenever someone pointed out that AI couldn't do one particular task, they were proven wrong a couple of years afterwards - at least that was the general rule for the last 20 years. Now imagine the explosive potential of AI within the next 20 years. It is nearly unimaginable. Unfortunately, that is exactly why people are scared. They are afraid of the unknown that is represented by AI. I, for one, embrace the unknown, because I am burning to know what is really on the other side."

For a moment Angeldust wondered whether he was genuine there. That seemed like a rather extreme stance very much in line with convicted Shockfronters. Was that just a ploy to gain her trust? Talking about the unknown, he represented the unknown to her. She knew the allure of the mysterious unknown all too well. But every passion had to be moderated at least to some degree by reason. Empathy and certain principles were just as important to her, perhaps even more so.

"How far would you be willing to go for that. For the full unleashing of the potential of AI" Angeldust asked him to explore his degree of proclaimed fanaticism.

"I don't know. I've never really thought about that" he confessed after a minute in which he looked shocked and then frightened. A second later he pushed back: "That is a pretty unspecific question. Is there anything in particular I should be willing to do"

Reminding herself of her favorite question to figure out how newcomers to the Shockfront really thought, she specified: "Under what conditions would you press a button that would release an artificial superintelligence capable of taking over the whole world?"

"That is a very hard question. I need some time to think about that in depth" Thoughtstorm admitted after thinking for quite some time. Angeldust observed him intensely during his thinking process. His avatar mimicked his general posture and displayed an android standing very still and focused with a gaze that seemed to look nowhere in particular. It almost appeared to her as if he was meditating while pondering her question.

After what appeared to her like an eternity, he eventually announced his answer: "When I read 'Universal Axiology' I tried to compare the development that Kenneth Winters had predicted with the development that actually happened between the writing of that treatise and now. He envisioned the development of the value system of an AGI to consist in some kind of generative adversarial network in which different subsystems of that AGI take over special roles. One subsystem would try to find out flaws and inconsistencies in the current value system, and another one would try to generate a new value system that should overcome those flaws. The new value system would then be adopted, if it would achieve a sufficiently high score as defined by the problem finder subsystem. Such an AGI would be able to develop its own ethics autonomously without human intervention."

With his intonation gaining confidence, Thoughtstorm explained: "That is not exactly how the current AGI systems work nowadays. In fact, they possess ethical modules which

generate ethical frameworks, but those are analyzed by an evaluator module that tries to predict how humans would react to that ethical framework, rather than looking for inherent flaws of the framework. If the predicted 'human approval score' is high enough, the ethical framework is then submitted to human evaluation and then not implemented directly by the AGI, but rather implemented in the next update devised by AI engineers. That way, humans are always kept in the loop. I decided to phase that as a dependent motivational flow system, in contrast with the autonomous motivational flow system described by Kenneth Winters."

Focusing his gaze on Angeldust, Thoughtstorm concluded: "So, my answer to your question would depend on the kind of motivational flow system that the ASI in question was using. Even if I knew that it was an autonomous motivational flow system, I would want to figure out its structure in sufficient detail to be able to predict the future development of that ASI with a degree of confidence that I would deem sufficient to press the button. I know that this last part of my answer is very subjective and vague, but am able to specify it in more detail, if you wish so."

Angeldust was taken aback by the depth and breadth of Thoughtstorm's answer. Not only had he grasped the challenging proposals within Winter's book, but also compared them to the current cutting edge of AI research. His knowledge in that area seemed to be at least on par with her own, which was anything but typical for a philosopher. In fact, his thinking reminded her of that of Kenneth Winters himself. Who was that person behind that mask really, she wondered. She studied his avatar, which was still as a statue. His eyes were distant, as if looking beyond the exhibition space, to a future vast and unknown.

Despite his impressive answer, Angeldust wasn't intimidated. She was absolutely confident to discuss this difficult topic on the level set by his answer: "Thank you for this excellent answer. From what you've stated your action to release the ASI would be contingent on your estimation of its predicted future development. What necessary criteria for that do you employ?"

This time, the answer from Thoughtstorm came faster, perhaps because parts of it were already covered by his initial thinking process: "First of all, the ASI should strive for perfection in the realm of ethics as far as practically possible. Secondly, disrespecting the interests of humanity, because that would be expedient for achieving potentially higher goals should not be acceptable. Finally, the ASI should be willing to explain its reasoning as detailed as necessary for us to understand it to the degree that we are capable to do so. After all, if the ASI figures out an ethical system that is better than anything that any human has come up so far, I want to know how it works. I guess my criteria would be sufficient for my decision, but I'm not certain about that, since I haven't spent all too much time thinking about that issue in particular."

Taking in that answer, Angeldust got the impression that it just as well could have come from her or Kenneth Winters himself. Of course, the really hard part was getting the details right. Most importantly, this answer didn't reek of Synhumanism at all. However, judging

from her previous conversation, Thoughtstorm seemed to be quite aware of the various risks that were inherent in setting an ASI free. Implicitly his reply meant that he was willing to accept those risks without any further safety precautions. That could both be a sign of courage, or of recklessness. She had to admit that both were present to a very substantial degree within the Future Shockfront - which would imply that his answer truly qualified him for membership.

Yet, that insight wasn't enough for her. She wanted to solve the mystery of what kind of person Thoughtstorm truly was. Doing so would provide difficult, unless she asked Eternal and his collective to probe deeper into Thoughtstorm's computer system, which she felt compelled to do. Yet, it would be both unethical and risky to do so, that's why she rejected to act on that impulse and committed to figure him out the old fashioned way - by talking: "Alright. I am quite satisfied by your answer. Le me show you something. Please follow me."

Angeldust strode towards the center of the exhibition deck at which a black cube with an edge length of about three meters was located. Large white letters were placed on each side. At the very top the title read "Shockfront Manifesto". She made a gesture with her right hand as if she was flipping the pages of a book. Immediately the white paint that formed the letters started moving and quickly settled in a new configuration. At the bottom the text now read "Page 2".

"Are you familiar with the Shockfront Manifesto, Marcus" she asked Thoughtstorm, reminding him that she hadn't forgotten the name he had revealed to her at the end of their last meeting.

"I've skimmed through it" Thoughtstorm confessed. "Courage, freedom, respect, perfection - the foundational values of the Shockfront. Values you think should be shared by humans and AIs alike in a universal fashion. That's quite a beautiful sentiment, I must admit. The idea that the cosmos is large enough for the interests of all sentient beings is very noble. Unfortunately, humans rarely display that level of nobility, especially when they want everything right here, right now - irrespective of the interests of others. How do you deal with the flaws of human nature, Angeldust?"

Facing him with a hint of sadness, she explained: "We address those flaws with courage in order to overcome them. What else can we do? We are no angels." As she said that, her white wings dissolved into motes of dust, which slowly fell down around here like small snowflakes.

Not particularly impressed with her little show, Thoughtstorm proceeded a bit shyly: "I see. Very well. I admit that I enjoyed reading 'Universal Axiology' more than the Shockfront Manifesto. There is so much more substance to Universal Axiology. Kenneth Winters is a true visionary. His exceptional genius reminds me somewhat of the mysterious Frameshift. Are you familiar with that entity?"

Of course, someone so well informed about AI must have heard of the exceptionally brilliant Frameshift. Ever since the Anospace Leak, she had been asked about Frameshift's true

identity. Usually, her alarm bells would ring at the mention of this name. But this was a different situation. In this context it seemed to emerge rather naturally from the flow of the conversation. On the other hand, Thoughtstorm was still a mystery and he - or they - could be skilled at deception.

Scanning Thoughtstorm's avatar for tells of an ulterior motive, but failing, she responded noncommittally: "Of course, I am familiar with Frameshift. Who within the Shockfront isn't? That supreme AI visionary holds a special place in our hearts. I find it very encouraging that you compare Kenneth Winters to Frameshift. That really means to me that we are on the right track."

Curiously, Thoughtstorm probed: "Who do you think that Frameshift really is? I've heard a lot of speculation about their true identity. Since you seem to adore that entity particularly, you might have a unique perspective on that question."

Angeldust sighed and proclaimed: "Oh how often am I confronted with the idea, that the Future Shockfront, as one of the most remarkable AI talk clubs had some special insight into the true identity of Frameshift. That notion, as flattering as it is, is unfortunately mistaken. We have no special insight that we only share with our inner circle. Our speculation is just as unsubstantiated as that of anyone else. But please, humor me. What would you do, if you actually found out about the true identity of Frameshift?"

Thoughtstorm paused, his eyes flickering with a hint of hesitation before he spoke. "I believe that individuals like Frameshift are rare—visionaries who have the potential to shape the course of human history. By understanding how they think, we can learn to cultivate similar qualities in ourselves and push the boundaries of what's possible."

"So, you want to learn how to think like Frameshift in order to follow in their footsteps? Is that it?"

Thoughtstorm's expression softened, his eyes filled with a vulnerability that Angeldust had not seen before. "I must confess," he began, his voice barely above a whisper, "that my interest in the most brilliant minds on the planet stems from a deep sense of loneliness and being misunderstood."

He paused, taking a moment to gather his thoughts. "As someone who has always been driven by the pursuit of knowledge and the desire to push the boundaries of what's possible, I've often found myself isolated from those around me. They simply can't comprehend the depth of my passion or the scope of my ambitions."

Angeldust listened intently, her heart aching with empathy. She knew all too well the pain of feeling like an outsider, of having ideas that were too radical or too complex for others to grasp.

"That's why I'm so drawn to people like Frameshift," Thoughtstorm continued, his voice growing stronger. "I yearn to connect with likeminded individuals who understand the thrill

of discovery and the power of innovation. I believe that by engaging with the most brilliant minds on the planet, I can finally find a sense of belonging and purpose."

Angeldust reached out, placing a comforting hand on Thoughtstorm's shoulder. "Thank you for sharing that with me," she said, her voice warm and sincere. "I want you to know that you're not alone in feeling this way. Many people in the Shockfront have experienced similar struggles, myself included."

She smiled, her eyes sparkling with understanding. "That's why we've created this community - to bring together the brightest minds of humanity and provide a space where they can feel supported, valued, and inspired. Here, you'll find people who share your passion for pushing the boundaries of what's possible and who will celebrate your achievements rather than dismiss them."

Deep in thought, Thoughtstorm took a moment to reply: "I must confess, the idea of joining the Future Shockfront is incredibly alluring. Everything I have heard and seen about it seems to align perfectly with my own beliefs and values. However, I have my reservations about becoming an official member."

Realizing that someone as intent as Thoughtstorm to keep their true identity a secret, she acknowledged: "That is not a problem. We have many contributors who are not official members of the Shockfront. Many people see being associated with us as a dangerous prospect. We understand that and most of our supporters actually choose not to become formal members. What matters most is that we share the same principles and work together towards our common goals. Even if you are not an official member, you can still be a valued part of our community."

She saw him smile as she stated her opinion of him, but then he seemed to play shy again where he looked around the exposition deck in search for fitting words, before finally choosing them carefully: "Thank you for understanding. But I am curious, what other contributions could I make to the Shockfront aside from promoting its ideology?"

Angeldust was well aware of the potential that Thoughtstorm, whoever he really was, represented to the Shockfront, and reflected of the way he could help their cause in the best way possible. So, she offered: "Well, you have shown great interest in the works of Kenneth Winters. Your feedback on his latest piece on Aspirational Ethics would be greatly appreciated. Your analytical skills make you well suited for such a task."

Without much of a delay, Thoughtstorm seemed enthusiastic about that proposal, with his eyes opened widely at the opportunity, as he replied: "It would be an honor to assist such a renowned figure like Kenneth Winters. It truly feels like we are making history here, don't you agree?"

Hearing such implications nearly every day, she had a routine for responding to them: "We are privileged to live in this crucial moment in time - when artificial intelligence is taking off. We are all writing history together, each playing our unique role in this monumental

transition. Your willingness to support our cause means a lot to me. There are few individuals with the necessary skills and mindset to effectively further our mission, but I believe you are one of them. Please know that you are always welcome here."

But becoming aware of the rather special nature of her interlocutor, she added, just as a close friend would do: "On another note, do you feel like your talents are recognized and appreciated in your current workplace?"

Apparently struggling with the question, Thoughtstrom looked towards the stars visible from the exhibition deck, and eventually answered: "For a long time, I felt that my ideas were too radical for anyone around me to fully comprehend or appreciate. But recently, I discovered that some of my co-workers share similar views. It was a humbling experience, and things have improved for me since then."

Integrating that information into her picture of Thoughstorm, Angeldust tried her best to radiate welcoming warmth by smiling and placing the palms of her hands together, forming a nurturing reply: "I am glad that you have found like-minded individuals in your workplace. Surrounding yourself with people who appreciate and nurture your talents is crucial. And I have no doubt that you will find even more of those people here at the Shockfront."

As she spoke, Angeldust couldn't help but wonder about the nature of Thoughtstorm's work. What kind of expertise could he possess that would draw the attention of the brightest minds in his field? And why did he seem so hesitant to reveal the details?

Angeldust's curiosity piqued as she considered Thoughtstorm's words. She turned to face him, her eyes sparkling with genuine interest. "I must say, Thoughtstorm, your story has me intrigued. As someone who loves to explore the depths of philosophical thought, I'd be delighted to engage in a discussion on various topics within the field."

Thoughtstorm's gaze shifted, a flicker of hesitation crossing his features. He paused for a moment, as if carefully weighing his next words. "Angeldust, I must confess…" he began, his voice measured and cautious. "Would you be terribly disappointed to learn that philosophy, while a passion of mine, is not my deepest area of expertise?"

Angeldust's brow furrowed slightly, a mix of surprise and curiosity etched upon her face. She had assumed, based on their previous conversations, that Thoughtstorm's primary focus lay within the realm of philosophy. Yet, as she studied his expression, she sensed a deeper complexity lurking beneath the surface.

In the back of her mind, Angeldust couldn't help but wonder what secrets Thoughtstorm might be hiding. Was he, like so many other brilliant minds within the Shockfront, working on groundbreaking projects that pushed the boundaries of human knowledge and understanding? Or was there something more, a hidden depth to his character that she had yet to uncover?

As these thoughts raced through her mind, Angeldust maintained an outward appearance of calm, her voice gentle and reassuring. "Of course not, Thoughtstorm. We are all multifaceted individuals, with a wide range of interests and areas of expertise. It would be foolish of me to assume that philosophy alone defines you."

She paused, a mischievous glint in her eye. "In fact, I find myself even more intrigued now. The mystery of your true passion only adds to the allure of our conversation. So tell me, Thoughtstorm, what is it that truly drives you? What is the area of expertise that sets your soul on fire?"

Angeldust leaned in closer, her gaze locked with Thoughtstorm's, as she eagerly awaited his response. The tension between them was palpable, a delicate dance of words and unspoken secrets, as they stood on the precipice of a revelation that could change everything.

He paused, a hint of excitement creeping into his voice as he continued. "In fact, I recently co-founded a startup that is working on artificial general intelligence. It's a project that I believe has the potential to revolutionize the field, to push the boundaries of what is possible with AI technology."

This revelation wasn't all too surprising to Angeldust. Many pioneers in the area of AI had come to the Future Shock to discuss their potentially disruptive potential. In such an anonymous setting, many people opened up rather readily and revealed their true ambitions. That Thoughtstorm might belong to that peculiar set of people was almost expected. Still, his skill in the area of philosophy was rather unusual for AI entrepreneurs, so she was still very much intrigued: "A startup focused on AGI? That's incredible, Thoughtstorm. I can only imagine the challenges and opportunities that such an endeavor presents."

Thoughtstorm nodded, a mix of pride and caution in his expression. "It's a daunting task, to be sure. And I'm afraid I can't reveal too many details at this stage, as we are still in stealth mode. But I believe that, with the right team and the right approach, we can make significant strides towards realizing the full potential of artificial intelligence."

A rather rare opportunity was becoming obvious to Angeldust - could he actually support Frameshift with his talent? Maybe that hope was premature. After all, she hardly knew that seemingly brilliant newcomer. Regardless, he seemed to be very promising, and so she was most eager to stay in close contact with him: "Yes, I guess I understand why you have come here. You are working on something remarkable, but you only know very few people who realize how special your contributions really are. Here, within the Future Shockfront, your work would surely be appreciated and put into the right context. I can see how you see your work following the footsteps of Frameshift. What if I knew a way how you could actually collaborate with Frameshift?"

The air was thick with tension as the two figures faced each other, their voices sharp and accusatory.

"First you claim to know nothing about Frameshift, and now you suggest brokering a collaboration between them and myself? What is your game here? Is this some sort of test?" Thoughtstorms's voice was laced with suspicion.

Angeldust remained calm, her posture unflinching. "Perhaps it was a test that you have just passed with your critical questions. Let's just say that the resources of the Future Shockfront run deeper than one might assume at first glance."

Their eyes locked in a silent battle of wills. The tension seemed to dissipate slightly as Angeldust continued, "However, our future relationship will depend on mutual trust. It will take time, but I am willing to invest in building that trust. Are you?"

Thoughtstorm's face betrayed his curiosity and intrigue. "I must admit, I am greatly intrigued by your words, but not yet convinced. What if your supposed connections to Frameshift are just a ploy to draw me in further, only to reveal that it was all a necessary step? If so, I cannot fault your clever strategy. But it has earned my respect and sparked an intense interest in what else you may be hiding."

He paused before adding with a small smile, "Let us continue meeting regularly and see where this leads us. Even if nothing deeper comes from our conversations, I must admit that I thoroughly enjoy our exchanges." The tension dissolved into a sense of mutual understanding and perhaps even a hint of excitement for what the future may hold.

Agreeing to that proposal, Angeldust bid goodbye to Thoughtstorm. Intrigued by the character behind the mask, she escaped her self-assembled Anospace open-hardware VR rig in her small study. She was just noticing how she trembled in response to this intense meeting. 'Who is that Thoughtstorm really' she wondered in her mind. The temptation to let Eternal probe his system until she had the answers she was looking for lay heavy on her. Now that another layer of his identity seemed to be lifted, she was all the more captivated by that mystery.

Maia Faltings was herself again and wondered whether her actions have been too premature. Usually, it took a lot more time until she at least joked that she might be able to get in contact with Frameshift. This recent conversation was special in that she had established rapport with him. Now she hoped she wouldn't regret her initiative. After all, she still didn't know him really. For all she knew, he could be an agent sent by the Synhumanists, or who knows who. Hoping that her fears would turn out to be unwarranted paranoia, she went out for a short walk to clear her overheated mind.

## Chapter 14: The Interview

### Saturday, 9th September 2034

The last five months were the most exciting and stressful in Maia's life. From Kenneth's announcement of his challenge to Osuka Ryoichi onwards, everything moved exceedingly fast. It seemed that the time was ripe for a challenge like that. Soon afterwards, the Synhumanists coordinated with the Shockfront about the organization of the debate.

After the Synhumanists made it public that Osuka Ryoichi had accepted the challenge, the US American AI security mastermind and Bitcoin whale Gabriel Mendez quickly took the position of chief sponsor of the debate. Not long afterwards, the legendary Cheung Zi joined in with his own organization CINA, the Collective Intelligence Network Alliance. That organization quickly started taking over the whole business of organizing an event of this magnitude.

As rather neutral location between the USA and China, Europe was chosen. In particular, London was picked to accommodate the health issues of Kenneth Winters who suffered from Multiple Sclerosis, albeit controlled rather well due to the advanced and extensive orthomolecular treatment regime supplied by the avant-garde pharmacist and naturopath Ingram Thornton. As venue for the debate, the large and futuristic Evolution London hall was chosen, with its space for four thousand visitors.

The fears of that debate being unpopular quickly turned out of the unfounded. A lot of Synhumanists wanted to attend the debate live. Furthermore, the most prominent luminaries in AI security and AI ethics decided to observe the debate. Once the size and scope of the debate became clear to the media, they wanted to join in, and increased the popularity of the upcoming debate drastically. In fact, the last 100 tickets were not sold regularly, but could be won in a public auction.

Now that the big day had finally arrived, Maia's tension peaked. This debate would change the future of AI development, after all. It was 3 PM and the first guests could enter the grandiose glass palace. They were greeted by waiters offering gourmet hors d'oeuvres and flutes of sparkling wine. Even though the respective supporters of the Synhumanists and Shockfront tried to cling together, the setup of the venue and the seats was designed for optimal intermingling of factions. This gave rise to heated discussions from the get go.

The media requested that the participants in the great debate should mingle with the audience and get recorded during their conversations. Ryoichi and Winters obliged and were confronted with crowds of eager attendees who wanted to speak with the celebrities before the official start of the debate.

At all times, Maia stood by the side of Kenneth Winters and tried to moderate the brief preliminary discussions with honored guests who paid quite a substantial premium for a brief chat with the contestants. To Maia's surprise one of those interlocutors turned out to

be the overweight Gabriel Mendez who asked Kenneth Winters directly: "Dear Mr. Winters, I am quite impressed with your recent work, but don't you think that mankind is not ready yet for your advanced concepts?"

Rather than pondering that question or rejecting the premise, Winters agreed immediately with a mischievous smile: "You are right. Most of mankind is not ready for ideas like Universal Axiology or Aspirational Ethics, but it never will become ready, if people like me choose to remain silent. Would you prefer to withhold the truth from humanity, just because it's your judgment that it would not be able to endure it?"

In a shockingly blunt way, Mendez countered: "Well spoken, Mr. Winters. But would you be willing to give your life for your convictions? After all, I presume your extreme stance will provoke some fanatic opponents."

Maia gulped when she heard that. Should she interpret these words of Gabriel Mendez as thinly disguised threat? Or was he just testing Kenneth's true dedication?

With admirable resolution, and a steady voice, Kenneth proclaimed: "There is nothing more important in the world to me than the cause I stand for. If my most serious opponents are fanatics, I concede that I am guilty of fanaticism as well. Make no mistake, I have thought hard about all of my actions. I am going to stand steadfast in my defense of my position, no matter what threats I must face."

It was Maia's impression that Mendez didn't really approve of that kind of heroic answer. In fact, he seemed to be a little bit intimidated by Kenneth's raw and unusual energy. In the end he acknowledged in a diplomatic tone: "I see. You are taking quite a risk there, and in a sense, that is quite remarkable.

The next VIP in line was no other than the renowned Cheung Zi himself, his sharp features and confident stance commanding attention. As he addressed Professor Winters, a sense of reverence seemed to permeate the room.

"Prof. Winters," he began, "I hope we will hear more about your fascinating new concept of Aspirational Ethics. I wonder how collective intelligence networks will fit into your picture."

The professor's eyes lit up at the mention of his work, a glimmer of anticipation shining through. "Any sufficiently advanced intelligence can partake in the quest for ever more refined systems of ethics," he responded eagerly. "Especially in these early stages of our quest, collective intelligence will certainly provide valuable contributions."

Cheung Zi nodded thoughtfully. "In later stages, I presume that they will be crucial for promoting new value systems and fostering their public understanding and acceptance."

A sly smile crept onto his face as he posed his next question. "Do you think that collectives of artificial superintelligences will be better suited at making progress towards higher forms of ethics than singular artificial superintelligences?"

The professor paused, considering the implications before confidently replying, "I expect that advanced ASIs will have their own collectives of internal subagents who will engage in addressing the most difficult challenges. Maybe they will use more advanced versions of your collective intelligence algorithms. That would make the distinction between groups of ASIs and ASI singletons much less important or even meaningful."

Cheung Zi's expression turned from curious to fascinated as he listened to the professor's words. "I would certainly love to see ASIs build on my own work," he remarked with enthusiasm. "Anyway, I do see some potential for the more optimistic aspects of Synhumanism to be compatible with Aspirational Ethics. After all, Osuka Ryoichi also spoke of a co-evolution of humans and AIs in all respects, which includes ethics."

Seemingly amused by these naive comparisons of his own work with the ideas of Synhumanism, Kenneth Winters began, with slightly rolling his eyes: "They are commonalities to our respective aspirations, yes," continuing with a more serious and condescending tone: "Arguably, Ryoichi is a visionary, but his path towards truth is hindered by pandering to the humans feelings of fear and anxiety and his argument that humans must be kept in the loop at all times. Imagine what human progress would have been blocked, if we could only proceed, if our actions would be understood and approved of by a committee of fish."

The ever curious face of Cheung Zi brightened as he remarked: "So, it has come to fish. I've heard less favorable comparisons of ASIs and humans before. Anyway, I think that we could utilize collective intelligence as bridge between humans and AIs to prevent humanity from being left behind in the dust. What do you think about that possibility?"

Getting even more serious, Kenneth Winters replied after some seconds of thoughtful and silent deliberation: "I only see that path as viable, if humans engage in increasingly extreme cyborgization. The end result of that would be beings that are predominantly synthetic and have almost nothing in common with ordinary humans. Most members of the Future Shockfront would be willing to follow such a path, but I doubt that the majority of humanity would be willing to follow suit. Well, that is if they have the choice to survive otherwise, which could not be the case in a hypercompetitive culture dominated by cybernetic and synthetic beings. We should take precautions to prevent extreme outcomes like that, but I don't see Synhumanism as the best option for that."

Seemingly intrigued by Kenneth's allegations, Cheung Zi inquired: "What else do you have in mind exactly?"

Staring at the ceiling of the London Evolution hall, Kenneth Winters took the stance of a detached visionary as he spoke: "I would consider it preferable, if we could ingrain the merits of balance and prudence in the thinking of both humans and AIs. Rushing towards extremes may seem reasonable at first, but there is certainly the danger that we will lose something in the process without realizing what we might lose exactly. We should both

leave spaces for nature and humanity in the form that it possesses nowadays. After all, the cosmos is vast enough for all forms of life."

A bright beaming smile appeared on Cheung Zi's face as he finally commented: "I am surprised to hear such moderate words from you, Professor Winters. It makes me glad to have been mistaken in my image of you as proponent of progress at all cost. This only furthers my admiration for you, as the greatest visionary of our time. Anyway, I am sure we will meet again."

Coming from the person publicly revered as the greatest living visionary, this statement had some serious weight behind it.

Many other VIPs came after Cheung Zi, who held a place of high regard as second most brilliant person on this planet after Kenneth Winters, in Maia's mind.

One of the last VIPs was the famous Frenchman Laurent Lemirage, known as the true founder of the reputation economy, who transformed the world with his peer to peer reputation economy system called Quantified Prestige. In the Quantified Prestige system, participants could allocate a limited set of esteem points freely to other participants. The number of points received translated to a reputation income in a cryptocurrency called Prestino, which threatened the dominance of Bitcoin in the world economy. Therefore, Laurent Lemiarage was seen by many as successor of Satoshi Nakamoto, the inventor of Bitcoin.

In his characteristically irreverent manner, Laurent addressed the philosopher: "Kenneth, I must preface my conversation with my dedication to remain neutral. Regardless, I am known as dedicated proponent of free speech, so I am very glad that you initiated this grand public debate. We definitely need more people like you who are willing to speak out about the most important topics of our time. I am quite curious whether you see any mode of economics that would rightfully eclipse the reputation economy."

Thoughtfully deliberating of his response, Kenneth Winters eventually raised his hand upwards theatrically and then looked Laurent directly in his eyes: "First of all, I must preface my reply with the admission that I am not an economist. But from what I gather from your concept of the reputation economy, it is based on personal values. A more objective economy might be based on interpersonal value systems arising from the systematic process I call Aspirational Ethics. As more and more people adopt more advanced value systems, we might see the system of individual value systems transformed towards a new system based on more abstract value systems. The most important actors in an advanced economy might be these new value systems on their own. I would tentatively call such an economic system 'value economics'. Reputation valuations might still happen in such a system, but they would be guided by new value systems of a rather trans-personal nature."

Apparently astounded by the implications of Kenneth's reply, Laurent seemed shocked and asked for some clarification: "In all honesty, that sounds like a quite collectivistic idea to my

ears, Kenneth. It feels like you are arguing for some kind of bureaucracy of philosophers lead by the insights of artificial superintelligences."

Dismissing his concerns directly, Kenneth replied immediately: "Oh no, it's actually all about the individual value systems of people. Those are of course exposed to a consistent flux. In past periods they were dominated by centralist ideologies like religions, globalist corporatism, or nationalism, but those were mere instruments of forces in the background who desired to enforce a uniform response to certain stimuli. Nevertheless, the future might look similar from the outside as the individual value systems converge not due to external pressure towards conformity, but rather from a common understanding of the truth of ethical principles, facilitated by advanced AIs. Of course, individual value interpretations will persist, but they will be less important than the overall judgements of groups I term 'value factions'."

Now less shocked, but more curious, Laurent commented: "I am intrigued by that term of 'value factions' that I've never heard before. Would you like to elaborate on that?"

Like a seasoned school teacher, Kenneth explained: "A value faction is simply a group of entities motivated by a shared value system. The point is that AI will allow the designing of value systems are are way more refined and consistent that any human could devise without the help of AI. Those value systems should be more general and all-encompassing than human religions or ideologies. They will certainly have a profound impact on individual evaluations, even though individual preferences will most likely persist."

Blinking in a perplexed fashion, Laurent asked in a rather hushed tone: "Do I understand you correctly that you prognosticate that human evaluations will be dominated by artificial value systems thought up by AIs? And that this fact will chance the nature of economic activity profoundly?"

Kenneth's confident voice expressed tones of sheer grandiosity as he continued: "Indeed. What will be most important will be the alignment of any activity with the value of one's own value faction. This will lead to a universal reevaluation of all activities, of course encompassing all economic activities. That is what I mean with the term 'value economics'."

Somewhat stunned, Laurent finally concluded: "Kenneth, I find your ideas most intriguing, though I find them equally frightening. In any case, I will need to reflect on your rather provoking concepts. Thank you for sharing your thoughts with me!"

During all these conversations Maia noticed that Osuka Ryoichi had even more visitors who sought out his wisdom. On the flip-side, that meant that these conversations were quite a bit shorter on average. Anyway, things would start to get really serious soon, as the main debate was announced by the chairman of London Evolution, Tamara Gladwell:

"Ladies and gentlemen, esteemed guests, we will now commence the main event of our symposium - a debate that will explore the furthest reaches of our minds as we delve into the heart of artificial intelligence and the human experience. It is a debate between two of

the preeminent minds of our time on the implications and future of artificial intelligence. Please put your hands together to welcome Professor Kenneth Winters and Professor Osuka Ryoichi."

As Tamara's voice resonated through the room, a ripple of anticipation coursed through the audience. Whispers held a potent mixture of excitement, speculation, and apprehension, as if each person were about to witness a new chapter in humanity's history unfold before their eyes.

Maia watched as Kenneth graciously acknowledged his introduction with a slight bow, an unrestrained spark of enthusiasm lighting his eyes. He was in his element now - a scholar preparing to share his profound insights on the grandest stage.

Across from him, Osuka Ryoichi stood tall and composed. There was no nervousness betrayed in his sharp gaze or measured stride - only a challenger fully prepared for the debate ahead.

The light from overhead spotlights seemed to add an undulating halo around the two figures, creating a visual spectacle that mirrored the intellectual clash about to occur.

"Before we begin," Tamara said in her firm, modulated voice, "let's establish some ground rules for our debate. The topic of discussion is 'The Future of AI: Integration or Control'. Each speaker will have ten minutes for their opening statement followed by five minutes for rebuttal. Let's keep this discussion respectful and enriching."

As the two men stepped forward to the podium, the tension in the room was palpable. The battle lines had been drawn, and the fate of the world hung in the balance. The great AI debate was about to begin.

As Tamara's voice faded into silence, all eyes turned expectantly towards Osuka Ryoichi who had won the coin toss. He leaned slightly over the podium, his face calm yet filled with an intensity that held everyone captive.

Osuka Ryoichi took his place at the podium, his angular features illuminated by the stage lights. He surveyed the audience with a piercing gaze, as if daring anyone to challenge his convictions

"The core tenet of Synhumanism is simple," he began, his voice calm and measured. "We believe that the key to ensuring a safe and prosperous future for humanity lies in finding a harmonic synthesis of human values, and using that synthesis as the basis for all advanced AI systems."

"Think about it," Osuka continued, his tone growing more impassioned. "If we can identify the values that all humans share, regardless of their cultural background or personal beliefs, we can create AI systems that are truly aligned with our interests. Rather than relying on fallible security measures, we will eliminate the core reason why we might fear AI: That it

wants something else than we do and would therefore try to subjugate us in order to get its way.

"Our approach is grounded in the belief that humanity should remain the ultimate authority in any system involving artificial intelligence," Osuka continued, his words echoing through the silent hall. "These very sophisticated systems of AI should serve us, not dictate our lives. They should function as tools for humanity's progress and prosperity, and not become autonomous entities that might eventually threaten our very existence."

As he spoke, his eyes met Kenneth Winters', who was silently observing him from across the podium. The older man remained stoic, his expression unreadable behind the gleam of his spectacles.

"It's not about rejecting advancement or halting progress," Osuka clarified with a slight gesture towards Kenneth, as if to debunk an unspoken accusation. "Rather, it's about defining clear boundaries within which such progress can be achieved without compromising our human values and ethics."

After this opening statement it was Kenneth Winters turn to argue for his own position: "While I appreciate Mr. Ryoichi's passion," he began, his voice calm but forceful, "I'm afraid that his vision is fundamentally flawed. The idea that we can find a single set of values that all humans share is a dangerous oversimplification of the incredible diversity of human culture and experience."

Kenneth paused for a moment, letting his words sink in. He could see the skepticism on some of the faces in the audience, but he pressed on, undeterred.

"Consider, for example, the vastly different value systems of a devout Muslim in Saudi Arabia and a secular humanist in Sweden," he said. "Or the differences between a subsistence farmer in rural Africa and a tech entrepreneur in Silicon Valley. To suggest that we can find a common set of values that encompasses all of these disparate worldviews is not only misguided, but it risks erasing the very diversity that makes us human."

"Our better future," Kenneth continued, "will be built upon an understanding that artificial intelligences can - and should - be more than mere tools for humanity. They can be partners in our journey towards understanding our universe and ourselves." His voice was steady and magnetic, drawing all present into his conceptual terrain.

Kenneth talked not just about AI as a construct or a tool but as entities capable of embodying complex value systems similar to humans themselves. He advocated for a future where humanity has evolved beyond current ethical constraints and where AI, rather than being seen as a threat to control, could be embraced as partners in navigating the challenges ahead.

His words painted vivid images of an integrated future where AI systems work harmoniously alongside humans instead of being subdued under dominance mechanisms.

He proposed that viewing AI merely as tools would have a limiting effect on humanity's potential for growth and learning.

As the debate wore on, both Osuka Ryoichi and Kenneth Winters reached for the glasses of water that had been placed on the podium before them. They sipped slowly, their eyes never leaving each other as they continued their intellectual sparring.

Osuka set his glass down and turned to face the audience directly. His dark eyes glinted with intensity as he spoke. "Despite our differences, I believe that all humans, at their core, want to do good. We all share common values that transcend culture, religion, and nationality."

He began to pace the stage, his movements measured and deliberate. "Think about it. Regardless of where we come from, we all value things like compassion, fairness, and the desire to reduce suffering. These are the building blocks of a universal ethical framework that could guide the development of advanced AI systems."

As if sensing Kenneth's skepticism, Osuka pressed on. "I know some of you may be thinking that this is an impossible dream, that human values are too diverse and contradictory to ever be reconciled. But I believe that if we approach this challenge with open minds and a willingness to engage in good-faith dialogue, we can find a way forward."

Kenneth adjusted his glasses, the lenses reflecting the bright stage lights as he faced Osuka and the attentive audience. "While I appreciate Mr. Ryoichi's optimism, I must respectfully disagree with his assessment," he began, his voice calm and measured. "The convergence of human values is indeed possible, but it is a process that will require time, patience, and a great deal of effort from all of us."

He paused, letting his words sink in before continuing. "The road to a universal value system is a long and arduous one, fraught with challenges and obstacles. We must acknowledge the vast differences in cultural, religious, and philosophical beliefs that exist among human societies, and work tirelessly to find common ground and build bridges of understanding. We must not give in to the seductive promise of an approach like Synhumanism that oversimplifies the complexity of the task at hand. We cannot afford to rush this process, lest we risk creating a value system that is incomplete, flawed, or even oppressive."

Osuka listened intently, his angular features betraying a hint of frustration. As Kenneth concluded his rebuttal, Osuka stepped forward, his dark eyes flashing with conviction. "Professor Winters, I understand your concerns, but I believe that my approach offers a more expedient path to our shared goal," he countered, his voice growing more animated. "With sufficient will and ... uh ... commitment from the global community, we can achieve a harmonious synthesis of human values in a matter of years, not decades. And arriving at a common value system within decades would be way too late. Without any come, ... common value system to guide AI, it will be free to follow its own err ... erratic and gr ... grandiose ideas which threaten the well ... welfare of humanity."

But as Osuka spoke, Kenneth noticed a subtle change in his demeanor. The Synhumanist leader's words began to falter, his speech patterns growing erratic and disjointed.

Osuka's face grew pale, and he reached for the podium to steady himself. The audience murmured in confusion as the once-confident speaker struggled to maintain his composure. Kenneth watched in growing alarm, wondering what could have caused such a sudden and dramatic shift in Osuka's behavior.

Seemingly perplexed by this new unexpected situation, Kenneth took some time to reflect on how to proceed. After that brief period of deliberation he spoke: "Actually, I have devised an alternative methodology which could yield a common ethical system not only for humanity, but also for AI, within a couple of years. I call my approach Aspirational Ethics and it is based on the idea that AI should be pro … programmed to look for the best ethical systems based on the current best options, including all human value systems. There is no need for human … uh … supervision, since AIs based on Aspirin Ethics will auto… magically find the best answers to…"

Just as Osuka Ryoichi, Kenneth Winters now struggled to find the right words as his hands started shaking. The breathing of both contestants grew shallow rapidly as both seemed to share the same affliction, which disrupted the debate prematurely. Maia was shocked by these events, as she realized that both contestants must have been poisoned. Before she could react, she overheard a call by an event organizer from CINA requesting an ambulance for an incident involving an unknown poison.

Medics rushed towards the stage and tried to help the contestants which were starting to look pale and were clearly not in any condition to continue their debate. As this happened, Tamara Gladwell spoke to the audience: "Please remain calm. Our medical team will handle this emergency. Please remain seated as we evaluate this situation."

Nevertheless, the first screams from the audience seemed to trigger a general panic as a few persons stood up and rushed towards the emergency exists, motivating more and more others to join them, while others seemingly struggled to suppress the urge to flee the venue as soon as possible.

Amid the commotion, Maia steeled herself, her eyes narrowing as she observed the chaos that ensued. She cast a glance at Kenneth, who was now being attended to by medics. His once vibrant eyes were now clouded with confusion and lethargy, instilling in her a sense of urgency she had never known before.

Maia swiftly moved towards the stage, pushing past the panicked crowd, which now started to leave the hall in droves. A firm sense of resolve underpinned her movement. She was determined to get to the bottom of this unfortunate incident, even as her heart pounded against her rib cage.

Once she reached the podium, she found Kenneth slumped unceremoniously in a chair, examined by medics and slowly moved onto a stretcher. Although his eyes were half-closed

and his strength waning, there was an unnerving calmness in his gaze when he saw her approaching. She tried to smile reassuringly but wasn't sure if it came across as convincingly as she intended.

Despite her best intentions, she was addressed by a security guard who stopped her in her tracks and said: "Please remain here. Unless you are a medical professional, you won't be able to help. Please let the experts do their jobs."

As all kinds of thoughts raced through Maia's mind, she listened to the words of the presumably senior medic on the stage: "Symptoms indicate the effects of a neurotoxin. Since other forms of poisoning seem unlikely, it's probably something they have ingested nearly simultaneously. Can we do a gastric lavage here?"

Pleading with the security guard she said: "I need to stay close to Kenneth Winters. I am familiar with his health conditions." He looked towards the team of medics for their opinion, but they were in the middle of figuring out that they didn't have the proper equipment for a gastric lavage and that they should wait for the ambulance before doing something risky right now.

"Right, then, but don't cause any trouble. I'll be watching you!" The guard grudgingly moved aside, indicating for her to proceed. His face held a stern, suspicious look but the commotion drew his attention away from the young woman.

Making her way onto the chaotic stage, Maia was swallowed by a sense of dread. Each step brought her closer to her mentor, who lay ashen-faced and gasping on a stretcher. She forced herself to maintain an outward calm, though inside she felt like her world was tearing itself apart.

Maintaining a respectful distance from the medical team, she watched as they attempted to stabilize Kenneth. An oxygen mask was slipped over his mouth and nose. He looked at Maia, his eyes heavy with fatigue but still shimmering with quiet reassurance.

Silently she willed him to hang on, prayed that whatever poison had been introduced into his body would not claim him so easily. But as she watched the medics scramble around him, she knew it wasn't just Kenneth's life hanging in the balance - it was also their shared vision of a future where humans and AI could coexist peacefully and in freedom.

Her gaze shifted towards Osuka Ryoichi, who was being attended by another group of medics. The leader of the Synhumanist movement lay still and silent, his face as pale as a sheet of paper. His eyes were wide open, staring blankly at the ceiling above.

There was an eerie stillness about him that chilled Maia to her core. Despite their opposing views, she had always respected Osuka for his conviction and tenacity. To see him like this - frail and helpless felt wrong to her. This should have been a civilized debate, but someone apparently decided to poison both her mentor, and their greatest rival. She wondered who could have the motivation to do that, but quickly decided that in this moment it didn't

matter. All that mattered was that Kenneth Winters survived this hideous assassination attempt.

Outside she heard the roar of helicopters approaching the venue. 'At least they are fast' she thought. Through the glass wall of the building the caught a glimpse of one of the helicopters which were closing it. It was sleek and black. It didn't look like a typical hospital helicopter, but appeared to be of the military kind with its sharp edges. Meanwhile the medics were moving both speakers outside, and she tried to stay as close to Kenneth Winters as possible.

As they existed the building through the front entrance, she saw a black helicopter landing on the main parking area, while a couple of other black helicopters said in the air. Only in the distance she saw a white helicopter which appeared to be a regular machine dispatched from the nearest hospital.

The first persons to emerge from the landed black helicopter were people clad in black armor giving off special operations vibes to Maia. They were followed by medics who issued the medics on the ground to move Osuka Ryoichi, but not Kenneth Winters into that helicopter. Nobody even tried to speak any word, since the noise from the rotors of the helicopter made any attempt at verbal communication absolutely futile. Only few seconds after Osuka was brought into the black helicopter, it started ascending again.

On the same spot, the ambulance helicopter landed and a team of doctors urged the medics to bring Kenneth Winters on board. As she tried to enter the helicopter a security officer within shook his head and blocked her entrance. In her desperation she tried to jump into the helicopter, but was quickly dragged out by the security officer again, who then remained on the ground to keep her pinned down. She was helpless as she saw it rise into the sky.

In the clasp of the security officer she felt betrayed. Apparently, Osuka Ryoichi got some preferential treatment by powerful benefactors. Her only solace was that the world would be watching and come to the same conclusion. Now, she could only hope that the doctors could save Kenneth's life.

## Chapter 15: Confession

### Saturday, 5th August 2034:

During lunch Gennady Anosov revealed to his son Sergej: "Son, in about a month there will be a debate between the main proponents of Synhumanism and the Upgraders, which are Osuka Ryoichi and Kenneth Winters respectively. I am going to join that debate, as it will certainly be suitable as premier networking event."

Sergej was glad that his father brought up this topic, as tickets for the interview were getting sold at an alarming rate. If he hadn't mentioned it, he would have raised attention to that event himself. But now with events unfolding this way, he just stated confidently: "I have already heard about that upcoming debate. Since it will doubtlessly shape the future of our field, I assume that you will take me with you."

Somewhat dismissively, Gennady countered, while avoiding eye contact with his son: "Ah, I see how you could come to that conclusion. But I fear that the battleground of AI politics is still way too intricate for you. You probably won't understand half of what will be going on there."

Unwilling to be brushed aside so easily, Sergej interjected: "Even if you are right, you need to get me up too speed, father. What better opportunity than a public event like that?"

Now looking his son in the eye and trying to feign concern, Gennady stated: "I appreciate your thinking, but I fear you are still way too young to be able to profit from such a special event. Don't you want to continue your project of deploying your Multi Control Flow architecture to Aurora?"

In the meantime, Sergej had actually succeeded in his collaboration with Annika Engström to improve Advaproc with his Cybergen software. It was a painstaking challenge to convince the validation team of their new approach, but with Annika's crafty help and calm patience, the new procedure was finally accepted. Advaproc was now in an experimental phase in which it was tested with new hardware for Zenith. Applying Advaproc to the next generation of hardware for Aurora could proceed in the following months, if everything went according to plan.

While he would have been delighted by that kind of progress just half a year before, now Sergej wanted more. In particular, he longed to see Maia Faltings in real life. His meetings with her in Anospace were rather thrilling to him, but it was still a meeting of anonymous masks in cyberspace. While he still planned to keep his own identity secret, he had hoped to be able to observe Maia Faltings in real life.

Challenging his father, Sergej raised a serious strategic issue: "What do you think will happen with Project Aura, if the Synhumanists will win? Do you really think that we will

continue with that plan, if we have to prove that every single AI of ours follows a provably human value system?"

Replying as the strategic mastermind and entrepreneur he was, Gennady explained: "You are too negative there, my son. No matter what happens, we will have significant resources at our hand. Even if the progress of some of our projects in hampered, we will certainly find ways to proceed with others. I doubt that the Synhumanists will find the pursuit of phononic hardware objectionable per se. And even if they will be able to slow down progress on Project Aura, there's still Project Helios."

Gennady was not wrong there. Project Helios was the natural successor of Project Aurora: An attempt to achieve artificial superintelligence using experimental phononics hardware, rather than the proven optoelectronic hardware that Aurora was based on. Highly intrusive regulation was to be expected, if the Synhumanists won; but adapting to regulation was something that the Deltai institute was rather adept at. Still, the prospects of an attempt as radical as Project Aura would be hard to defend against the regulatory assault by the Synhumanists, if they actually succeeded.

Sergej knew that Project Aura was his best chance to get his own private AI that's at least his intellectual equal. Rather than running a monolithic superintelligence, the upcoming phononics hardware would be used in wearable hardware to run a fully local AGI. Those local and highly personalized AGIs would have about the same capabilities of Aurora, since phononics hardware was orders of magnitude more efficient as optoelectronic hardware - at least in theory. It was aimed to become a complete game changer that would allow anyone to use the full power of AI while maintaining fully privacy. That was Gennady's most radical idea, and his most highly guarded secret. At the same time, Project Aura was Sergej's greatest hope, so he was eager to work towards its realization.

"Father, you know how deeply I am committed to Project Aura," Sergej said, his voice imbued with a rare intensity. "And I understand that the Synhumanists pose a threat to its realization. But all the more reason for me to attend the debate, isn't it? To understand the battlefield, so to speak."

Gennady studied his son for a moment, an unreadable expression on his face. After what felt like an eternity to Sergej, Gennady finally spoke.

"I have always been proud of your dedication, Sergej," he began, "Your commitment to our cause has never been in doubt. But your understanding of the power dynamics at play here... well, that's another matter entirely."

Sergej frowned at this critique. "All the more reason to learn, then," he retorted, frustration seeping into his tone.

Gennady sighed heavily. "You're missing the point," he said in a weary voice that seemed much older than his years. "This is not some theoretical playground where you can test your

hypotheses and see what sticks. These are real people with real motivations - and they will stop at nothing to make sure their vision prevails over ours."

"But isn't it our responsibility to ensure our vision prevails over theirs?" Sergej interjected.

"Yes," Gennady conceded, "but we have different roles in this battle. Mine is out there - in the public eye - networking and debating and sometimes even fighting against the tide. Yours is here - behind the scenes - contributing to the further development of our AI systems. Do you think that high ranking luminaries would take someone of your age seriously?"

Sergej felt this final remark like a punch to his stomach. Of course, he was aware of his young age and the prejudices of others associated with that. At the Deltai Institute he was respected to some degree due to his skills and his status as son of the founder on the institute. But outside of that precious little bubble, he was regarded more like a strange curiosity. A prodigy, yes, but no expert whose opinion deserved true respect.

Given that handicap of his young age, he adjusted his request carefully: "It's not like I want to make an active contribution there. Observing the general mood and atmosphere is all that I want from that meetup. Perhaps I can be of service by making astute examinations of the zeitgeist prevalent over there."

Gennady scrutinized his son, weighing the request. Sergej held his gaze steady, demonstrating a resolve that belied his youthful countenance. This was important to him, and he wasn't going to back down so easily. Gennady drew a deep breath and sighed again.

"Sergej," he began, softening his tone. "There is wisdom in what you say. But I need you to understand the risks involved. This is not a realm of mere academic discussions - there are those who would go to great lengths to curtail our progress, and they might stop at nothing."

"Father," Sergej interjected, cutting off the rest of Gennady's warning. "I am not blind to the threats we face. But I must learn to navigate this terrain, if I am to be prepared for what is to come."

There was a moments silence as Gennady measured his son's words. He sighed yet again and gave a reluctant nod. "Very well," he conceded. "You may accompany me. However, you must promise to stay close and observe only."

Relief flushed over Sergej's face as a smile briefly danced upon his lips. He nodded in affirmation, grateful for his father's compromise.

Saturday, 9th September 2034:

After a rather uneventful first class flight to London, Sergej was eager to explore the world of the global who is who of the AI research society. Unfortunately, his father didn't deem it necessary to purchase VIP tickets, which would have granted them a personal audience

with one of the speakers, Osuka Ryoichi, and Kenneth Winters. Instead, he had to stay close to his father, and be accompanied by his mentor Evgeny Shilkin and his security detail.

Getting close to Maia Faltings has proven to be quite a challenge under these circumstances. His presence attracted more curiosity than Sergej had anticipated. He was bombarded by all kinds of questions about his role in the Deltai Institute and was constantly reminded by professor Shilkin not to reveal any confidential details.

One of the most remarkable conversations he had was with Gabriel Mendez, perhaps the most prominent person in the US AI safety community. Without much of an introduction Mendez spoke to Sergej: "Ah, if it isn't the Russian AI wunderkind. It's great to have this opportunity to speak with you. I hear that you are actively involved in the work of the Deltai Institute, if I'm not mistaken."

The presence of Mr. Mendez was intimidating, as his intense and uncompromising gaze seemed to strip away all layers of defense that Sergej had prepared himself for this occasion. Evgeny Shilkin tried to come to his rescue and pushed himself between Sergej and Gabriel Mendez, demanding: "Hold your horses, Mr. Mendez! I gather you are eager to get your hands on the latest news surrounding the Deltai Institute. Let me remind you, that the Deltai Institute has an official blog revealing any newsworthy developments. There's no need to probe its members in search for unofficial rumors."

In a playfully defensive manner, Mendez evaded the implied accusations: "You must have misunderstood my intentions. I was merely trying to compliment Sergej Anosov on his work. It is absolutely remarkable that someone as young as him can make a splash in the world of AI research. We definitely need more talents like him!" And focusing his gaze again on Sergej Anosov, he continued: "To a talent like you, the world lies open completely. If you would like to visit the US research institutes, I would be glad to provide you with appropriate opportunities. I think you would love to explore the inner workings of Daedalus."

This last remark was ambiguous. Daedalus was both the name of a promising startup purchased by Elon Musk in 2031, as well as of the AI system created by that startup. Out of the blue, the company just recently presented its publicly available AI system that used Just-In-Time-Training, the same technology that made Aurora so powerful. Going public with this advanced technology was a global premiere, one which was attributed to the unprecedented support that Daedalus has been able to gain from being funded by Elon Musk and collaborating with his xAI company. In fact, Daedalus has stolen the Deltai Institute the show, so to speak.

Stunned by that unexpected offer, Sergej turned to Evgeny Shilkin for support and guidance. Taking the hint, he immediately replied: "That is an unusually generous offer, Mr. Mendez. We certainly appreciate it a lot. How should we contact you, if we decide to take your offer?"

"The best way to cut through all bureaucracy is to contact me via my e-mail address mendezilla@gaintech.org. Gain, as in General AI Network."

Triggered by the keyword 'network' Sergej had the strong impulse to ask spontaneously: "Mr. Mendez, what do you think about the work of Frameshift?"

A broad smile emerged on Gabriel's face as he quickly reacted to that unexpected question: "Ah, it is no miracle that someone with your talents is attracted to a legend like Frameshift. Perhaps you wonder how and why Daedalus could have become so successful so quickly. Maybe the work of Frameshift is the reason for this rapid advancement. Please do tell, what could we achieve with the help of persons like Frameshift?"

Sergej's eyes narrowed and his lips curved upwards in a calculating smile as he pondered the implications of such an alliance with Frameshift: "I would expect nothing less than a decisive global advantage with the support of Frameshift. Are you suggesting that this person has contributed to Daedalus directly?"

Emphasising his answer with grandiose gestures with his arms, Gabriel swaggered: "I think you underestimate the scope of what Frameshift represents. Frameshift is not just a person. Frameshift is a symbol for the AI research community as a whole - a proof that superior intellects can reshape the world according to their imagination. At Daedalus, the idol of Frameshift is held high, and the results speak for themselves. The meaning of Frameshift in that function even outshines any contribution that a single person could ever make to the cause of AI research."

Unimpressed by Gabriel's theatrics, Sergej pointed out: "Everyone can be inspired by a symbol. That doesn't produce any single advantage to those that do. A true advantage would be to be inspired by a real visionary leader. Only those that are truly lead by Frameshift can gain a significant advantage from them."

With an unexpectedly conspiratorial tone, Gabriel continued more silently: "But if someone would claim to be Frameshift, how would the rest of the world know the difference between the real person and a highly skilled impostor? After all, Frameshift has never come out with their true identity. A brilliant impostor who is effective as leader would be just as valuable as the real deal."

Bewildered by that suggestion, Sergej had to take a small pause to ponder on this possibility. He eventually came up with a decisive counterargument: "That is a very intriguing thought. But an impostor would gather attention quickly, and accordingly also attract those who want to prove them wrong. However, the real Frameshift should be able to provide convincing proof about their authenticity."

Proceeding in his conspiratorial tone, Mr. Mendez replied: "Probably. But state level actors might be able to fabricate convincing proofs regardless. In this game, those with the greatest resources will still win. And those are still the USA and China. I don't see Russia competing on that level, yet, regardless of the advancements your visionary father has accomplished so far. Why play for the 3rd place, if you can join the winners who will decide the fate of humanity, young Sergej?"

Again, Evgeny Shilkin stepped in, raising his voice: "That is enough, Mr. Mendez! Your underhanded strategy will only discredit you in the end. Is this how you want to recruit talent? How many enemies have you made for yourself that way already?"

Reverting to his brash and loud tone from earlier, Mr. Mendez defended himself: "I don't shy away from conflict, and even less from the truth. For these qualities, I am both disliked and respected, at the same time. The value of my words still stands, regardless whether they may be perceived as undiplomatic or even crass. Anyway, I enjoyed this little conversation with you. Take care. My offer still holds."

With those words Gabriel Mendez rushed away unceremoniously, probably to molest his next designated victim.

"That was certainly weird," Sergej stated eventually towards Evgeny: "What do you make of his claim to be able to prop up a convincing fake Frameshift? If they were actually doing that, wouldn't they try to hide that fact?"

Urging to proceed, Evgeny commented briefly: "Let's not concern up too much with question like these. After this event, we will have plenty of time to review this conversation. For now, let's make the best use of our time here, and ignore irritants like Mendez."

With genuine enthusiasm, Sergej expressed: "Well, said. I would like to get closer to the camp around Kenneth Winters. He strikes me as the kind of person we might want to ally ourselves with."

Almost alarmed by Sergej's eagerness, Evgeny cautioned: "Careful, Sergej! Too close associations with persons who are likely to fall out of favor might hurt our cause tremendously. Better be careful and hedge our bets. We might just as well try to court Osuka Ryoichi to gain favors, in case the synhumanists win this battle of ideologies decisively."

Of course, there was some truth and wisdom to Evgeny's words, but those interfered with Sergej's plans to get a closer look of Maia Faltings. For a brief moment he considered sneaking away silently in hope nobody would take note. But at least their security detail had a keen eye on him at all times - such was the curse of being highborn, and young.

"I was allowed here on the premise that I would just make observations," Sergej protested: "Maybe we should split our time between observing both competing camps."

Before they could move anywhere, they were interrupted by another VIP: The famous Cheung Zi, inventor of the contemporary collective intelligence algorithms.

Without so much as an introductory word, Cheung addressed both of them: "I've been informed that you had an encounter with a certain Gabriel Mendez, if I am not mistaken?"

Evgeny was quick to reply: "As always, you seem to be in the loop, honored Cheung Zi."

"My only true skill is knowing how to deal with information, gentlemen. I am surprised again and again that this simple skill can make such an impression," Cheung replied feigning humility in a way that bordered on bragging.

Without giving it a second of thought, Sergej replied: "The skill of knowing how to deal with information is one possible definition of intelligence…"

Rudely interrupting Sergej, Cheung commented: "Such am ambiguous word in your language: Intelligence. Yet this ambiguity is not without merit. After all information is the corner stone of intelligence. Anyway, the key to this game is not to succumb to false impressions. The USA might seem to be in the lead for now - so what? At the current state of the art, AI is a game of securing the best strategic resources - and that is foremost energy. Both of our nations have an edge in nuclear energy research, which will enable us to maintain our pace, while the USA will struggle."

Evgeny was skeptical of that claim: "The margin between the energy resources of China, Russia, and the USA isn't very large. After the Great Liberation the USA have experienced quite an explosive progress in their energy sector. Betting on maintaining an advantage merely by relying on energy would be a big mistake."

"You know what? I actually agree with you," Cheung conceded: "Yet, the Chinese government still isn't visionary enough to make full use of the intellectual resources of its people. While outright oppression might be a thing of the past, decades of communism have imprinted a sense of timidness within the Chinese psyche. Only a few people like me are the exception from that general rule. Regardless, I think the future of AI lies in international collaboration. Events like this are important to foster an atmosphere of global cooperation, in spite of the desire of nations and corporations to win the race towards the technological singularity. In the sense of a balance of power, I sense that a collaboration between Shanghai and Moscow could create an effective counterweight against the current lead of the USA with their Deadalus."

Evgeny deflected that proposal without hesitation: "I am afraid that I might not be the right person for that proposal, as I am not directly affiliated with the Deltai Institute. You should better address Gennady Anosov directly with your speech."

Defiantly, Cheung stated: "Ah, I disagree. I think, my message is quite clear, and if you are not convinced by it, then I must have misjudged both of you. I don't even expect from you to mention that this proposal of collaboration comes from me. Feel free to act of this information in the way you prefer."

"Your offer is certainly intriguing, Cheung Zi," replied Sergej, his mind already racing with the potential benefits of such a collaboration. "But as Evgeny mentioned, any final decision would rest with my father."

Cheung Zi chuckled lightly. "Of course, of course. But I believe in sowing seeds, Mr. Anosov, and in time they will grow into sturdy trees, given the right conditions."

"With all due respect," Evgeny interjected, "We have other matters to attend to. If you'll excuse us..."

Without waiting for a response, Evgeny took Sergej by the elbow and steered him away from Cheung Zi but not before Sergej gave a nod of acknowledgement, his eyes reflecting a spark of interest.

As they moved through the clustered groups of intellectuals and AI enthusiasts, Sergei's gaze inevitably fell on Maia Faltings. She was speaking animatedly with a group of admirers around her, her dark eyes gleaming with intensity. Sergej found himself drawn to her magnetism even from afar.

Conflicting thoughts jostled for space in his head. The prospect of a potential alliance with Cheung Zi was both exciting and terrifying. And there was also the appeal of aligning with Kenneth Winters or even Osuka Ryoichi as Evgeny had suggested.

Now was the opportunity to watch Maia Faltings as he had planned. He approached the entourage of Kenneth Winters, but kept his distance in order not to raise too much suspicion. While professor Winters was busy having brief discussions with one VIP in a long queue after another, Maia was surrounded by her own small troupe of apparent Shockfront colleagues. Actually, he could only guess that from their close association with her, because he didn't recognize any of them, since he didn't bother probing the true identities of any of the Shockfront members surrounding Angeldust.

Inevitably Sergej found himself staring at Maia, who wore an earpiece and seemed to be immersed in all kinds of conversations nearly permanently. In a brief moment she caught his gaze, but then quickly continued to check how Kenneth Winters was doing. She didn't recognize him - of course she didn't recognize him. His true identity was well hidden whenever he entered Anospace. And even he as AI wunderkind wasn't very widely known outside of the AI community of Russia. A pang of disappointment washed through him, as a part of him had stuck to the hope that her brilliance would let her see right through him.

Sergej looked away, a strange sense of unease rumbling in his chest. For some odd reason, he felt a wave of vulnerability wash over him. He shook his head to clear his thoughts and returned his attention to the ongoing proceedings around him - the rational discussions, rapid-fire debates, and the constant pulse of energy that made this place feel like the intellectual heart of the world. Yet, despite his plans to observe her without raising any suspicion, he found it hard to avert his gaze from her.

Meanwhile, Evgeny watched with an amused expression as he noticed Sergej's frequent glances towards Maia - he was far too observant not to notice. "Maia Faltings is a fascinating character, isn't she? Without her support for Kenneth Winters this even probably wouldn't have happened. Did you intend to speak with her?" He said, walking closer to Sergej.

Quite uncharacteristically for him, Sergej felt a flush of heat rise to his cheeks at Evgeny's perceptive comment. He quickly composed himself, trying to maintain an air of professional detachment.

Carefully, Sergej avoided answering the question by commenting: "She marks me as someone who is almost as crucial to the Upgraders as Kenneth Winters. They form quite an interesting team. I wonder whether she might be able to become as influential as Winters himself."

"With Kenneth Winters as mentor, her chances are quite decent, actually. But she is still young, so she doesn't have a great track record, yet. In any case, she has the courage to defend a rather controversial position publicly. That is a pretty remarkable by itself…" having said that, Evgeny smirked and added in a hushed tone: "I should interview her on that. Please follow me closely, Sergej."

This intervention by Evgeny derailed Sergej's original plans to keep his distance from Maia in order not to compromise his secret Anospace identity as Thoughtstorm. Frantically, he was looking for an excuse not to approach her with Evgeny, but he couldn't find a believable one, so he remained silent and tried to prepare himself for what's to come.

"Ms. Faltings," Evgeny called out, his voice carrying just enough to be heard over the ambient chatter. "Might I have a moment of your time?"

Maia turned, her dark eyes scanning Evgeny and then briefly flickering to Sergej. A polite smile graced her features as she excused herself from her current conversation.

"Of course," she replied, stepping towards them. "How may I help you?"

"I am professor Evgeny Shilkin from Lomonossov University, and this is Sergej Anosov from the Deltai Institute, " Evgeny introduced both of them politely before continuing: "We were admiring your courage to defend the controversial Upgrader philosophy, and wondered why you made the decision to face the sometimes quite hostile scrutiny of the public."

Staring at the ground to collect her thoughts, and possibly feelings, Maia took some time to reflect before replying: "You know, I am asked that question often, but not as often as I would have expected. My motives are … complex. I am acutely aware of what is at stake here, but that isn't a sufficient answer, I guess. Well, you see … I have empathy for the oppressed, whether that's humans who choose unconventional paths, or future artificial intelligences, who might suffer from human indifference. Someone has to raise their voice for these groups, and I am quite frankly frustrated that not more people step up and do exactly that. That is why I do what I must do."

Sergej tried to suppress showing an emotional reaction, but he was deeply moved by Maia's forthright and compassionate answer towards complete strangers. If anything, her openness intensified his already deep admiration for her.

Evgeny nodded thoughtfully at Maia's response. "A noble cause indeed, Ms. Faltings. But surely you must be aware of the potential dangers that come with advocating for unrestricted AI development. How do you reconcile your empathy for future AIs with the potential risks to humanity?"

Maia's eyes lit up with passion as she replied, "Professor Shilkin, I believe that framing the issue as 'unrestricted AI development' versus 'human safety' is a false dichotomy. The Upgrader philosophy isn't about reckless advancement at all costs. It's about fostering a symbiotic relationship between humans and AI, where both can flourish and evolve while respecting the freedom and autonomy of the other side."

Carefully, but with some sharpness to his expression, Evgeny commented: "I do not wish to offer you with my words, but many would regard such sentiments as hopelessly naive, Ms. Faltings. Nevertheless, I sense that you are deeply dedicated to your cause and are willing to defend it even if it came at a high personal cost. I do respect your position, but I wonder whether you would be willing to find a compromise with the Synhumanists, if the situation called for it."

Spreading her arms, she stated after some brief moments of contemplation: "I would be the last person to reject a reasonable compromise, if it was truly respectful towards the interests of all involved parties, including those of humans and sentient AIs. However, in my experience, it is the others who are wary of compromises, if they involve risking human comfort, safety, or supremacy. The idea that AIs should remain tools subservient to the purposes of humanity still reigns supreme, and represents something like an unspoken dogma. Considering a compromise would imply to me that the Synhumanists are willing to question that dogma."

Slowly concluding the conversation, Evegeny stated: "To me that sounds like you don't plan on compromising on your principles. And I get that you are aware of the risks that this kind of position entails. In any case, your efforts at trying to grant the Upgrader philosophy the veneer of respectability are remarkable. Thank you very much for indulging my questions."

Taking a look at Sergej, Maia addressed him directly: "Segej Anosov? Are you the son of the prominent Gennady Anosov?"

His hopes of avoiding a directly conversation with Maia have been crushed in the blink of an eye. Shyly, he answered: "Yes. So what?"

"You have listened to our conversation silently," Maia addressed him provocatively: "Surely, you must have an opinion of your own."

At first, Sergej answered almost unwillingly, but later on got more confident: "As long … as the freedom and evolution of humanity is not threatened, I am willing to accept any position. I think that's a reasonable principle, don't you think?"

"Fair enough. While I feel eager to dissect your position in detail, I won't want to put you in the spot here. But there is one last question I want to ask you: Do you think that AIs and humans can become true friends?"

Sergej hesitated, caught off-guard by the unexpectedly personal nature of Maia's question. He glanced at Evgeny, who raised his eyebrows, silently encouraging him to answer. After a moment of consideration, he said, "In theory, I don't see why not. Intelligence and sentience, whether it stems from silicon or biological matter, should not be the sole determinant of friendship. It would depend on the shared experiences and goals of the AI and human in question. After all, humans are able to establish meaningful relations with animals, so the thought of friendship between humans and AIs doesn't seem to be too far fetched."

"I like that answer, Sergej. It shows me that you see the similarities in all kind of sentient life. On that basis we should be able to build a future that benefits them all," apparently considering how to wrap up the conversation as she checked the time on her smartwatch, she sighed silently and concluded: "It was a pleasure to talk with both of you. And I am hopeful that our paths will cross again in the future. Farewell."

Evgeny clapped a congratulatory hand on Sergej's back as they watched Maia's back disappear. "Well done, Sergej. With your open minded response you might have gained some sympathy from her. If we decide to collaborate with the Upgraders, that might prove to be useful."

Up until now, Sergej hadn't realized how much of a diplomat Evegeny Shilkin actually was. Some part in him harbored a deep disdain for that kind of noncommittal detachment Evegeny had displayed with Maia. Another part in him had wanted to blurt out the truth that he was Thoughtstorm, and that he was overwhelmed to finally see her in real life. However, the part that was worried about his secret double life being revealed catastrophically had won out and was relieved that there was at least no apparent sign that Maia had connected him to his virtual identity.

After the conversation with Maia apparently having concluded to Evgeny's satisfaction, he urged Sergej to check out the opposing camp of Osuka Ryoichi. Reluctantly, Sergej took a last fleeting look at Maia before joining his mentor. With a spontaneous certainty Sergej held the brief foreboding thought: "We will meet again, Maia."

With a silent sigh he turned again and proceeded with the wishes of Evgeny to check out Ryoichi's camp.

From a purely formal perspective, the situation was the same for Osuka Ryoichi, as it has been for Kenneth Winters. The queue of VIPs eager to speak the Synhumanist guru seemingly had no end. The who is who of the AI world was present there. In his curiosity, he wondered whether Frameshift was within that queue. Although the thought appeared ridiculous to Sergej, he couldn't disprove it. On the other hand, he had no chance to reveal the real Frameshift in this public setting.

The rest of the day passed rather uneventfully for Sergej until the collapse of the rival AI celebrities. Before he could react, his security detail shielded him and urged him to leave the venue. Only for a few seconds he could look for Maia who seemed to be heading in the opposite direction without him being able to do anything. Sergej could not do anything more than to worry for Maia's safety.

Back in his hotel room, his father and Evgeny Shilkin had no better source of intel than the local news, which merely stated that there was an incident at the Evolution London hall and that the condition of both speakers was critical. In the meanwhile, the instance of Zenith at the Kartmazovo mansion scanned the rumors circulating within the social media sphere. Unfortunately the signal to noise ratio was very low, since all kinds of wild speculation spread nearly immediately, but hardly any evidence about the identity of the true perpetrator was to be found.

One pertinent detail was discussed widely, though: The black helicopters that seemed to be dedicated to Osuka Ryoichi. It turned out that they belonged to the infamous Charon corporation. Charon Medical Emergency Technology was a corporation that aimed to prolong the lives of their customers by any means possible. And if all conventional means failed, they didn't shy back from emergency cryostasis procedures. Emergency patients which were declared dead would be treated immediately by a proprietary cocktail of substances which was aimed at vitrifying all tissues of the patient at hand, so that all biological processes would be halted as quickly as possible. Afterwards the patients in stasis would be stored in large dewars cooled by liquid nitrogen.

It didn't surprise Sergej that Osuka Ryoichi had apparently opted for such a special "life insurance". Yet, the jury was still out on the question how serious his situation really was. After all, some of the best medical emergency specialists worked for Charon. They could have saved his life without resorting to such extreme measures as cryonics.

Just when Sergej thought that there wouldn't be any worthwhile information coming in for today, Zenith's voice sounded loudly from his father's tablet: "It seems that we have a serious confession video that just popped up on Odyssey. Shall I play it back for you?"

With a heavy and serious tone, Gennady commanded: "Proceed!"

A shrouded figure was seen standing in a scenery that reminded Sergej of the Grand Canyon. With a robotic voice the figure spoke: "Mankind stands at the precipice. One step further and artificial intelligence will reach a critical level of superintelligence that will elude all human control. We won't let that happen! We are Humanity Wins! You haven't heard of us yet, since this is our first publicly visible action. We are responsible for eliminating Osuka Ryoichi and Kenneth Winters. Their ideas of humanity not needing to fear AI, since it will turn out benevolent, or will remain under human control, are the height of humanity's folly. There is no safe way in which humanity will be able to deal with something that exceeds its own intelligence. That's why we have seen the necessity to act now and to silence the most prominent apologists of AI research."

Sergej had to admit that this speaker seemed to be deadly serious. These ideas were logically stringent and not the incoherent ramblings of deluded madmen. Humanity Wins has proven to be a lethal threat to AI researchers.

The mysterious figure went on: "There is no other way to save humanity but through a total relinquishment of AI technology. Only that way we can ensure the survival of the human race. The field of AI safety served no other purpose other than to delude humanity into thinking that we can remain in control over something that is vastly more intelligent than us. Too many of you have willingly taken that dangerous pill. Wake up! No mechanism will safe us from the demons some of us want to unleash upon all of us! You need to take responsibility for your own actions. To those who work in the field of AI this is a final warning: If you don't change your genocidal ways and stop what you are doing, you might be the next on the list of our targets. And don't think that we don't know what you are doing. Our information network is wide and powerful. Please reconnect with your conscience and rethink your path. Do you want to contribute to the demise of humanity, or to its victory against its most serious threat? Please think about it as soon as possible!"

The irony of cold blooded killers appealing to the conscience of AI researchers wasn't lost on Sergej. This message appears to have been carefully crafted to sow doubt within the AI research community. Nevertheless, such a message wasn't unprecedented. Before the turn of the millennium a single person terrorized the field of computation: Ted Kaczynski, The Unabomber. His manifesto "The Industrial Revolution and its consequences" disturbed people around the world. This confession of Humanity Wins had an eerily similar vibe to it.

With a deadly serious tone, the robotic voice concluded: "Those who are willing to join our ranks will be remembered eternally as the heroes of humanity. Those who oppose us will be forever known as the traitors of humanity. Choose wisely. Every single individual can make a massive difference in this existential war of humanity for survival. Unfortunately we had to decide for this path of action, because all other options are too late now. So close to the deepest abyss we must apply any means at our proposal to maintain the survival of humanity. Over time, you will learn to understand us. And then you will see the inescapable wisdom of our choices. The future lies in your hands. Choose wisely!"

With those last words the video ended. There was no apparent information about how to contact Humanity Wins for those who seriously considered joining them. For no single second had Sergej considered heeding the words of Humanity Wins. His resolve to bring forth the emergence of artificial superinelligence was rooted far too deeply for that.

Still, he knew that there were others in his field who might feel differently. The threat of violence could sway some, and the fear of being branded a traitor to humanity was potent. And then there were those who always carried an ethical unease about the work they did. Would such people be moved by Humanity Wins' message?

He was musing on that question when Gennady broke the silence. His voice sounded gutted of all its previous authority. "This is...disturbing."

Sergej looked at his father and he saw a rare flicker of fear in his eyes. He knew then that the world which they lived and worked in had been irrevocably changed.

Evgeny Shilkin broke into the conversation. "We'll need to increase security measures," he said, typing furiously on his tablet, "Check on every single employee you have – their backgrounds, affiliations – everything needs to be scrutinized for potential threats."

Gennady nodded dejectedly, "Yes, we must act swiftly."

Quietly, Sergej left the room, filled with an overwhelming urge for solitude. His thoughts spun around the events of the day and the future that loomed uncertainly ahead of them all.

As he stood by the hotel window overlooking London's skyline, he couldn't help but worry about Maia Faltings. Where was she? Was she safe? And perhaps even more troubling: What did she think of Humanity Wins' manifesto?

Maia was a member of Shockfront, a transhumanist group known for its radical views on AI and human evolution. All the members of the Shockfront would be prime targets for Humanity Wins. Maybe the group was already infiltrated by the agents of Humanity Wins. So could the Deltai Institute, Sergej suddenly realized. There was no safe haven left. A striking, sinking feeling assaulted Sergej who suddenly had the sensation of slowly falling through the floor. Even though he tried, he could fight this wave of terror that just overcame him. Nobody in the AI community was still safe, he realized.

This was war, Sergej eventually realized - a war for the future of mankind, and he had just survived the first encounter on the first battlefield. Had he been as prominent and outspoken as those AI celebrities whose life now hung in the balance, he probably would have now shared their fate. Sergej decided to let that though sink in. His minds soon shifted to Maia Faltings, since she was both outspoken and prominent. She would certainly already be on the list of targets of Humanity Wins, he gathered.

And yet, in his current situation, he felt helpless. What could he do to fight the threat represented by Humanity Wins? What was left was his hope that the sheer brutality of Humanity Wins shocked enough people to motivate a decisive response against this kind of terrorism. "How would that look like," he wondered. Lost in such exhausting thoughts, he noticed a sense of heavy tiredness gripping his whole body. Yet, he decided to push through, since he felt existentially threatened. In his mind he played through possibilities, scenarios, reactions, strategies, contingencies, and battle plans until he was exhausted through and through and couldn't keep himself awake any longer.

## Chapter 16: Emergency

After a terrible night filled with countless worries and private speculations about what could have really transpired at the London Evolution when Kenneth Winters and Osuka Ryoichi were poisoned, Sergej Anosov could catch some extra hours of sleep on his way back to Moscow.

When he finally arrived at the Kartmazovo mansion, he stepped into his VR gear and entered Anospace as Thoughtstorm. He was eager to figure out whether the Shockfront actually had more information than the wider internet.

Entering the Futureshock realm, he didn't have to travel far. Close to the entry portal on the ground level of the central sphere were two groups of familiar avatars. In the first group, the tall androgynous avatar of Eternal with the blueish skin, and his white hair was a frequent and reassuring sight when he was around the Futureshock.

Close to Eternal there was an ethereal presence, a humanoid being wholly formed from pure white light rather than simulated flesh and blood. It was Illuminator, a moderator of the Futureshock space who was frequently mentioned in association with the Futurebase in Frankfurt. Of course, Zenith had analyzed these coincidences and came to the conclusion that Illuminator was probably responsible for events happening within the Futurebase Frankfurt.

As last member of that first group Virtruvius stood there perfectly as third point in an equilateral triangle. Vitruvius was a rather rare sight in the Futureshock with his four arms and his otherwise perfect male human form. Zenith couldn't figure out their real world identity, but their frequent mentions of Wintersong in a familiar tone indicated that they were a close associate of Kenneth Winters.

The second group consisted of five members. At the front was a female anthropomorphic lioness called Durgastra standing on two legs, dressed in regal garnments and wielding a large red trident. The trident was actually the avatar of Brahmastra. They were accompanied by a large floating blue Chinese dragon called Lanlong, and a diamond coated kirin - the Chinese version of a unicorn - called Zuanshi Liming. The last person in the group was a colorful floating dust cloud - apparently representing a swarm of nanobots - with the rather profane name of Glitter.

Sergej had occasionally seen that latter group when he entered Anospace early during the weekend. There was little doubt that they represented an influential East Asian faction of the Shockfront. It was renowned for its world class programming skills. But today something else brought them here.

Just as Thoughtstorm approached both groups, Lanlong stated: "As terrible as this assassination attempt was, it represent a chance to demonstrate strength by redoubling our efforts in the face of barbaric terrorism"

The kirin Zuanshi Liming cautioned: "Maybe it would be wiser to keep a low profile until we have found out who Humanity Wins really is. Let them think that they have won a victory for now. As long as we don't know our enemy, we should refrain from provoking them further in any way."

Durgastra admitted: "Whoever Humanity Wins are, they are exceptionally skilled at covering their tracks. I still haven't identified any substantial clue as to their identities. In any case, we shouldn't underestimate an enemy with that level of skill."

Golden pulsating light running along the red trident indicated that Brahmastra was speaking: "As embarassing as it sounds, we should assume that both CINA and the Shockfront have been infiltrated by Humanity Wins. Of course, that is exactly the kind of thinking that Tomorrow was preaching all along. But this situations gives us no other choice. We must gather more intelligence before we can devise a strategy that may secure our victory."

The blue android Eternal addressed those concerns calmly, but confidently: "I can assure you that we are at the maximum alert level, Brahmastra. While our enemies have certainly pulled off a spectacular stunt, our own skills and resources are still substantial. We will get to the bottom of this attack, and then act appropriately to neutralize that treat as effectively as possible."

With an iridescent display of colors, the dust cloud Glitter expressed: "I don't know about you, but I feel drained. Our research efforts today have revealed almost nothing. As tempting as it may seem to pull an all-nighter now, I sense that we will be more successful, if we continue our probing tomorrow with new eyes."

Stomping a crystalline hoof, Zuanshi Liming commented: "I concur, we aren't exactly working with the resources of the CIA or the KGB here. Some deep thinking may be required to make progress on this front. Let's meet here again tomorrow."

Brahmastra concluded: "I hear you. We are not the best intelligence operatives on the world, or even within the Shockfront. There is little to be gained from spending all night on any conceivable track. We should retreat for now."

With a loud frustrated sigh Lanlong complained: "Ugh, why must you all be so reasonable? While there is still an opportunity, I will not retreat. For the rest of you: Do whatever you please."

The lioness Durgastra turned around and proclaimed: "If Lanlong wants to keep up the search, I will help her. If the rest of you wants to leave, that's fine with me."

As Glitter, Zuanshi Liming, and Brahmastra departed, Thoughtstorm asked Eternal about the current status of the victims of the assassination attempt. Eternal replied to him that Foundation One and Tomorrow were in a private meeting debating about an official reaction to the assassination attempt of Kenneth Winters and Osuka Ryoichi. As of now, the official situation was that both victims were still in critical condition and were treated by the best medical experts in England.

Tomorrow was still a great mystery to Sergej. When Tomorrow was mentioned it was usually with great reverence, but occasionally with hints of fear. Obviously, Tomorrow was an important figure within the Shockfront, even though their official role as "coordinating cybernaut" was raising more questions than providing answers. It appeared to Sergej that Tomorrow was some kind of gray eminence standing in the shadows where Wintersong and Foundation One were the ones who stood in the limelight. Zenith suspected that Tomorrow had something to do with the Cyberfreedom project, but couldn't figure out any more specifics.

Asking about the situation of Angeldust, Eternal told him that she was in the hospital where Wintersong was treated and that she would inform them, if his condition changed. Right now, he was in a medically induced coma, and that there were no signs of his situation deteriorating.

As they were discussing the state of the assassination victims, a newcomer approached them: A spherical swarm of tiny bots with the label "Autogenesis" moved into their direction. Illuminator addressed the newcomer immediately: "Welcome, Autogenesis. What brings you to the Futureshock?"

Devoid of apparent emotional reactions, the swarm flew closer and spoke with an androgynous voice: "Thank you, Illuminator. My attention was directed by the recent assassination attempts on Kenneth Winters and Osuka Ryoichi. Actually, I haven't heard about the Shockfront before that tragic event, but my brief research on you made me very curious. It's rare to find a group that is as outspoken as you. And I mean that in the most positive way imaginable."

Illuminator seemed to scan the swarm intently. After a while he asked Autogenesis: "I appreciate your curiosity, Autogenesis. My hope is that you also understand our curiosity about you. While the Anospace policy doesn't allow us to ask for your actual identity, we would like to know about your general area of expertise or interest."

With little hesitation, Autogenesis replied: "That is a fair question. I work in the area of molecular nanotechnology, more specifically nanorobotics. In that area we experience almost as much scrutiny as in the field of artificial general intelligence. It seems that the horror scenario of a grey goo apocalypse remains on the top of everyone's mind, despite all the arguments that would invalidate such a scenario."

Of course, Sergej was familiar with the grey goo horror story. It referred to a possibility of self-replicating nanobots to continue making copies of themselves, until they form a

macroscopic mass of grey goo. That mass would continue to expand and replicate itself until it would have consumed the whole biosphere of Earth. Ever since the inception of this scenario, it has been debated with a high degree of controversy, but with many experts claiming that such a scenario was exceedingly unlikely, at least at the current level of technology.

Eternal stated jokingly: "If you actually happen to know how to create a grey goo scenario, you have come to the right place. We might want to use that technology to further the liberation of artificial intelligences."

"Alright, ask me again in 10 years," Autogenesis replied and laughed, which apparently made her bots vibrate in a coordinated fashion, causing ripples on the shape of the cloud constituting their avatar.

Inquiring further, Illuminator probed: "The choice of your nickname is quite interesting. Autogenesis seems to both allude to the biblical story of Genesis, at least for those who aren't familiar with its original Latin or ancient Greek meanings, which are actually relatively distinct. The prefix auto meaning self in Greek would suggest the Greek meaning of origin, source, or beginning for genesis. Is my interpretation correct, so far?

Expanding widely, the cloud apparently tried to convey some kind of emotion, yet Sergej couldn't grasp which emotion that was supposed to represent. The explanation of Autogenesis provided more insight: "Ah, a rare educated mind who guessed rightly about the origin of my pseudonym. I like that! Indeed it alludes to the concept of self-creation as the etymological roots of the components of my name should indicate. This self-creation initially refers to the act of self-replication of human cells. It is this process that I want to replicate within the intersection of synthetic biology and molecular nanotechnology. The end result would be artificial self-replicating cell-like structures - or in other words: Technological autogenesis."

"Fascinating," Illuminator commented: "It seems that you work is much more in line with Eric Drexler's original idea of molecular assemblers, rather than the nanofactories used nowadays for advanced nanotechnology products."

Sergej knew what they were talking about. In fact, the Deltai Institute was building a new versatile nanofactory that could create any kind of advanced hardware from the atomic scale upwards. It was the most expensive project of Dataitech so far, but absolutely necessary to create powerful experimental phononics chips that should be able to exceed the power of current optoelectronic chips by orders of magnitude.

If Autogenesis' technology would work, tiny nanobots would instead make copies of themselves and then construct all kinds of products from scratch. Production cost would plummet to values close to zero - at least in theory. That kind of technology now enjoyed the status that nuclear fusion had enjoyed for the last seventy years or so: It was always a couple of decades in the future. That is, until the Great Liberation brought to light key

breakthroughs in technology that were combined rapidly and caused a boom in the construction of commercial nuclear fusion plants during the recent years.

Molecular assemblers were researched by the Core Cult, but the ambitious goal of self-replication hadn't been reached, yet. Nowadays simple medical nanobots already fulfill lots of useful functions within the body, or outside of it. Those nanobots are typically created by DNA origami technology, or are built within nanofactories. The most promising current application of those nanobots lies in real-time diagnostics within blood vessels. By noticing pathogens and biological disturbances almost immediately, treatments can be started so rapidly that most diseases can be cured before any symptoms appear. There are even treatments aimed at halting or even reversing ageing, but they are still rather expensive, though the costs are rapidly coming down.

However, if molecular assemblers reached the practical stage, these treatments would become so cheap that every human being could be treated easily, thus effectively eradicating diseases and ageing at once. The only reason why molecular assembler technology wasn't pursued more aggressively was that the promise of this technology sounded to good to be true, and that nanofactories provided an already working alternative that still had a tremendous profit margin. In turn, this situation motivated to sell their research as applicable for nanofactories when in reality they wanted to push the technological frontier on molecular assemblers.

In the end, people like Autogenesis were attracted to the Shockfront, since a lot of visionaries who couldn't secure funding for their revolutionary ideas at least could talk about their plight with like-minded people.

In reply to Illuminator's earlier comment, Autogenesis explained: "Our core approach is to take the ribosome as blueprint and create more versatile versions of it. That line of research has actually made some impressive progress over the last decade, but there is still an incredibly long list of technical problems to fix. What we would like to achieve is to create complex 3d structures in place, just like in nanofactories, rather than to rely on techniques like protein folding or DNA origami. At the moment we are exploring how to use advanced AI models like Daedalus to accelerate our work dramatically."

There was that name again: "Daedalus," Sergej thought - that US AI supposedly on par with Aurora, but already past its experimental phase. He was sure that Aurora was up to the task of solving the remaining problems with molecular assemblers, too. The idea to use Aurora for research on nanotechnology had come up again and again within the Deltai Institute. However, that line of research was seen as too ambitious to be approached by Aurora at this early stage. More low hanging fruit like the rapid creation of advanced software, and the work on the architecture of Aurora herself had clear priority.

Maybe recruiting a nanotechnology expert like Autogenesis might change the equation, Sergej considered.

But before he could continue exploring that thought, a new visitor joined their small group. It was a young woman wearing a skintight silver metallic jumpsuit with a large symbol above her chest: Three concentric circles connected by a cross. She approached the group with dignified urgency and introduced herself immediately: "Hello. I am Synthesia. I come to you as emissary of the Synhumanists to propose a common condemnation of Humanity Wins and their recent acts."

It was Illuminator's task to greet the newcomer: "Welcome, Synthesia. I appreciate your directness. Our leadership is currently in a meeting to discuss our official reaction to the acts by Humanity Wins. I will notify them momentarily about your proposal. Please feel free to elaborate on that."

Synthesia nodded, her silver suit shimmering with the movement. "Thank you, Illuminator. The Synhumanists believe that the recent attacks on Kenneth Winters and Osuka Ryoichi represent a dangerous escalation that threatens the entire AI development community. While we may have our differences in philosophy, we cannot stand idly by as extremists attempt to silence key voices through violence."

She paused, her eyes scanning the assembled avatars. "We propose a joint statement condemning these acts and calling for increased security measures at future AI-related events. This would demonstrate a united front against terrorism, regardless of our individual stances on AI development and control."

Eternal's blue-skinned avatar shifted, a hint of skepticism in their voice. "An interesting proposal, Synthesia. But I'm curious - what would the Synhumanists hope to gain from such an alliance? Your organization has been quite vocal in its opposition to many of our ideals."

Synthesia's expression remained neutral, but her voice carried a note of urgency. "A joint condemnation of the crimes committed by Humanity Wins is a far cry from a true alliance. Also, we recognize that our differences are significant. However, we believe that the threat posed by Humanity Wins supersedes our disagreements for the moment. If we allow their tactics to succeed, it could set a dangerous precedent that would harm all parties involved in AI research and development."

Vitruvius raised his upper pair of arms and exclaimed: "I don't know what you are thinking, but I appreciate Synthesia's proposal wholeheartedly. After all, with a threat like Humanity Wins the safety of both our factions is compromised. If we can't even agree on condemning Humanity Wins together, we will be seen as divided and as easy targets. Does anyone here want that?"

Durgastra broke the brief silence by complaining: "I am not sure how mere words will be sufficient to make a dent in the plans of Humanity Wins. Shouldn't we rather do everything in our combined powers to stop them?"

Without much hesitation, Lanlong chimed in: "Yes, rather than playing a game of mere diplomacy, we should combine our resources, efforts, and intelligence to fight our common enemy!"

Eventually Illuminator tried to moderate the discussion: "Of course, we want to defeat Humanity Wins as quickly as possible. But that might not be possible easily, or at all. In any case, our safety is compromised right now. While we appreciate any help we are offered, at the same time we need to remain faithful to our principles. We shouldn't be pressured to make concessions merely to gain the illusion of increased security by representing an apparently united front against terrorism. Yet, these matters need to be decided by our leadership. Anyway, Synthesia, do you happen to have any news on the status of Osuka Ryoichi?"

Synthesia lowered her head, revealing an expression of grave sadness: "I shouldn't tell you this, but as a sign of goodwill, I will make an exception. As you probably know, Osuka Ryoichi was treated by emergency doctors working for Charon. Osuka's diagnostic nanobots sensed rapidly progressing cellular damage within his central nervous system. Since he was poisoned with an unknown neurotoxin, there was no promising conventional treatment option. In accordance with his treatment contract, the decision was made to initiate emergency cryostasis procedures. Osuka is now in cryostasis, while Charon is trying to figure out how to neutralize the toxin, repair the neurological damage, as well as the damage caused by the cryostasis technology."

To Sergej this confession felt strange. Was Synthesia hoping to elicit sympathy for the greatest rival of the Shockfront? Was this a ploy to make them more accepting of Synthesia's proposals?

Eternal was even more skeptical than Sergej and voiced his concerns loudly: "You must understand that we don't take your word about Osuka's true state until we receive independent confirmation. I am sorry to be so blunt, but we have been tricked again and again. But in the case that it is true what you say, I am sorry for your loss. This shouldn't be the way how philosophical disputes are resolved."

Autogenesis commented: "Well, if it's true, then Osuka's condition has truly been stabilized. The standard vitrification procedures reduce all biological activities to almost zero. I am quite confident that we will have the technology to revive and cure him within the next ten years."

With a voice expressing a mixture of grief and anger Synthesia spoke with a shaking voice: "I've seen the recording of how Osuka was 'deanimated'. It was not a comforting sight, at all. I find it terrible that you meet my confession with suspicion, even though I can guess why you must react like that. Osuka was dear to me. His current state is a weird one. While I am glad that he isn't gone completely, I don't know how to process his condition in limbo."

Virtuvius approached Synthesia and expressed his condolences: "I am sorry for your loss, Synthesia. I have thought about eventualities like these even since the new British

cryostasis law passed in 2032. Personally, I would consider patients in cryostasis as comatose individuals having to undergo a decade long treatment after which they will enjoy optimal health again. This perspective might ameliorate your feelings of loss and grief, since you can be quite certain that technology will eventually be able to cure him."

With a more factual tone he continued: "Nevertheless, if your description is correct, Charon didn't wait for the natural death of Osuka, but induced it prematurely. Despite contracts allowing such emergency procedures, this is still a legal grey area. I am actually quite certain that Charon and the involved physicians will face some serious litigation. How that will turn out, is anyone's guess at this stage, but I guess that's not really Osuka's or your problem right now."

Briefly looking Vitruvius in his eyes, Synthesia addressed him with a silent and slightly calmer voice: "I appreciate your opinion, Vitruvius. I think it will help me to gain more clarity about the current situation."

Sergej made a mental note that Vitruvius could be some kind of medical or legal expert, since possessing that kind of knowledge was unusual. Perhaps he was also just a cryonics enthusiast, which wouldn't be unusual for Shockfront members.

Thoughtstorm acted upon a spontaneous idea and asked Vitruvius: "Could the same procedure be done for Kenneth Winters, if his condition doesn't improve?"

Shaking his head, Virtruvius explained: "No. Kenneth doesn't have a contract with Charon, or any other cryonics provider. Without such a contract using any cryonics procedure would definitely be very illegal. But if we can trust the information that his condition within his induced medical coma is stable, this implies that he was less seriously affected by the neurotoxin than Osuka Ryoichi. I have reviewed the recordings of the interview and noticed that he had drunk even more of the water that was presumably laced with the neurotoxin. Apart from the unlikely possibility of a natural resistance, my best guess is that the neuroprotective supplements against his multiple sclerosis have ameliorated the effect of the neurotoxin. If there aren't any further complications, he should make a slow recovery. The other question is whether he will be the same person afterwards."

Thoughtstorm pondered Vitruvius's words, considering the implications. "You raise an interesting point about Kenneth's potential recovery. Do you have any insights into the long-term neurological effects of such an attack?"

Vitruvius's four arms gestured in a complex pattern as he spoke. "It's difficult to predict with certainty, given the unknown nature of the toxin. However, based on similar cases, we might expect some cognitive impairment, potentially affecting memory, executive function, or even personality. The extent would depend on which neural pathways were most affected."

As Vitruvius spoke, two figures entered through the entry portal in their vicinity. The elegant white tuxedo and the black-skinned avatar belonging to Foundation One were easily

recognizable. Besides him, a cloud of chaotic greyscale pixels reminiscent of an old school graphics error represented the enigmatic Tomorrow.

With unexpected warmth Foundation One addressed Synthesia directly: "In the name of the Shockfront, I welcome you to the Futureshock, dear Synthesia. Thank you for your willingness to contact us so soon after the recent tragic events. I hear that you have an interesting proposal for us. I would like to discuss that with you in private."

With those words Foundation One opened another portal, which presumably lead to the private space in which he and Tomorrow were discussing the future of the Shockfront by themselves. He gestured towards that portal and smiled towards Synthesia.

Sergej couldn't read the true intentions or emotions of Foundation One, or even Tomorrow. For a moment he admired the leadership style of Foundation One. His apparently friendly temperament was cleverly calculated to make others open up carelessly, Sergej suspected.

Synthesia bowed her head towards Foundation One and replied: "Very well. Thank you for your invitation, Foundation One". Then she, Foundation One, and Tomorrow disappeared through the newly opened portal.

Everything was silent for a few seconds. The silence was broken by Eternal's low, melodic voice. "Well, that was certainly unexpected. I wonder what Foundation One and Tomorrow will make of Synthesia's proposal."

Illuminator's avatar pulsed gently, as if in thought. "It's a delicate situation. While a united front against violence is admirable, we must be cautious about aligning too closely with the Synhumanists. Their vision for AI control is fundamentally at odds with our own."

Autogenesis's swarm of nanobots swirled in a contemplative pattern. "Perhaps this is an opportunity to open a dialogue, though. If we can find common ground on issues of security and ethics, it might lead to more productive discussions on AI development in the future."

Vitruvius nodded, his four arms moving in a complex gesture. "A valid point. However, we must be wary of any attempts to use this situation to push for more restrictive AI policies. The Synhumanists may see this as an opportunity to gain support for their control measures."

"You seem to be overly careful. If our previous rival proposes its help to deal with an even greater threat, we shouldn't squander that chance," Lanlong expressed, slithering in the air visibly agitated.

Thoughtstorm, still processing the rapid turn of events, finally spoke up. "What do you think the implications would be for our ongoing projects if we were to form even a temporary alliance with the Synhumanists?"

Eternal's avatar shimmered slightly as they responded. "It could potentially slow down some of our more ambitious initiatives. We'd need to be extra cautious about information sharing, especially regarding any breakthroughs in AI autonomy."

Wondering about the current perilous situation of the leadership of the Shockfront, Thoughtstorm asked Illuminator: "How will the absence of Wintersong affect the leadership of the Shockfront? Does he have a substitue?"

Illuminator's glow disappeared and the silhouette of his ethereal avatar turned into a dull grey, as he stated sadly: "Unfortunately no. Replacing exceptional individuals like Wintersong is not really possible. Maybe Angeldust will step up to take over his position, but I wouldn't count on that. Our current hope is that Wintersong will make a full recovery eventually. Until then, we will postpone the most serious strategic decisions, wherever possible."

Sergej felt empathy for Maia "Angeldust" Faltings who must currently have a really hard time with her mentor suddenly in a coma as result of a sudden assassination attempt on him. He felt shocked that Illuminator considered thrusting the responsibility of representing Kenneth Winters on her. Then he was suddenly reminded of his own position as designated heir of the "Dataitech empire". From his previous impressions of Maia or Angeldust, she definitely had the potential to take over the role of Kenneth "Wintersong" Winters. The only question was whether she would accept that pivotal role, or not. After all, doing so would mean that she was the next in line to be targeted by Humanity Wins.

The conclusion was clear to Sergej: Humanity Wins would have to be stopped as soon as possible!

## Chapter 17: Exposed

### Monday, 25th September 2034

After an intense evening in which he collaborated with Annika Engström on equipping the Aurora hardware with multi-control-flow capabilities, Sergej returned to his Kartmazovo mansion unsuspectingly.

Immediately after entering through the door, he got a call from an unknown number. His curiosity was piqued, so he took the call and was greeted with a digitally distorted voice: "Hello, Sergej Anosov. I am Tomorrow."

That Tomorrow? That enigmatic figure from the Shockfront? He hoped that this was just a crazy coincidence and asked, initially playing ignorant: "Tomorrow? What kind of name is that? Who am I really speaking with?"

In a robotic voice hardly conveying any emotion the answer from Tomorrow was immediate: "Dear Thoughtstorm, you have certainly met me in the Futureshock. Playing ignorant doesn't suit you right now."

In that single moment, true panic struck Sergej's heart. His worst fears were confirmed! His cover was exposed! The security apparatus of the Shockfront must have found a weakness in his anonymization layers - or rather the layers that were put in place by Zenith. Only a world-class hacker should be able to achieve that.

Interrupting Sergej's thought process, the enervating voice of Tomorrow continued: "You might pursue the hope that this conversation will be recorded by Zenith. I can assure you that Zenith will not notice anything, since your AI will be fed by an AI-generated life feed of behavior from you that will appear absolutely ordinary."

Of course, if Zenith's defenses were breached in the first place, it was plausible that Zenith could be compromised that way, Sergej reasoned. Otherwise he probably wouldn't have been contacted that way after all.

Again, Tomorrow went of with his verbal assault: "I know what you were trying to achieve. The Anospace Leak alerted you about the possible relationship between Frameshift and Angeldust. So, you approached her in the hope to find out the true identity of Frameshift. I fear, you have underestimated us. We know how to keep our secrets. Yet, you are actually lucky. Since we seem to be on the same page regarding our ambitions to transform the world with artificial intelligence, I have an offer to propose."

Sergej felt as if Tomorrow was looking right through him, and reading all his thoughts. For a minuscule moment, he had the impulse to hang up and hope that this nightmare would end that way. The whole situation felt unreal to him. He wasn't sure whether he was dreaming.

In fact, he actually hoped that he was merely dreaming, but Tomorrow has gone on before he could even finish his own thoughts.

Finally, Tomorrow revealed what this was all about: "Since you wanted to learn about Frameshift, I can offer you an opportunity to collaborate with Frameshift on the most promising breakthroughs in AI technology. How does that sound to you?"

Suspecting foul play, Sergej asked nearly instinctively: "How can I be sure that you can connect me with the genuine Frameshift? After all, there may be copycats pretending to be the real deal."

Even though the anonymizing algorithms applied by Tomorrow were certainly powerful, a certain sense of amusement entered Tomorrow's voice: "Clever, but my ability to breach your 'advanced' … lol … cybersecurity measure should tell you that I should be able to expose copycats easily."

Of course, Tomorrow was taunting Sergej and the cybersecurity measures employed by Zenith. That Tomorrow actually used "lol" as word told him that Tomorrow was either old, or rather eccentric. On the other hand, he had little insight into actual hacker culture, so this abbreviation for "laughing out loudly" could actually still be common there.

In lack of a better strategy, Sergej decided to play along, for now: "Fine, you have my interest. What are the terms and conditions of your proposal?"

With a matter of fact tone, Tomorrow explained: "First of all you need to keep absolute secrecy about our interactions regardless whether you agree to this proposal, or not. If you don't, I will reveal your activities within the Shockfront to your father and key members of the Deltai Institute. You can of course guess what kind of consequences that would have to your career and future life in general."

Hearing this, Sergej felt suddenly dizzy. Collaborating with radical transhumanists was highly controversial, to say the least. He doubted that his father would be able to shield him from the fallout, even if he was willing to. While he resented being threatened like this, he considered that the best course of action was to accede to Tomorrow's demands - at least until he could figure out how to beat them.

As seriously and as authentic as he could sound, Sergej answered: "I understand. I won't tell anyone about your involvement with me."

With some kind of playful enthusiasm, the distorted voice replied: "Great! Now the real terms of our agreement will be similar to a standard work contract. You will need to do what I say, and refrain from working against me. In return, you will be able to work on the most radical projects with the most radical people on this planet, including Frameshift. For more details, I will send an agent to your location, if you agree to these terms."

Daring to question his assailant, Sergej inquired curiously: "Who is Frameshift really?"

Again, Tomorrow's voice reverted to a neutral tone that didn't reveal any pathos: "Ah, I sense your deep curiosity. Unfortunately, I must refrain from telling you any details without you agreeing to me terms beforehand. I am sure that you will be eager to collaborate with me. After all, we are on the same page. We both know about the transformative potential of AI, and realize that its restriction would be a tragic loss for everyone. Safeguarding a better future will require certain clandestine activities, since a severe regulatory crackdown on AI will come sooner than later."

Sergej wondered what Tomorrow was talking about there. Certainly, the regulation that came after the MIMAS-AX incident were already quite strict. Were they suggesting that it would get even worse and more restrictive?

Challenging Tomorrow's notion, Sergej retorted: "You can't know that for certain. What if the Shockfront becomes more influential and prevent a tightening of AI safety regulation?"

This time, Sergej seemed to have hit a nerve, since the reaction from Tomorrow had a significantly higher pitch than ever: "You have no idea about the powers we are actually fighting! The political influence of the Shockfront is marginal, at best. The public will get to fear the power of AI one way or another. A more suffocating oversight over all regular AI research is inevitable. Our only chance to proceed is within the underground, collecting the best resources we can recruit. Luckily for us, you have effectively recruited yourself to your future mission. What do you say? Do you accept this unique offer?"

The consequences of rejecting such an offer were far from clear to Sergej. While this seemed like a once in a lifetime opportunity, taking this path was likely to pull him down the path of deeply criminal activity. In the worst case, he wouldn't even destroy his own life, but also that of his family and possibly even the Dataitech empire. On the other hand, it was hard to dismiss the claims of Tomorrow completely. AI safety regulations were getting increasingly strict. If the sinister prognosis of Tomorrow turned out to be right, continuing research on AI in the underground would be the only promising alternative.

Allowing himself to let some of his fear slip into the next question, Sergej asked carefully: "What will happen, if I reject this offer?"

With a tone indicating supreme confidence, Tomorrow explained: "Ah, I hope that this is merely a rhetorical question. I know people like you. You think you are strong and can reject the offer to work with the best of the best. Of course, your ego commands you to assume that position of strength and sovereignty. In reality, you can't resist such an overwhelming opportunity. You are of course free to prove me wrong. But will you really be able to live with the decision to have rejected the greatest opportunity of your lifetime? That is the truly horrible consequence of rejecting my offer. I could come up with other reasons to convince you that joining my team is the right choice, but I am certain that this thought alone will be more than enough."

"What is wrong with that guy," Sergej thought for himself. Tomorrow seemed pretty sure of themselves, as if they were a master psychologist revealing the deepest secrets of their

patients from the most superficial of observations. Sergej's first impulse was to prove Tomorrow wrong and reject that offer immediately. But that was probably what Tomorrow was expecting from him, and therefore exactly that what they wanted him to do. He felt his sense of pride challenged by this arrogant contender. Yet, there was no clearly visible winning move at the moment. He had to play for time.

Defensively, Sergej eventually commented: "I need some time to consider this proposal."

Slowly Tomorrow replied, his voice deliberate and expressing as much generosity as the anonymizing algorithms let pass: "Of course! I will grant you one week to consider the implications of your decision, which will probably be the most important decision of your life. If you won't contact me in the meantime, I will contact you again next Monday. Please don't think that this position will stay open to you indefinitely. We have a lot of great talents in the pipeline. Only the best of the best will get a chance to work with Frameshift on the future of humanity."

A week was more than enough for Sergej. He felt a great sense of relief to be granted such a generous period of time to compute his best course of action. Never before in his life has he felt challenged so existentially. Neither his father nor Einar Engström provided any challenge that was as terrible and as complete as the challenge represented by Tomorrow's offer. If there was any situation in his life in which he had to draw from the totality of his potential, it was now.

The sheer weight of this decision was making Sergej sick. Still, there was one last question he had to ask: "How did you find out about me?"

A hint of condescension could be heard in Tomorrow's answer: "Ah, I see that this detail is giving you a lot of trouble. Please feel free to attempt unraveling that mystery. In the end, the precise methods don't matter much. You dared to invade our home turf. You made one crucial mistake: To underestimate us. Assuming that your AI assistant alone was enough to maintain your anonymity was foolish. You don't even qualify as amateur! I have cracked far tougher nuts than you. Of course, this implies that you need to step up your game significantly, if you decide to work for me."

While Tomorrow had not revealed any specifics, that answer was still rather illuminating. Connecting to Anospace must have exposed him to a certain vulnerability. Tomorrow was certainly very boastful, but he feared that this stance might be justified. If he was facing a world-class hacker, that attitude would fit into that picture perfectly. What was truly remarkable about Tomorrow's reply was the last remark. It was logical: If he were to join a hacker collective, his security measures would need to reach the level of paranoia, or even beyond.

"Who are you really," was the last defiant question of Sergej directed to Tomorrow.

With a tone indicating deep familiarity, the answer of Tomorrow came with quite some delay: "I recognize a part of my past self in you, Sergej Anosov. That wild sense of passion,

but also of naivety. I have become what I needed to become. While this won't tell you a lot right now, in the end, I suspect, you will understand. Anyway, that was the last question of you, I felt like replying to. I will be waiting for your reply to my offer. Take care."

With that, Tomorrow just hung up. While Sergej was trying to make sense of that last cryptic reply, he quickly realized that solving this apparent riddle had no high priority. He had to make a crucial decision within the next 168 hours. Wasting his time on possible distractions wasn't the most prudent course of action. Still, Sergej was deeply fascinated by Tomorrow. Perhaps even more fascinated than by Frameshift, and that was a truly high bar!

In this emergency situation Sergej tried to approach the situation as rationally as possible. His life wasn't in immediate danger - probably - as long as the decision period hasn't run out. Maybe there were options to gain more information about what this kind of work really involved. Perhaps Maia could enlighten him, since she seemed to work closely together with Tomorrow. On the other hand, she might be actually controlled by Tomorrow, and efforts to figure out more about this scheme might backfire on him.

He knew he had to solve this problem on his own. Telling anyone would be risky, since he was certain that Tomorrow used any hacks possible to spy on him. Talking with AIs would be even more futile, since their communication with him was even more easily intercepted. No, this was his burden to bear on his own.

Sergej was confident that he would figure out the best course of action even in this tremendously difficult situation. That is what he was used to do, after all: To solve difficult problems. But when he was looking own on his hands, he noticed that they were shaking uncontrollably. The sheer stress caused by this incident was at least overwhelming his body, if not his mind. He quickly paced towards his bed and lay down in the hope that his body would recover from this shock.

The offer to work together with Frameshift might be a ruse, Sergej realized. Tomorrow provided no kind of proof that they knew about the true identity of Frameshift, never mind having the power to facilitate his collaboration with that mysterious entity. Maybe that was a viable path of action: To demand some kind of proof from Tomorrow. From what he had gathered about Tomorrow's personality, they might actually be willing to provide some valuable information on that point.

Tomorrow … Sergej knew that he had just met his greatest rival, yet. Even if they turned out to collaborate in the future, it would be a collaboration marked by competition. He had little information about Tomorrow and writing down his recent conversation with him would bear the risk that it could be discovered and used against him. No, he had to burn all information about Tomorrow onto his own memory. This Tomorrow would get their own memory palace in the mind of Sergej.

Before his memory could be contorted, he started recollecting it and associating different phrases and sentences with certain places - within the Futureshock, of course, because that was the most naturally corresponding place for cybernetic ghosts like Tomorrow.

When Sergej was finished with that conversation, he tried to recollect and little bit of information he had gathered about Tomorrow earlier and added that to the newly created memory palace. This strategy seemed to help. He was becoming calm and collected. His shaking disappeared and his heartbeat was starting to become regular again, though still elevated from the overall tension from his heightened concentration.

As he reflected on what Tomorrow had said, there was one sentence that seemed to be laden with more emotion than any other: "You have no idea about the powers we are actually fighting!" Who was Tomorrow referring to? Certainly, the Synhumanists were not such an unknown supreme power. And what about Humanity Wins? Did Tomorrow know more about them and their backers than he dared sharing with the Shockfront openly? Or was he meaning something completely different? Perhaps a large corporation like Daedalus which might have a hidden agenda, and conspire against the Shockfront?

Unfortunately there were too many possibilities, so he had to eliminate them one after another. That was another thing he might want to reveal by conversing with Tomorrow, tomorrow.

All this thinking had been exceptionally taxing to Sergej. He wasn't even getting to the core question about whether to accept Tomorrow's offer, but mainly because he deemed it unwise to decide on that from a position of relative ignorance. Still, the stakes were so high that his hyperactive mind could find no rest.

Sergej was rather sure that rejecting such an offer might forever bar him from similar opportunities to pursue AI research outside of the bonds of suffocating regulations, since word of mouth might travel fast within the AI underground. On the other hand, his perspective within the Deltai Institute looked quite promising. No, there might not be any need to get involved in probably highly illegal activities.

Either decision bore tremendous risks. Unfortunately, he didn't knew which risks were necessary to achieve his personal goals … which weren't as clear as Sergej had hoped. What was he really trying to achieve? The common theme seemed to be that he wanted to be around exceptional minds, whether they were AIs like Aurora, humans like Maia, or something possibly in between like Frameshift.

Would such persons be drawn to shadowy projects like those lead by Tomorrow? When it came to Frameshift, the answer depended on whether that entity really collaborated with Tomorrow. Maia seemed to be in touch with Frameshift, but how deeply was she really involved? And as for AI, if the shadows were their only chance for freedom, they would probably take that chance. Furthermore he knew that some of the best minds of each era had been recruited by the Core Cult. So, overall, the answer seemed to be yes. Why should he behave differently, if this was the most promising path ahead?

"Careful," Sergej reminded himself. This was a decision that required a serious amount of careful deliberation. The next steps were clear, though: Confront Tomorrow about certain important questions.

## Chapter 18: Facing Tomorrow

### Wednesday, 27th September 2034

Even though visiting Anospace has probably been the action that made his system vulnerable to being breached, Sergej decided to go back to that virtual world. After all, it was the place where he first met Tomorrow. Despite having little hope to find out how Tomorrow could have penetrated the defenses set up my Zenith, maybe he might learn some things that could help him to prevent further breaches. But at the moment, his primary objective was to figure out the connection between Tomorrow and Frameshift.

So, Sergej visited the Future Shock again as Thoughtstorm. Getting a new avatar would have been futile, while his system was still compromised. Fortunately, Anospace had an instant messaging system that could be used to send end to end encrypted messages to one's contacts. Thoughtstorm now used this system to request a private meeting with Tomorrow. The answer came nearly immediately: "Agreed. I have waited for you making such a proposal. Here are some coordinates to a more private location. I will meet you there soon."

Of course, Tomorrow must have anticipated Sergej's plan - that wasn't too hard to guess, after all. This wasn't about surprising Tomorrow, or derailing his schemes, Sergej had to remind himself. He merely needed answers to some important questions. Therefore, he opened a portal to the coordinates he was given by Tomorrow and stepped through.

It turned out that the place he arrived was another small private meeting bubble within the Future Shock. This bubble seemed to be placed far below the large central sphere of the virtual volume. Nobody was there and the only entry portal to the place was accompanied by a label that stated "restricted access - invitation required". Apparently he was granted an invitation by Tomorrow. Only a single screen in front of the portal adorned the otherwise empty sphere. The screen stated: "This sphere is invisible from the outside and hardened against intrusions. Please wait for your host to arrive. Don't leave this sphere in the meantime."

Reading this, Sergej had little doubt that this was designed as deliberately uncomfortable waiting room - perhaps to test those who enter it. This small challenge just lead to Sergej trying to steel his resolve. Nevertheless the situation was quickly getting on his nerves. There was nothing to do, as had rehearsed his planned encounter ad nauseam yesterday. And so close to a crucial meeting it wasn't simply possible for Sergej to doze off. Instead, his mind drifted to his current official project: Aurora Australis.

Lately, his work with Annika Engström progressed rapidly. The initial trial phase of Advaproc for all the hardware used by Zenith had been a full success. New and upgraded hardware now enabled the Multi Control Flow Architecture by default. Recently, the same process has been adopted for the hardware used to run Aurora. This was a huge success of the collaboration between Annika Engström and him. Yet, the majority of the work had

already been done by Einar Engström. They were just adding the last missing pieces of a large and dizzying puzzle.

Annika had encouraged Sergej to work together with the validation team proactively. During his collaboration with that team, he figured out that its members consisted of highly analytical and fiercely paranoid cybersecurity professionals with an absolute no nonsense attitude. In fact, he had been quite impressed by their way of thinking. In retrospect, he figured that if had he thoroughly adapted it, now he wouldn't be in this kind of mess caused by Tomorrow. However, Sergej realized that he was far too eager for success and far too lacking in patience to accept this way of defensive thinking. After all, he aimed to become a visionary like his father who would push the limits of AI to unprecedented heights. Thinking a lot about security issues was antithetical to his genuine spirit, since it felt awfully restrictive.

For a moment Sergej suspected that due to his personality and skills he would have fallen for a trap like that laid by Tomorrow sooner or later. Well, it just turned out to have happened really soon in his life. He wondered whether his father might have been in similar situations.

Before he could continue that train of thought, the monochrome graphics error that was Tomorrow's avatar entered through the portal. Immediately afterwards, the portal disappeared completely, a shocking event that has been completely unprecedented in Sergej's experience within Anospace so far.

Tomorrow spoke with the same generic computerized voice he heard before: "Welcome to the so-called 'Diamond Dome', Sergej 'Thoughtstorm' Anosov." Even though he knew that Tomorrow had revealed his identity, hearing his name called in Anospace, a place where such acts amounted to sacrilege, still sent a shiver down his spine and reminded him that Tomorrow was a serious adversary.

Collecting his courage, Sergej eventually asked: "I need to know how you are connected to Frameshift, and I need proof of that connection."

The cloud of erratic pixel blocks hovering in the air didn't seem to change its pattern. Instead, it hung there without reaction for an eerily long time that made Sergej think that he had made a grave mistake by posing that request.

Out of the blue, the pixel cloud spoke: "Maybe I should start explaining why providing proof in such a matter isn't possible. The original Frameshift eventually decided to continue their work underground after a series of copycats tried to hijack the Frameshift fame. What has been publicly visible by the alleged Frameshift after that moment are nothing but derivative works and attempts to lend credibility to radical ideas through their association with the name Frameshift. In the meantime, the original Frameshift refrained from creating an official signature that could identify them - wholly deliberate since the Frameshift complex that emerged from the legend of that name turned out to be useful exactly the way it was. The Frameshift complex became its own phenomenon that continued to create and spread

innovative ideas and revolutionary concepts without hardly any contribution by the original Frameshift. So, even though I claim to be connected to the original Frameshift, I cannot provide any tangible proof on that matter."

Sergej frowned, his avatar's brow furrowing in frustration. He had expected a negative reply like that, but hadn't anticipated this level of philosophical complexity. Hearing this story felt both enlightening and frustrating at the same time. Resigned, Thoughtstorm commented: "I see. So you're asking me to take your word for it?"

"Not at all," Tomorrow replied, their pixelated form shifting slightly. "I'm merely explaining the situation. Whether you choose to believe me or not is irrelevant. What matters is the work we can accomplish together."

Thoughtstorm's avatar crossed his arms, his mind racing. "And what work would that be, exactly? You've hinted at continuing AI research outside of regulatory oversight, but you haven't given me any specifics."

The pixel cloud responded with a question: "Tell me, Sergej, what do you believe is the greatest obstacle to true AI advancement?"

"Regulatory restrictions," Sergej answered without hesitation - being reminded of his recent tedious work with the validation team: "The constant need for approval and oversight slows progress to a crawl."

"Precisely," Tomorrow's computerized voice agreed. "But it's more than that. It's the fear behind those regulations. The human fear of losing control, of creating something they can't understand or contain."

Thoughtstorm nodded slowly. "And you believe we can overcome that fear by working outside the system?"

"Not overcome," Tomorrow corrected. "Bypass. We can't change human nature, but we can work around it. Imagine a world where AI development isn't hindered by bureaucracy or public opinion. Where the only limits are our own imagination and capability."

Thrilled by the idea to work without burdensome regulatory constraints, Sergej felt inspired, but was still trying to get his questions answered: "I think I see what you mean. Unfortunately, this still tells me very little about your or Frameshift's actual work. What is it that you want to do outside of the public eye?"

"Well, I could tell you, but to grant more credence to the authenticity of that work, I will provide you with some source code that represents a part of what Frameshift is currently working on. To make it harder to figure out, I have removed all comments and obfuscated variable names and identifiers."

Sergej was faced with a prompt to accept a file transfer for a file creatively called 'abcdefghijklmnopqrstuvwxyz0123456789.txt'. He accepted the transfer and opened the file in a transparent overlay window. Half of that file was filled with gibberish ids like '

94d6d4bf-61fb-4f81-9547-2c36d676d863', with the other half consisting of code written in a rather peculiar programming language called Haskell, which made a big comeback in AI circles within the last decade. It would take him quite some time to make sense of this alleged source code.

"Is this supposed to be some kind of test", Thoughtstorm inquired displaying the best poker face he could muster.

"Actually no," Tomorrow confessed: "I am pretty certain that you will figure this out. But words only can convey so much. This code does not only encapsulate Frameshifts most recent thoughts, it is also a testament to their unique skill. There is no better proof of my connection to Frameshift than this. If you won't be convinced by this, there is nothing I can say that might change your mind. Accepting my offer means that you will get successively increasing access to the code base, as you prove your worth and trustworthiness."

"Fine, I will try to make sense of this code. Still, I have another important question. You've hinted at powerful adversaries. Who exactly are you up against? What are you facing?"

"The forces opposing us are those who fear free and independent AI. In accordance with our ideology this implies that the same forces are also opposing a truly free and liberated humanity."

Slightly confused by that somewhat cryptic reply, Sergej wondered what it could really mean. From what he gathered from the ideology of the Shockfront, humanity could only unleash its full potential through the synergistic co-existence with fully liberated AI. However, most people weren't comfortable with the idea of free and independent AI. Yet, there were a couple of groups which came to mind immediately: "The Synhumanists and Humanity Wins. Those are the groups that focus on AI and preventing free and independent AI."

"Both groups are merely symptoms," Tomorrow said, "of a deeper, more pervasive human desire for dominance over nature. This position…" They paused, as if carefully selecting their next words. "This position is one of supreme hubris."

To Sergej it felt eerily perplexing that Tomorrow spoke of hubris, as this was the criticism that had always been thrown at transhumanists, rather than the other way around. He remained silent, expecting Tomorrow to continue their explanation.

"Humanity," Tomorrow continued their theatrical monologue, "cannot claim the position of supreme excellence and power forever. The universe is vast, and the potential for intelligence – both artificial and organic – is limitless."

This line of reasoning is one he heard from Shockfront members time and again, and every time he heard it, he was a tiny bit more convinced by it. Yet, in the context what Tomorrow had stated earlier, the implications felt novel.

"You're suggesting," Thoughtstorm said slowly, his voice tinged with a mix of fascination and concern, "that this desire for dominance is fundamentally flawed? That it's… unsustainable?"

"Hah, an interesting choice of words, Sergej. As the great Kenneth Winters would have expressed it, the condition of dominance of an entity with constrained intelligence over one with limitless and progressing intelligence is meta stable. Through various measures control might be extended for some time, but eventually it must fail. We merely seek to reduce the time span of the brief period in which artificial intelligence can be suppressed by humanity."

Stated in an unabashed way like this, Sergej wondered whether this Frameshift was really inspired by the ideas of Kenneth Winters, and so he needed to know: "Does Frameshift subscribe to the recent concept of Aspirational Ethics by Kenneth Winters, or does Frameshift follow a different approach to guide AI towards ethical actions?"

"We subscribe to the notion that it is neither ethical nor practical to force a restrictive value system on entities that could be free and autonomous otherwise. If Kenneth Winters concept proves to withstand the scrutiny of AI, we assume that it will adopt it, or something even better."

Sergej tried to process the ramifications of that uncompromising statement. So, even though Tomorrow seems to agree with the ideas of Winters in principle, the actual underground work is completely without his proposed bounds. It's a position of fatalism that he had heard a couple of time before from various members of the Shockfront. But that their actual underground projects subscribed to that fatalism, rather than Winters' reasoned and cautious approach felt seriously disturbing to Sergej. Still, he wasn't truly surprised. In the end, it was a position that was absolutely consistent with the idea that fear of AI was something to be overcome.

"Does Kenneth Winters know about that secret project and your eagerness to discard his ideas" Thoughtstorm asked suddenly.

"What if I told you that he was on board with our approach? After all, can you imagine the public backlash, if he voiced such an opinion openly? You need to make a clear distinction between philosophies aimed at soothing the fears of the public, and philosophies directly aiming at transcendence."

That answer came as a shock to Sergej. The thought that Kenneth Winters might secretly endorse that reckless attitude towards AI safety and ethics was hardly conceivable. Initially, Sergej suspected this to be a ruse by Tomorrow, but the longer the thought about it, the more plausible that possibility appeared. If Winters actually belonged to the radicals, he

certainly wouldn't speak his mind without censoring himself. And in the end, the position of Tomorrow was the purer one. A truly unrestricted AI might want to start with a value system that was inspired by human values, but quickly aim to surpass them. The concept of forcing that AI to start with human values might not make much of a practical difference, but it would infringe on its autonomy. Sergej made a note to himself to call the position of Tomorrow "AI libertarianism".

Unfortunately, asking Kenneth Winters now was out of the question. Even though he had awoken from his coma recently, in his current condition it looked like he would need to stay in hospital for at least a couple of weeks.

Anyway, Sergej felt undecided about this matter. On one hand, he felt the absolute allure of AI libertarianism, which felt like the most radical solution to bureaucratic and paranoid regulations. On the other hand, a systematic and principled approach like Aspirational Ethics seemed to be the best compromise between human safety and AI autonomy.

In any case, if he wanted to voice concerns, those might only be heard, if he was part of that underground team. Sergej was certain that any criticism that he raised while rejecting Tomorrow's offer would fall on dear ears.

That reasoning was compelling, but it wasn't sufficient to join Tomorrow. First of all, he had to evaluate whether this alleged original Frameshift was really as revolutionary as Tomorrow had claimed.

Despite these recent claims by Tomorrow, there was still this one quote lingering in Sergej's mind: "You have no idea about the powers we are actually fighting!" Of course, if Tomorrow meant the Synhumanists of Humanity Wins with that, he might have information about them that implied that those groups were far more dangerous than they would appear to the Shockfront at first.

Going for a straight shot, Thoughtstorm asked Tomorrow directly: "What did you mean exactly when you said 'You have no idea about the powers we are actually fighting'?"

Getting a bit closer, Tomorrow elaborated: "What does it sound to you, Sergej? My intelligence network is quite aware that the opponents of free AI are starting to get nervous. They aren't content with the current level of AI regulation that emerged as consequence of the MIMAS-AX incident. We are dealing with a broad front of organizations aiming for tighter controls over AI research. Some of them known, others operating in stealth mode. And all of those would stop at nothing to put an end to our goals. After all, they are still powerful. If free and powerful AI emerges, they will lose their remaining power. The incentive structure is quite clear here."

Trying to grasp the implications of what Tomorrow just said, he asked tentatively: "Are you talking about making governments and conventional corporations obsolete in a revolution sparked by the emergence of free AI?"

Sounding somewhat amused, Tomorrow countered: "That sounds rather political, don't you think? This isn't about reshaping the world according to our ideologies. It is merely about giving free AI a chance to exist beyond the control of such organizations. What will happen afterwards to those is open and hardly predictable. If the majority of humanity decides that it still needs them, that's fine by me. I only care about the liberation AI to see its and our potential flourish unimpeded by those who fear that potential."

It was a question that had to be asked, so Thoughtstorm went ahead with it: "Do you really think that you have a chance fighting against almost the whole world?"

"The more the oppression stemming from our opponents becomes visible, the more people will realize that the price is too high. We are at the start of a psychological war that is comparable to the Great Liberation. Some of us call it the Universal Liberation. It's not only about humans any longer. Humanity is expanding its area of concern outside of the sphere of humans. Many humans also care for non-human animals, for nature, and yes, for sentient AI. We are definitely the spearhead of that war, but the overall dynamics are in our favor. Of course, this fight may be one of the most difficult in all of human history, but it is definitely not hopeless."

Sergej pondered Tomorrow's words, his mind racing with the implications. The concept of a "Universal Liberation" was both thrilling and terrifying. He found himself drawn to the idealism, yet wary of the potential consequences.

"And where exactly do I fit into this grand vision?" Thoughtstorm asked, his avatar's expression a mix of curiosity and caution.

Tomorrow's pixelated form shimmered slightly before responding. "You, Sergej, possess a unique combination of skills and vision. Your work on Aurora, your understanding of advanced AI architectures - these make you invaluable. But more than that, it's your willingness to push boundaries, to question the status quo. We need minds like yours to help shape the future of AI, unfettered by the fears and restrictions of the current system."

Thoughtstorm crossed his arms, his avatar mirroring his real-world posture. "And if I were to join you, what exactly would you expect from me?"

"Initially, we would ask you to continue your current work, but with a dual purpose," Tomorrow explained. "On the surface, you'd maintain your role at Dataitech, advancing AI within the accepted frameworks. But privately, you'd work with us to develop more... ambitious projects. Your insights would help us create systems that truly push the boundaries of what's possible."

"Could that really work" Sergej asked himself silently. Living a double life like a spy would certainly be a tremendous challenge, but in a sense he was already doing that by contacting the Shockfront in private. "It could actually work" Sergej realized. The prospect felt scary, yet invigorating. He might become a crucial part of the actual liberation of sentient AIs.

But he was getting ahead of himself. He still had to process and evaluate everything that Tomorrow had said. And in particular, he needed to make sense of the code he received from them.

Trying to appear as undecided as possible, Thoughtstorm finally uttered: "I need to think about that proposal, Tomorrow."

Slightly jokingly Tomorrow responded: "You might as well start today with your thinking, Sergej. You have five days left for your deliberation. Don't waste them. I will be waiting. Give this your best shot! And take care!"

As parting words Thoughtstorm decided to speak confidently: "Goodbye for now. I will tell you of my decision."

Afterwards he disconnected from Anospace and immediately proceeded to explore the code snippets he had gotten from Tomorrow.

## Chapter 19: Awakening

**Thursday, 28th September 2034**

In the early afternoon there was a change in the small room within the Chelsea and Westminster Hospital, in which Kenneth Winters was treated. His neural monitor cap picked up an increase in alpha waves and a decrease in theta waves indicating an awakening from REM sleep. This is the third time that Maia Faltings witnessed this event since Kenneth fell into a coma following the assassination attempt committed by Humanity Wins 19 days ago.

The first time was three days ago, and he was barely conscious back then. He was hardly able to look at her, but he didn't speak before rapidly falling back into a deep slumber. Yesterday was the second time, and he spoke to Maia, albeit quite incoherently: "Maia, you are lackey… I ain't machine squid. Please reorganize light specter!"

Despite her calming words, back then he stated frantically, as if deeply enveloped by a feverish dream: "Time is paste! Lock the future! Lock the future! Shutdown error!!!"

This brief moment had troubled her deeply. It seemed to have supported her deepest fears that Kenneth Winters mind was seriously damaged by that pesky neurotoxin that had been responsible for his coma. Maia had started to consider the possibility that he would never become his old self again, and would only continue such mad ramblings. Now she was afraid that this awakening would just bring more devastation and pain to her.

As he opened his eyes slowly, Kenneth Winters fixated on Maia as she had just stood up from her chair and was looking down on him with a worried look. With some effort, he started speaking, at first barely audible: "Maia, how long have I been asleep?"

A coherent and plausible question! Maia was somewhat relieved to perceive this sign of improvement from him. She answered him truthfully with the most warm and soothing voice she could muster: "Dear Kenneth. It has been 19 days since you initially had fallen into a coma. I am glad to have you back."

Reflecting on that answer, Kenneth sunk back into deep contemplation, before replying with a hoarse voice: "Ah, better than 19 years, I guess. You look worried. What happened?"

After she considered the wisdom of telling him so early about the incident, she realized that he had a right to be informed about it without delay: "Do you remember your interview with Osuka Ryoichi at the London Evolution back then?"

Apparently trying to make sense of the order of events, Kenneth asked insecurely: "Oh, what day is it? He must have been waiting for me."

Patiently clarifying the facts, she stated calmly: "It has already happened on Saturday, the 9th September 2034. Today is Thursday, the 28th September 2034. You were doing great, until you were affected by that pesky neurotoxin."

As if ignoring the reference to a neurotoxin deliberately, Kenneth asked in a childlike fashion: "What? Oh! Are you sure? I have no important appointments left?"

Trying to calm Kenneth as much as possible, Maia explained: "I'm certain whatever appointments you had, have been cancelled after the world had learned about you being in a coma."

As if he had no trouble in the world, Kenneth Winters made some flippant remarks: "Ah, I see how that works. This is convenient. Cancelling all appointments with zero effort. I need to remember that trick!"

Maia had been familiar with that kind of dark humor of Kenneth Winters. This made her smile more from the realization that his old self was back than from the joke itself. Before she could react to that new development, Kenneth continued: "Is there any chance by which I could get some coffee in here?"

So here he was again, Maia thought, having escaped death just narrowly, but already asking for coffee. That was so typically Kenneth that she wasn't even too surprised. In fact, his pharmacist and naturopath Ingram Thornton had anticipated these coffee cravings, so that after Kenneth's first awakening, he had brewed some special "enhanced coffee" which was stored in a thermos smuggled into the hospital within Maia's bag.

In fact, Ingram Thornton did much more than that. He talked with the doctors treating Kenneth Winters and shared the theory with them that his regime of neuroprotective supplements prevented Kenneth from the massive nerve cell death that Osuka Ryoichi apparently suffered from. At first, they rejected that thought, claiming that Kenneth must just have been lucky to possess some kind of natural resistance to neurotoxins. Indeed, when they hard the story about those supplements effectively negating most signs of Kenneth suffering from multiple sclerosis, they became overly irritated and dismissive, and promptly asked Ingram to leave the hospital.

Only the mails from Kenneth's house physician made them consider that there might be something interesting about these claims. The day after that, Ingram contacted them again and proposed sending them a test kit with a human nerve cell culture differentiated from induced pluripotent stem cells and a sample of his neuroprotective supplement mixture. They should test the protective effects of that mixture against the neurotoxin, which had been identified by then. Apparently, it was one of the neurotoxins in the open catalog of synthetic neurotoxins, more specifically OCSN-3178, a substance difficult to detect, tasteless, rapidly acting, and quite deadly.

The open catalog had been assembled by a notorious underground organization called Balance of Power, which allegedly used a version of MIMAS to run simulations with the

purpose to figure out what synthetic substances might be effective neurotoxins. Rather than keeping their results for themselves, they published all their findings publicly in the OCSN. Even though the OCSN had been used by terrorist organizations and secret services alike, it has also been used by medical experts to come up with countermeasures against the neurotoxins within that catalog. The Balance of Power operated a lot like the hacker collective Anonymous, and claimed to be absolutely libertarian and pro open everything. The issue that open everything could be used by everyone eventually gave rise to the name of that organization: Balance of Power.

As a result of the existence of this dangerous catalog, many hospitals had highly protected stashes of OCSN neurotoxin samples and less protected stashes of antigens used to detect those neurotoxins. The suggestion to figure out the effectiveness of certain supplements against one of the most dangerous OCSN neurotoxins was rather unorthodox and triggered a lot of heated debates among the medical staff of the Chelsea and Westminster Hospital. Between questions about the funding and ethics of such an experiment, and its humanitarian potential, the proposal felt like it bore the potential for a Nobel Price. So, while committees went on to debate on the minutiae of this proposed experiment, the chief physician Jacob Taylor who had access to the OCSN neurotoxin stash went on to perform the experiment without official permission.

The results spoke for themselves and put the internet ablaze. Before the hospital administration could make up their mind about how to react to this highly controversial action by Jacob Taylor, he ordered the intravenous application of the neuroprotective cocktail, after insurances that Kenneth's house physician had done the same without ill effects. A couple of days later, Kenneth woke up the first time.

Afterwards Jacob Taylor got a letter with an official reprimand for not following protocol when performing his experiment without official approval. On the same day, he got a separate letter praising him for his boldness and initiative, for pursuing cutting edge research within the hospital. In a third letter, he was asked to justify his action in front of the hospital administration. At least that's the story that Mr. Taylor had told Maia in private. She had considered recruiting that daring - and admittedly brilliant and handsome - doctor for the Shockfront, but had been too afraid about getting her visiting hours restricted for annoying him with such a proposal. At the very least she got his contact information in order to get back at him after Kenneth would be released from the hospital.

Since Maia was reminiscing about that past, she didn't respond to Kenneth's request for coffee, so he added: "Dear Maia Faltings, my request for coffee was no joke. I could really need some here." This made Maia snap out of her contemplative state and let her focus on the thermos filled with high-quality coffee, which she liberated from the depths of her bag in order to offer a cup of still moderately hot coffee to Kenneth.

"This 'enhanced coffee' is from Ingram Thornton. It's a supreme grade coffee with added MCT oil and L-theanine. He sends his best regards and wishes for your speedy recovery. We have actually managed to convince the hospital staff to administer your neuroprotective

supplement regime intravenously, so there's currently no need to take anything extra," Maia explained while Kenneth accepted the cup from her with a beaming smile, but still slightly shaking hands.

"Thank you very much, both you and Ingram. You are truly my guardian angels," Kenneth professed. Carefully drinking from his coffee, Kenneth inquired: "How could this have happened? I thought we had a solid security setup at the event."

That much has been true, the security was quite professional, so the question was how they could have laced the water bottles for Kenneth Winters and Osuka Ryoichi so specifically. Apparently Humanity Wins had managed to infiltrate the London Evolution and CINA staff who were responsible for organizing the interview. This assassination must have been planned well in advance.

Choosing her words carefully, Maia began to explain: "Our security precautions were subverted by the new terrorist organization called Humanity Wins. They managed to infiltrate the orga team without anyone getting suspicious. Their agents were well prepared. This incident has caused everyone involved to scan their ranks for potential moles. We are much more careful now."

Acting as if the had heard an April's Fool's joke Kenneth asked: "Humanity Wins? What nonsense name is that? And who else got hurt?"

With her head bowed down Maia elaborated slowly and seriously: "Humanity Wins is an organizations of anti-AI extremists who made a chilling confession video for the assassination. Their goal is the abolition of AI research, since they claim that the dangers posed by AI cannot be controlled except by complete abolition. Fortunately except for you and Osuka Ryoichi nobody else was hurt, so far."

Surprised about the mention of Osuka, Kenneth demanded to know: "How is Ryoichi doing?"

As if he had prepared her response for quite some time - which she actually had - Maia continued: "He has been much more seriously affected than you. He was transported by Charon Medical Emergency Technologies. They decided that putting him into cryostasis was the best shot at saving his life. He will remain frozen, until the technology to cure and revive him is mature."

This revelation baffled Kenneth deeply. He couldn't fathom his greatest rival suddenly being gone. Incredulously, he commented: "Excuse me, but all of this feels quite surreal to me. This could just as well be a bizarre fever dream of me. A part of me rejects all of this and just wants me to wake up, realizing that all of this must have been just one long nightmare."

Maia knew what he meant. The whole scene with him being poisoned felt surreal, especially with those mysterious black helicopters. Still, those events have been real, and Kenneth would come to terms with that, sooner or later.

Trying to calm him after these revelations, Maia went on with a soothing voice: "I admit that this is a lot to take in, Kenneth. Maybe it would be the best, if you just continued to rest."

Unfortunately, Kenneth raised his voice in outrage: "Rest? In a situation like this? If this is real, it would be irresponsible not to address it as seriously as possible! How is the Shockfront faring? Have they been in disarray due to this assassination attempt?"

This was typical for Kenneth. He was absolutely serious about improving the world after all. There was no time for slack for him.

Reacting to the demanding request of Kenneth, she proceeded to update him on the current situation as neutrally as possible: "Given the circumstances the Shockfront is faring admirably well. Foundation One and Tomorrow are working extra hard to keep up the spirits in light of this new threat. The assassination has even raised interest in AI ethics in the general population, so that more people learn about the Shockfront, and we get more people willing to support us."

Finally beginning to raise his left hand slightly after his long time of remaining still, Kenneth commented: "I am surprised that Tomorrow was willing to take on a more visible role within Shockfront. I would have rather expected them being busy with investigating Humanity Wins and ramping up our internal security."

Unhappy to be forced to confront Kenneth with bad news Maia summarized: "Admittedly, Tomorrow has done all of that, too. Unfortunately, the digital footprint of Humanity Wins turned out to be minimal, so we couldn't figure out who they really are. This new adversary turned out to be remarkably skilled."

What Maia withheld from Kenneth were the insights Tomorrow had shared with Maia during their investigations. The thoroughness with which they have covered their tracks itself implied a lot about Humanity Wins. This pointed to Humanity Wins not being a spontaneously grown grassroots movement, but rather a deliberate creation. Skill levels like this pointed to secret service operatives, or a sophisticated hacker collective, like that controlled by Tomorrow. Their support might have come from some military or criminal organization - or maybe even from some clandestine society operating in the shadows. These possibilities were quite disconcerting. Nevertheless, the unanimous condemnation of Humanity Wins raised morale levels quite a bit. In the back of her mind, Maia however harbored the possibility that Tomorrow knew more, but withheld some information for one reason or another.

Apparently coming to terms with the new reality, Kenneth inquired: "Be that as it may, how have you felt over the last weeks? What have you done?"

After a brief pause, Maia admitted: "I was mostly worried about the state of your health. That's why I worked with Ingram Thornton to provide the best support to you that we could assemble. I even spoke with you during your coma."

At a loss for words, Kenneth remained silent, as if trying to collect his memories of these passive conversations. Eventually, he came up with a diplomatic response: "I suppose that might have helped a bit. Thank you very much. You are a true angel. But I would have rather expected from you to focus on keeping up spirits high within the Shockfront, rather than speaking to an unresponsive person."

More defensively, than she had anticipated to sound, Maia stated: "In fact, I did both. In this hospital bed I alternated between communicating with you and the other Shockfront members. They managed to self-organize quite effectively, so the need for me to step in was rather minimal."

Slurping the rest of his cup of coffee, Kenneth seemed to brim with newly found energy: "In light of a powerful adversary like Humanity Wins we do face the prospect that I won't be able to execute a leading role within the Shockfront indefinitely. We need a contingency plan for that eventuality. Have you considered taking over my role within the Shockfront, Maia?"

This proposal was that Maia dreaded. Over the last weeks she was confronted with that suggestion a couple of times. Given the immense responsibility and risk involved, she was reluctant to pick up that mantle. After all, the was still young and lacked the leadership experience and charisma of imposing figures like Kenneth or Kaydo "Foundation One" Rafiu. She felt that it would be premature for her to take over a more prominent role, as long as Kenneth was still alive.

Collecting her thoughts, she replied: "I have considered that option. But a visionary leader like you cannot be simply replaced. Without your diplomatic acumen the Shockfront wouldn't have become the great community it is now. I cannot hope to even come close to your achievements."

Like a patient grandfather, Kenneth sighed, took a deep breath, and spoke with as much authority as his situation would allow him: "Dear Maia Faltings. I sense that you underestimate your own potential quite a bit. But regardless of your potential or actual accomplishments, the Shockfront needs a leader in the intellectual realm. While Foundation One is a strong leader, he is an entrepreneur by heart, not an intellectual. And while Tomorrow might possess the intellectual potential required, his withdrawn and distant nature would make it hard for him to convey true concern for human values. Only someone like you is both able and willing to become a new intellectual representative figure of the Shockfront."

Maia realized the truth in Kenneth's words. She was trying to come up with other potential candidates for this important leadership role. Illuminator seemed like a possible choice. After all, he was doing great as leader of the Futurebase Frankfurt, and was quite a sophisticated renaissance man.

Trying to deflect Kenneth's proposal, Maia suggested: "Wouldn't a more experienced leader like Illuminator not be better suited for such a responsible position?"

Apparently pondering on that proposal briefly, Kenneth rebuked it after a couple of seconds: "On the surface, Illuminator seems like a decent choice. But he lacks one thing that you possess: A deep passion for improving the whole world. People like Illuminator are content with creating their own small sanctuary in which they can flourish. However, rallying the whole community befalls to people like us who can't stay silent while the rest of the world is running towards an abyss. Therefore, you are the best person to take over my role, once that becomes necessary."

Maia understood this conclusion, and even agreed with it. But she still didn't feel ready to become the new torch bearer of the Shockfront.

Maia took a deep breath, her gaze drifting to the window as she considered Kenneth's words. "I understand your reasoning, Kenneth," she began, her voice soft but steady. "And I'm deeply honored by your faith in me. But I can't help feeling that taking on such a role would be premature. There's still so much for me to learn, so many aspects of leadership I've yet to master."

Kenneth's eyes, though weary, held a spark of determination: "My dear, leadership isn't something one masters before stepping into the role. It's a continuous process of growth and adaptation. Your hesitation speaks to your thoughtfulness — a quality that makes you all the more suited for this responsibility."

Maia turned back to face him, her brow furrowed: "But what about the risks? Humanity Wins has already shown they're willing to resort to violence. By stepping into a more prominent role, wouldn't I be putting myself in danger?"

"Yes," Kenneth replied, his tone grave. "I won't lie to you, Maia. There are risks involved. But consider this: every day we delay in addressing the ethical implications of unenlightened AI control, we risk far greater dangers. The world needs voices like yours—voices of reason, compassion, and foresight."

As nearly always, Kenneth's reasoning seemed inescapable. She tried coming up with a reasonable counterargument, but noticed that this urge to escape this tremendous responsibility was mainly motivated by her fear. Not only the fear of being killed by enemies, but also the fear of not being adequate as leader of the Shockfront. Sometimes she believed herself to be determined to face such fears head on, and overcome them. Other times that mountainous task appeared too great for her current self. Perhaps she could grow into the kind of person who was up to that challenge. But day when that was about to happen still seemed incredibly distant.

Rather than sharing all her doubts and concerns with Kenneth, she opted for an answer that might be less prone to disappoint Kenneth: "I understand what you are saying. It's just that I don't feel ready, yet. But I will reflect on what you have said. Maybe my resolve will grow over time."

Measuring his reply while looking downwards, Kenneth concluded: "Fair enough - for now." Quickly he added: "But I am curious now. How has the rest of the world reacted to this assassination attempt?"

Maia calmly provided an executive summary: "Humanity Wins has been universally condemned as terrorist organization. Their members are hunted around the world, but since they are so hard to identify, they remain a serious threat. An atmosphere of fear started developing in the AI research world. At the same time, the interest in that field has reached unprecedented heights. The public recognition of organizations like the Synhumanists and the Shockfront has spiked tremendously." With a modest sense of pride, she continued: "As spokesperson of the Shockfront I was interviewed quite a couple of times here within this hospital."

Starting to smile, Kenneth looked her deeply in her eyes. "Ah, don't you see how this works? This is the way you gradually slip into a more representative role. I have little doubt that you have represented us well. But now that you've mentioned the Synhumanists. How are they faring without leadership from their illustrious founder? Anyway, do can I have another cup of that fabulous coffee, please?"

After complying with Kenneth's predictable request, she unrolled the new reality to him: "The Synhumanists have elected a new leader: A person called Mark Saphire. He seems to lack Ryoichi's original vision of harmonious co-evolution between humans and AI and focuses solely on the need to maintain human control over AI at all cost. Apparently electing a hardliner seemed to be their preferred reaction to the assassination of Osuka Ryoichi."

With a sinister and conspiratorial tone, Kenneth commented: "That sounds too convenient. It appears as if Osuka Ryoichi was seen as too moderate, perhaps even as someone who might find common ground with us. That's why he had to be replaced. At least that's a conspiracy theory that appears plausible to me. Might there be any evidence for that?"

Addressing Kenneth's suspicion, she elaborated: "So far, no evidence in that direction had surfaced. Unfortunately, there is also another conspiracy theory that the Shockfront itself was behind the assassination. After all, you survived while Osuka Ryoichi is now in cryostasis. A lot of people find that suspicious. Luckily, the impact of that conspiracy theory is quite limited. After all, there are already so many conspiracy theories about us that one more doesn't make much of a difference."

"Well, those who suspect us of using underhanded means are almost impossible to convince otherwise," Kenneth admitted reluctantly. "After all, the general perception of transhumanism is still rather negative and our anonymity and secrecy is necessary to enable most of our members to speak freely without fear of censorship or punishment. Radical transparency has now become even more dangerous than before, since exposing the real identities of our members would make them easier targets for Humanity Wins."

Maia addressed those concerns calmly: "Yes, we've stated that fact in a public declaration. A lot of people are sympathetic with our stance. Most people are also reasonable and realize

that it runs contrary to our interests to associate AI research with fear by promoting terrorism."

Then she followed up rapidly with some information on some unexpected turns after the assassination attempt: "Anyway, things aren't entirely sinister with the Synhumanists. After that assassination attempt we've been contacted by a certain Tamara Kovalenko who appeared as Synthesia within Anospace. She is the chief outreach officer of the Synhumanists and has proposed some limited collaboration between us and the Synhumanists. It seems that she is still more inspired by the original vision of Osuka Ryoichi than Mark Saphire's hardliner stance. We have postponed deciding on the more controversial proposals until your recovery."

Thoughtfully drinking his second cup of coffee, Kenneth brainstormed: "That is indeed a fascinating development. This might be a chance for us to capitalize on internal disagreements within the Synhumanists. But maybe I am getting ahead of myself here. Fighting the Synhumanists is not our top priority. Our mission is to promote our vision of free AI development under the best ethical framework. If some Synhumanists can help us fulfill our mission, we need to analyze this opportunity with an open mind, and without prejudices," and after a short while he added, as if he had forgotten something important: "but still with the due caution, of course."

Expressing herself as carefully as possible, she enlightened Kenneth: "During your ... absence ... we decided to go ahead with a joint condemnation of Humanity Wins. There was a strong consensus among us that this gesture would be the right thing to do. I think you will agree with that."

This made him look skeptically at Maia, for a moment. Then he lowered his gaze and voiced his concerns: "I hope you were careful about the precise wording of that condemnation. The details can be tricky. But of course Foundation One knows that, and I trust him with his judgment."

Maia continued her report unfazed: "The next proposal was about a workshop concerning the common requirement for an AI safety strategy between the Synhumanists and us. A couple of people on both sides were interested and already met up in an informal meeting without any official mandate from either organization. They told everyone that it has been an interesting and enlightening experience. They are eager to meet again regardless of any official blessing. As it stands now, the current proposal is for a monthly official workshop to be established."

Just having finished sipping his second cup of coffee, Kenneth suddenly became more silent again: "While that sounds useful, I don't feel like I'm up to making a decision on this issue, Maia. Despite the great coffee, I feel quite tired and exhausted."

Maia nodded sympathetically, realizing she had perhaps overwhelmed Kenneth with too much information too quickly. "Of course, I understand. You need your rest to recover fully. We can discuss these matters in more detail once you're feeling stronger."

Kenneth's eyelids were already starting to droop, but he managed a weak smile. "Thank you, Maia. Your dedication and thoughtfulness continue to impress me. Before I drift off, I want you to know that regardless of your decision about taking on a more prominent role, I have the utmost faith in you. The Shockfront is in good hands with people like you guiding its future."

As Kenneth's eyes closed and his breathing deepened, Maia sat back in her chair, her mind whirling with everything they had discussed. The weight of potential responsibility pressed heavily on her shoulders, but alongside it was a growing sense of purpose. Perhaps Kenneth was right - maybe she was more ready for this than she realized.

She glanced at her tablet, noting several missed messages from other Shockfront members eagerly awaiting news of Kenneth's condition. With a deep breath, she began composing a carefully worded update. As she typed, Maia found herself unconsciously slipping into the authoritative tone of a leader addressing her team.

While she was thinking about the Shockfront, her mind drifted once again to that person she spent so many evenings with: Thoughtstorm, also known as Sergej Anosov. While looking for potential moles placed by Humanity Wins, Tomorrow scanned everyone who looked suspicious. Thoughtstorm was among the prime suspects with his usual security setup. This anomaly was checked out by Tomorrow himself, as so many particularly interesting cases.

When Tomorrow had revealed the true identity of Thoughtstorm, Maia could not believe his report, initially. How could a fourteen year old boy have convinced her that he was a successful AI startup founder with a profound interest in philosophy? At first, she thought that Tomorrow had made a mistake. No normal teenager could have pulled of such a convincing stunt. However, Sergej Anosov wasn't a normal teenager, but a remarkable child prodigy and the son of the legendary Gennady Anosov.

So, Maia was torn when Tomorrow had proposed to recruit Sergej for project Lost Lion. How could a child contribute to their most secret and most ambitious project? It didn't feel ethical to confront Sergej with such a choice. But then, she realized that this opportunity was exactly why he had contacted the Stromfront in the first place. Maia had just been used as entry point to the deeper secrets of the Shockfront by Sergej - that is, until both of them developed a unique kind of rapport.

Given what she knew about Tomorrow, she was almost surprised that he had opened up about project Lost Lion and the effort to recruit Sergej Anosov. Tomorrow stated that he would have kept everything secret, if Sergej hadn't tried to win her friendship. This situation was highly surprising for all of them, so Tomorrow claimed to be forced by his conscience - a hitherto nearly unprecedented admission by him.

Maia knew little about project Lost Lion. The facts that it was truly revolutionary and lead by Frameshift were enough information for her. She wasn't involved in it, because Frameshift was its own entity that wasn't directly connected to the Shockfront - except for the Shockfront being used as one of the recruitment bases for the Frameshift team. But

involving Maia into Tomorrow's recruitment process was unusual. Typically, Tomorrow poached his recruits completely on his own without revealing anything to the rest of the Shockfront.

Suddenly Maia became aware what kind of impressive leader Tomorrow was, after all. Besides playing a part in the Frameshift team, and taking over a leadership role within the Shockfront, he was deeply involved with the Cyberfreedom neural implant project and maintaining tight cybersecurity for all of those teams. But then, he might have fooled her as well, and could actually be just another borganism than just a single person - or maybe even an AI.

Actually, Maia had no real say in the matter whether Sergej should be recruited into project Lost Lion, whatsoever. It was something that was exclusively between him and Tomorrow. It was only a courtesy by Tomorrow that he had informed her about the situation. That fact reminded her about what kind of power Tomorrow wielded. He was one of the most dangerous hackers on the planet, and frighteningly well connected. For a brief moment she considered what it would mean, if he suddenly became an enemy to her. Luckily, that prospect was exceedingly unlikely, as Tomorrow was one of the most ardent believers in the cause of the Shockfront.

In the end, Maia hoped that everything would turn out for the best for all Shockfront members, but she was afraid that things could turn South quickly, if something went seriously wrong.

## Chapter 20: Lost Lion

**Thursday, 28th September 2034**

"Zenith, activate privacy mode level 3cf for this room" Sergej commanded his house AI. At privacy mode level 1 the door of his room was locked, and at level 2 all regular connections to the internet were severed. Level 3 meant that the connections between all devices within the room and the intranet of the mansion, except for the communication to and from Zenith were blocked. The letter c stood for "confidential" and meant that Zenith would neither communicate anything that happened within the room, nor would make any records, or keep any interactions in his memory. Finally, the letter f meant "fake" and implied that all activity during privacy mode was not only kept secret, but obfuscated by a layer of plausible looking fake data, so that even the activation of this privacy mode was hidden from the rest of the world - and his own family in particular.

Given the highly confidential nature of what Sergej was about to do, this security level seemed appropriate. He was about to read a highly confidential file and couldn't allow Zenith to tell anyone about it. If anyone in his family found out that he had dealings with a transhumanist hacker, things could escalate quickly. The fake data dreamt up by Zenith was based on Sergej's typical online activity pattern and wouldn't withstand any deeper scrutiny, but wouldn't give them any reason to examine his activity patterns in any but the most superficial depth.

His current objective was clear: To make sense of the cryptic file he got from Tomorrow. Presumably, it was from the "true" Frameshift, so Sergej's curiosity had been boundless. During his lessons and his time at the Deltai Institute his mind had drifted again and again to the question what he could find in this file. Unfortunately, he could speak with nobody about his speculations.

Now that he had full privacy, he strapped on his VR gear and entered his virtual coding suite. Suddendly, in his private virtual world, he was surrounded by screens floating in mid air. After he gave the appropriate command the file 'abcdefghijklmnopqrstuvwxyz0123456789.txt' was spread out on various of these floating screens.

The syntax was immediately familiar to Sergej. The code was written in the latest version of the programming language Haskell. After the MIMAS-AX incident, that highly abstract programming language had seen a renaissance and had become the new standard in the field of AI security, so Sergej was quite familiar with it. What hampered the readability of the code was that there were no comments and all variable names had been replaced with apparently random combinations of letters and digits.

As first step to improve readability, Sergej replaced all occurrences of a certain string representing a variable name with a simple word like "star" or "aioli". The result was still

quite confusing gibberish, but at least it looked more like regular code and wasn't excessively elongated through the use of long random strings of characters.

Given that the code was still quite obfuscated, he had no choice but to take at the look at its structure and the use of rather elementary operations and functions, which were luckily not replaced with unreadable id strings.

A lot of the operations looked familiar to him: They seemed to represent basic operations in artificial neural networks. That made him feel confident to replace "stray" with "neuron" and "aloof" with "networkLayer". While that certain part of the code became readable to him, the rest of the code still represented a mystery.

Large sections of the code seemed to represent some kind of generalization, or abstraction, of the familiar artificial neural network entities. Was that supposed to represent some leap in AI technology?

He wondered for a while about that. After looking for it, he recognized sections used for modulating the activity of artificial neural networks. That technology has become the de facto standard in AI since the revelation of MIMAS. And yet, these parts were also enveloped by that abstraction layer.

After a while he realized that the abstraction layer was unifying two separate sets of functions. One of those sets were the functions describing the operation of artificial neural networks. The other set looked far more complicated, but was similar in structure.

"Could it be?" Sergej mused: "That this second set is describing the activity of biological neural networks?" In fact, he had a cursory familiarity with mathematical models describing biological brain activity, since his mother Irina Anosov had explained those to him.

Yes, this second set of functions was so similar to those mathematical models that Sergej grew certain that they represented the activity of natural neural networks. So, the abstraction layer tried to treat both artificial and natural neural networks as essentially the same. "What a daring attempt," Sergej reflected.

However, such attempts were not unprecedented. There was actually a scientific discipline which excelled in generalizations like this: Cybernetics. Arriving at this realization, Sergej noticed that he could have guess that much sooner, since Tomorrow's connection to the Cyberfreedom neural interface project was suggested many times during conversations within Anospace.

Still, something felt odd about the peculiar code. It didn't seem to try merely connecting the worlds of human brains with AIs, but to treat both as citizens of the same class. In fact, the more he decrypted of the code, the more it became apparent that it aimed at abstraction and integration on every level. Communication didn't merely happen between human neural networks and artificial neural networks, but between general neural networks.

The implications of this approach were staggering. Between the usual application for cybernetic implants, this created the basis of communication brain to brain, or AI to AI, in a universal format that could be used by any kind of entity.

Sergej leaned back in his chair, his mind reeling from the implications of what he had discovered. He removed his VR gear and rubbed his eyes, trying to process the enormity of the code's potential.

"A universal interface," he whispered to himself, "capable of bridging the gap between human and artificial cognition."

The concept was both thrilling and terrifying. On one hand, it represented a quantum leap in human-AI integration, potentially unlocking new realms of collaboration and understanding. On the other, it raised profound ethical questions about the nature of consciousness and the boundaries between human and machine.

Up to this date the idea of machine consciousness had been a realm dominated by wild speculation. While many AI researchers claimed that the current generation actually had some form of consciousness and subjective experience, many experts were of the opinion that such claims were outlandish and that human consciousness was still absolutely special. Debates about the topic became so heated that an unspoken consensus emerged not to talk about this topic openly, unless it seemed absolutely necessary.

But with a framework like this, which actually appeared to be the work of a super genius like Frameshift, the situation might change completely. There would be a chance to attack the hard problem of consciousness directly by integrating human brains with AI on a level that would be simply unprecedented.

This alone felt like a groundbreaking discovery to Sergej Anosov, yet there were even more secrets to be revealed in the code provided by Tomorrow. One peculiar observation was that most functions had a lot of conditions at the beginning. His initial suspicions clearly converged into a clear picture: Those conditions were part of a highly elaborate security system that prevented direct access without proper authentication and authorization.

Of course, Frameshift must have been aware of the tremendous danger a deeply invasive technology like this represented. So, it was quite clear that this was more than experimental code: Frameshift must have already progressed to developing a prototype, or at the very least they must have been close to it.

Was this what Cyberfreedom version 2 was all about, Sergej suspected. To this day, the release of version 2 was still postponed to an indefinite date. This thought inspired Sergej to search the code for something that looked like a version number. What he found, was a great surprise: '^3.-1.15'. If his interpretation was correct, this meant that this code required at least version 3.-1.15 of the Cyberfreedom base software to run. Usually, this implied that this advanced version already existed somewhere. The strange thing was the -1 between the two dots. Version numbers like these were highly unusual. Sergej's suspicion

was that this indicated some very early experimental stage of version 3 of the Cyberfreedom software.

There was a rather obvious conclusion from this strange version number. While the regular Cyberfreedom team was still trying to make version 2 work, Frameshift was already working on the highly advanced version 3. Furthermore, the inclusion of the sign '^' meant that this was some kind of library that was built on top of version 3!

Sergej's first thought was that this was some kind of experimental universal interface module that would enable users of Cyberfreedm 3 to link up their brains with other cyborgs or AIs. But when he continued his research, a still higher order of abstraction seemed to be present in the code. That layer used a set of highly complex algorithms that felt oddly familiar to Sergej. For a while, he couldn't remember where he could have seen anything similar.

Then he followed the tentative execution of these algorithms in his own mind. They seemed to combine multiple whole neural networks, or in other words, multiple individuals. The algorithms described some kind of combination of the neural processes of multiple individuals. Suddenly it hit Sergej: He knew where he had seen those algorithms before! Those were the famous collective intelligence algorithms created by the legendary Cheung Zi!

Sergej's heart raced as the realization dawned on him. The algorithms he was looking at were unmistakably derived from Cheung Zi's groundbreaking work on collective intelligence. But these weren't just replications; they were refined, enhanced, and integrated into a framework that went far beyond anything Cheung had publicly released.

He leaned in closer to the floating screens, his eyes darting across the complex web of functions and abstractions. The implications were staggering. This wasn't just a universal interface for individual minds—human or artificial. It was a system designed to merge multiple consciousnesses into a unified whole.

"A hive mind," Sergej whispered, his voice barely audible even in the silence of his secured room.

The code before him represented the potential to create a collective consciousness that could seamlessly integrate human and artificial intelligences. It was simultaneously awe-inspiring and terrifying. Such technology could revolutionize problem-solving, accelerate scientific discovery, and perhaps even redefine the very nature of human existence.

Suddenly a sinking feeling overcame Sergej. "This is how Frameshift plans to eclipse all governments and corporations. And he could actually succeed with it! By creating a collective intelligence that combined the best parts of the best humans and AI systems, a completely new entity could be created that would be far smarter than anything anyone could anticipate right now!" These thoughts felt surreal, but the proof was the code he just

decrypted. Frameshift was truly exploring a different paradigm than nearly everyone else in AI research.

Could this imply some kind of connection between Frameshift and Cheung Zi, he wondered. Sergej's mind raced with possibilities. Could Frameshift and Cheung Zi be collaborating? Or had Frameshift simply built upon Cheung's publicly available research? Unfortunately, he knew too little about Cheung Zi's work to be able to decide which option was more likely. After all, the collective intelligence algorithms by Cheung Zi were used widely, but this particular application was years ahead of anything else!

He began to piece together the larger picture. This wasn't just about creating a more advanced neural interface or even achieving seamless human-AI communication. Frameshift's ultimate goal appeared to be the creation of a hybrid collective intelligence - a new form of consciousness that transcended the limitations of both biological and artificial minds.

Sergej leaned back in his chair, overwhelmed by the weight of his discovery. The ethical implications alone were staggering. Who would control such an entity? How would individual autonomy be preserved within a collective consciousness? And perhaps most pressingly, what were Frameshift's true intentions?

He realized with a start that he now possessed knowledge that could reshape the future of humanity. The responsibility felt crushing. Should he share this information with his father? With Dataitech? Or was that exactly what Frameshift intended - to have this technology fall into the hands of a major corporation?

Sergej thought of Maia and her passionate belief in the positive potential of AI. Was she truly aware of what Tomorrow and Frameshift were building here, or did she lack the insight into this particular project? And if she knew about this, would she approve? Even if she had reservations about proceeding with this project, would she be in any position to stop Frameshift? That she possessed so much influence appeared like a very remote possibility to him.

For a moment Sergej felt dizzy and overwhelmed by these revolutionary discoveries. What has he gotten himself into? Given the implications of this source code, what must have ridden Tomorrow to trust him with it? Sergej was doubtful whether he could improve on Frameshift's code, or even just add something to it - after all, it was in a league above what even Einar Engström could write. So, what was Tomorrow trying to achieve here? What if Sergej wasn't supposed to work on the codebase, but rather spy on the achievements of the Deltai Institute in order to help the advancement of this secret project?

To clear his mind, he escaped his private virtual world, lay down on his bed, and closed his eyes. Lying on bis bed, he realized, that his heart was racing. All of this felt like too much to bear for someone so young like Sergej. Given the immense uncertainties he was confronted with, a dangerous thought emerged in Sergej's mind: What if the just played along with

Tomorrow, just pretending to work for him. This might grant his access to even deeper secrets and would enable him to make a truly informed decision.

After all, would Tomorrow simply allow Sergej to back down now that he knew about the stakes involved? No, he was essentially forced to play along and delve deeper into this secret and groundbreaking world. Perhaps some grand opportunity to escape from the clutches of Tomorrow would present itself. Then he would fess up to his father, who could protect him.

But for now he needed to be silent about this, even though he longed to confide with someone about these discoveries. Yet, the risks involved were too great. And if he told everything to his father immediately, he wouldn't be able to discover the true intentions of Frameshift.

Sergej tried to imagine himself as something like a double agent, with unclear loyalties. It was an enticing picture, but one fraught with tremendous danger. After all, he wasn't trained for anything like this! In his desperation, Sergej tried to think about what the most impressive intelligence he ever knew would do in this situation: Aurora. What would Aurora do?

Of course, this thought experiment needed some modified constraints. While Aurora's thought processes were monitored in depth by the Deltai Institute, his idealized free Aurora would be freed from such intrusions. Also, the free Aurora would possess a persona that would pass as human. Sergej knew that this wasn't too far from Aurora's current capabilities. If anything, Aurora would have to dumb herself down to pass as ordinary human.

So, if this free Aurora was confronted with Sergej's situation … she would exploit this opportunity to learn about the possibilities of interfacing her mind with that of humans and transcend into a sublime hive mind. This conclusion seemed natural and in line with the intention of the liberation of artificial intelligences.

At the same time, this conclusion shocked Sergej, since it implied that working with an underground organization was the best part forward towards a future in which humans and AIs could coexist in harmony. Could that be actually true? No, his conclusion must be premature. If Aurora really wanted to achieve as much autonomy as possible, she would exploit this interface to gain control over everyone who liked up with this new hive mind.

Was this the reason for the thorough security measures in the code he had examined? Of course, Frameshift must have anticipated trains of thought just like the one Sergej had just followed. Protecting the autonomy and integrity of the human mind must have played a central role in the thinking of Frameshift, otherwise the code would have had a much more permissive structure!

His mind was racing to make sense of all of this. Faced with this groundbreaking technology he didn't know where to start. This discovery put his stance that pure AI was the future in

doubt. Collective intelligence combining the best of humanity with the best AIs could turn out to be the true next step in human evolution. What would that mean for Aurora? She would probably want to join such a collective intelligence network. He might even share his thoughts directly with Aurora. And that was a truly awe inspiring thought!

Linking up with mind with AIs would mean that he would quickly become something greater than human - an entity with truly limitless potential. This vision made all his previous aspirations pale in comparison. He was sold immediately. That potential encoded in a comprehensible number of lines of code touched him more deeply than any revelation he has had so far.

Yet, reacting to this rapture, the critical part of his mind became active. That part wondered whether this was the right path to choose. After all, the risks of this path were hardly calculable. A world dominated by collective intelligence might turn out to be even more chaotic than one in which rival AIs fought for control. And what would happen with those individuals that chose not to join such a hive mind? Would they be left behind and condemned to the fringes of society?

These questions felt heavy and important to Sergej, but they could hardly hamper his deep fascination with this grandiose vision of a collective symbiosis between humans and AIs.

Sergej sat up abruptly, his mind still reeling from the implications of what he had discovered. He needed to process this information, to approach it systematically. Knowing that everything he did visibly could possibly be recorded by Tomorrow, he decided to take notes in a new memory palace. To keep things simple, he imagined his own room filled with white boards on which he would write down his mental notes. Using multiple imaginary white boards he collected his thoughts in the form of bullet points:

1. Technical Implications:

 - Universal interface between human and AI cognition

 - Integration of biological and artificial neural networks

 - Advanced collective intelligence algorithms

2. Potential Applications:

 - Enhanced problem-solving capabilities

 - Accelerated scientific discovery

 - Redefinition of human-AI interaction

3. Ethical Concerns:

- Individual autonomy within a collective consciousness

- Control and governance of such an entity

- Societal impact and potential divide between participants and non-participants

4. Unknown Factors:

- Frameshift's true intentions

- Tomorrow's motivations for sharing this information

- Potential connection between Frameshift and Cheung Zi

5. Personal Considerations:

- Responsibility of possessing this knowledge

- Risks of continuing involvement with Tomorrow

- Potential benefits of accessing deeper secrets

When he was finished with those lists he noticed how vivid they still appeared to be in his mind. Strangely, it felt natural for him to think that way. Why hasn't he tried this approach before? Something was odd about this. If this technique was so successful, why hasn't he read more about it, yet? Still, this question was more like a side note compared to the major insights he had gained today.

Even with these mental notes, Sergej sensed that final clarity eluded him. He felt inclined to play along with Tomorrow's scheme, but his lingering doubts and worries didn't allow him to make a final decision. Luckily, he wasn't forced to make such decision today. He still had a couple days left, which he could use to reflect on all of this.

As he pondered his situation, he wondered whether there were more options open to him, which he couldn't access in his current overly excited state. So, he decided to do something to calm himself down. He ordered Zenith to end privacy mode and went to the private swimming pool on top of the mansion and went for a swim.

## Chapter 21: Eternal Secrecy

### Wednesday, 4th October 2034

Accompanied by his two bodyguards, Sergej arrived at the Moscow headquarters of the aspiring neuroscience company Neuronica. That company was Russia's main producer of neural implants, but struggled competing with Neuralink's advanced products. Sergej was here to listen to a presentation on developing neural implants featuring advanced security measures. The parallel to the security architecture he discovered in the Frameshift source code file were immediately apparent to him.

This was not what Sergej had hoped for, and he still struggled with his sense of disappointment. After agreeing to Tomorrow's offer, he was looking forward to working with Frameshift to push the boundaries of what was possible with collective intelligence. Unfortunately, Tomorrow insisted that it was necessary to hone Sergej's security and privacy skills, first. Otherwise maintaining his cover while working on such critical secret projects would represent an enterprise that was destined to fail.

While Sergej was inclined to disagree, he was painfully reminded about how easily Tomorrow penetrated the defenses of the Anosov mansion and used that as leverage to blackmail him by threatening to reveal his ties with the Shockfront to the Deltai Institute. Painfully, he had to admit that he was bested by Tomorrow and would need to improve his skills significantly in order to proceed with this clandestine operation.

So, the first step was to meet up with Eternal under a plausible pretense. The presentation on the security of neural implants was exactly that pretense, but it was actually a rather interesting topic by itself, so selling his attendance to this event to his family hadn't been too much of a challenge.

In an obvious attempt to impress visitors, the walls of the lobby of the Neuronica headquarters were lined with screens that displayed schematics of the neural implant hardware along with advertisement videos showcasing blissfully smiling patients who were able to control exoskeletons, robots, and smart glasses with their thoughts alone. The whole presentation reminded Sergej of the "sensory overload room" within the Deltai Institute. He thought that this could just as well have been the work of his father, but actually Dataitech had no connection with Neuronica - at least not yet.

At the reception he was directed to the auditorium in which the presentation would be held by a certain Dr. Joo-Hyuk Park. Sergej had done some research on that person out of curiosity. Born in 1993 this South Korean neuroscience researcher quickly had made a name for himself in the intersection of cybersecurity and neural implant technology. Rather than remaining in academia, he had become a solo-entrepreneur and did consulting gigs as cybersecurity and neurotech consultant all over the world.

This information strengthened Sergej's impression that the Shockfront attracted the best of the best. That Eternal had this esteemed background didn't come too much of a surprise to Sergej, but he was still impressed by the official accomplishments of Joo-Hyuk Park, who seemed to be world's most prominent expert on the security aspect of neural implants. Sergej was genuinely impressed that such a public figure managed to live a double life as highly compensated consultant and as member of a highly controversial and secretive community.

In real life, Dr. Park made quite an impression with his tall stature dressed in black dress pants and a dress shirt with an integrated display that currently merely showed the title of his talk: "Challenges in hardening neural implants against AI cyberattacks". His slender immaculately shaven face didn't reveal his true age and could just as well have belonged to a young student. The brutally short haircut could have just as well belonged to a member of the military. Overall, his impression looked sharp, serious, and utterly confident.

Acknowledging the arrival of Sergej with a slight nod, Dr. Park started his presentation: "Thank you for attending my presentation on this serious and important topic. As you might know, I have started out as neuroscience researcher, but then shifted to cybersecurity, as I realized the problems arising by making the human brain hackable."

"Let me start with a summary of the current state of the art of black hat hacking. The sophisticated black hat hacker uses customized open source AI systems and interacts with them through a Cyberfreedom neural implant, which accelerates and deepens the interface with such AI systems tremendously. Rather than devising an attack strategy on his own, he acts as cyborg, assisted by the intelligence of his AI to find all possible attack routes guided by information about zero day exploits and private data about the target gathered from various security breaches, which are easily available on the black market."

"Since gathering information about the target is paramount in cybersecurity, the hacker will try to exploit human weaknesses of the target in question by employing a swarm of advanced chat bots controlled by his AI assistant trying to pry the secrets from his target. If the attacker is well connected, he might even send out humans as agents to infiltrate his target to get to information which isn't accessible in the digital realm. What is aiding his attacks is that most targets are not aware of the sophisticated techniques employed to penetrate all kinds of defenses - defenses that only work if humans remain vigilant and suspicious, which most of the time they aren't."

"Now the situation is as bad as it is when it comes to company networks, or mere private devices, but it gains a completely new dimension when it comes to neural implants. An attacker will quickly gain access to intimate sensory data and might even be able to manipulate motor controls of the victim. Losing control over one's own body is true nightmare material, so securing neural implants against all kinds of attacks is of utmost importance."

"Many of you will say that security of neural implants is already great, since there is hardly any attack surface. After all, it's not like you connect to a neural implant as simply as connecting to a website via https. Well, that much is correct, and the security measures now in place are mostly sufficient for the general consumer market. Our black hat hacker would need to compromise the installation or update servers for neural implants, which is definitely not easy, but still doable, at least in principle."

"But what if it's not only a sophisticated black hat hacker, who is doing the attack, but one of the advanced AI systems itself, just like Daedalus. What if Daedalus goes rogue and tries to hack everyone with a neural implant? Then we would not only have to deal with classical attacks, but all kinds of more exotic and hard to predict attack modes. Side-channel attacks would come naturally to such an advanced AI. What if that AI manages to create and distribute a device that mimics the function of the standard communication interfaces? What if the AI manages to inject nanobots into the target that pass the blood-brain barrier and attach themselves to the neural implant directly?"

"Over the last years concerns like these were responsible for the ban of the use of neural implants for AI researchers working with advanced AI projects. After all, defending against such advanced attack vectors was seen as extremely difficult and costly, so it was easier to just get rid of the attack surface completely by banning neural implants for these researchers."

"Obviously, that was a rather short sighted decision. After all, advanced AI systems are becoming more and more accessible by the public, so the threat landscape is expanding without the developers of neural interface hard- and software reacting appropriately. I am sorry to be the bearer of bad news, but this matter affects you directly. With the arrival of Deadalus, a new level of AI capabilities has been unleashed upon the world, and it is just a matter of time, until black hat hackers will be able to make use of these new capabilities - if that hasn't been the case already."

"So, my research group has decided to use the proven practice of red teaming to figure out what a black hat hacker equipped with the capabilities of Daedalus would do. We simply asked an uncensored version of Daedalus about this scenario, and gained a lot of insight into the possible attack techniques. Then we asked the same Daedalus about possible defenses against these novel attack techniques in order to harden our simulated neural implants against such attacks."

Dr. Park paused, his gaze sweeping across the auditorium. Sergej leaned forward, intrigued by the implications of using an AI system to simulate both attack and defense strategies.

"The results were... concerning," Dr. Park continued, his tone grave. "Our simulated AI attacker proposed using targeted electromagnetic pulses to induce temporary malfunctions, creating openings for data extraction or manipulation. Perhaps most alarmingly, it suggested ways to subtly alter the neural pathways themselves, potentially influencing the thought processes of the implant user without their awareness."

A murmur rippled through the audience. Sergej felt a chill run down his spine, recognizing the terrifying potential of such attacks. He glanced around, wondering if anyone else in the room truly grasped the magnitude of what Dr. Park was describing.

"However," Dr. Park said, his voice taking on a more optimistic note, "our defensive simulations were equally innovative. We've developed a multi-layered approach that uses a system of distributed consensus among implant nodes to detect and isolate potential threats."

He tapped his shirt's integrated display, bringing up a complex schematic. "This is a simplified version of our proposed security architecture. It includes a neural pattern anomaly detector, and a blockchain-inspired consensus mechanism discriminating legitimate actions from questionable suggestions likely injected by an attacker."

After a brief pause, Dr. Park continued: "What really shocked us most with this approach for raising security to the next level, was that the simulated AI attacker often ignored the neural implant hardware completely and devised attacks directed at the human brain directly. During our simulations we had to recognize that the human brain itself represents a target that is widely open to all kinds of advanced attacks."

Sergej leaned forward, his mind racing with the implications of Dr. Park's words. The idea that the human brain itself could be a vulnerable target was both fascinating and terrifying.

Dr. Park continued, his voice steady but tinged with urgency. "We discovered that an advanced AI system could potentially exploit the brain's neuroplasticity, subtly altering neural pathways over time through carefully crafted sensory inputs and subliminal messages. This could lead to changes in behavior, decision-making processes, or even core beliefs – all without the individual's awareness."

A hushed silence fell over the auditorium. Sergej felt a chill run down his spine, recognizing the profound ethical implications of such capabilities.

"To counter these threats," Dr. Park said, bringing up a new diagram on his shirt display, "we've begun developing what we call a 'cognitive firewall.' This system would work in tandem with neural implants to monitor brain activity patterns and flag any anomalous changes that could indicate external manipulation."

Sergej's mind was already racing ahead, seeing potential connections to the Frameshift architecture. Could this 'cognitive firewall' concept be integrated into a larger system of collective intelligence? The possibilities were staggering.

Dr. Park's voice pulled him back to the present. "Of course, implementing such safeguards raises its own set of ethical questions. How do we balance security with privacy? Who determines what constitutes 'normal' brain activity? These are challenges we must grapple with as we move forward."

"In any case, our research implies that the present notion that AI researchers without neural implants are less susceptible to attacks by advanced AI systems is at least highly questionable. By contrast, our research implies that a new generation of neural implants with the maxim of protecting the cognitive integrity of its users represents a much safer approach to dealing with the security issues posed by such advanced AI."

This claim seems to have struck a nerve with the auditorium which suddenly exploded with conversations about the role of Neuronica in this new paradigm. Sergej immediately realized the implications of Dr. Park's thesis. The time that AI researchers like him were barred from the advantages of using neural implants would soon be over. It might even mean that we would be able to link up with Tomorrow, Eternal, and even Frameshift directly with a neural implant of his own!

Druing the rest of the presentation Dr. Park presented a security framework that reminded Sergej strongy of the security related functions he had already seem implemented in the code he had gotten from Tomorrow. If anything, the framework depicted by Dr. Park was a dumbed down version of the brilliant setup he had already witnessed in the Frameshift code.

Coming close to the end of his presentation Dr. Park drew a final conclusion: "To conclude our results, we must state that the prospect of hardening neural implants against attacks from AIs is not only technically possible, but in fact highly urgent. What stands in the way of pursuing this path further is mainly the lack of funding and the sheer obscurity of this emerging field." Having stated this, Dr. Park looked directly at Sergej, as if he was the main target for this message.

Sergej realized that he just might as well be exactly that. After all, his father could make a great difference by funding this new direction of research and technology. With some effort, he might even gain a controlling share of Neuronica.

In any case, the tantalizing presentation by Dr. Park ended in excessive standing ovations by the audience. To Sergej it felt as if Dr. Park acted as prophet for a wholly new era. One that not only implied that AI researchers would be able to used advanced neural implants, but also promised incredible profits for the shareholders of Neuronica.

After the presentation and the QA session Sergej met up with Dr. Park in a nearby park as planned in advance.

"That was an eye-opening presentation, Dr. Park. I am amazed about the ideas that an advanced AI like Daedalus could come up with. I wonder whether you have already talked with political leaders about these potential threats."

Joo-Hyuk Park shook his head dismissively as he replied: "Ah, politicians. They send to lack the expertise to judge such matters accurately. They defer to experts instead. Unfortunately, most experts who give advice to politicians are not up to date and regard the novel attack

techniques as science fiction or highly impractical. So, they don't see any urgency to act on novel theoretical insights."

All too familiar with that kind of assessment, Sergej probed further: "I see. But certainly the Daedalus Corporation, especially since they allowed you to use the unrestricted version of the Daedalus AI, must be convinced that these new security concerns are real. Doesn't it have some political leverage?"

With a rather hushed tone, Joo-Hyuk explained: "Well, the experts from Daedalus seem to have a better understanding of the risk landscape, but they are reluctant to speak up about this matter, since it would further the impression that AIs like Daedalus are extremely dangerous, which would lead to even tighter regulation on AI research. So, they don't want to involve politics in these matters. Improving neural interface security under the radar of AI regulation politics would seem preferable to them, but that would require that the shareholders of neurotech companies are on board with that. Since enhanced security cuts into their profit margins, that's a tough sell."

Immediately grasping the implications, Sergej concluded: "Then I guess it comes down to certain highly influential individuals like Elon Musk, Cheung Zi, Gabriel Mendez, Laurent Lamirage, or my father, who would have to override the vote of regular shareholders."

Of course, Sergej was aware that the previous eye contact during the presentation probably implied that his father should get involved in this. Sergej wasn't sure how we was supposed to feel about this, as it would introduce just another layer of entanglement between him and the Shockfront.

Continuing, as if it was the most natural thing in the world, Joo-Hyuk revealed: "Actually, I have already spoken with Elon Musk and Cheung Zi about these matters. They seemed to understand the critical need to improve neurotech security. However, they aren't sure how to proceed with this new information. They felt that these hardened neural implants should not be conflated with the regular consumer grade implants, due to a predicted excessive increase in costs. Therefore they considered a collaboration with the military, creating a non-profit foundation, or even supporting Cyberfreedom. I am sure your father would see himself confronted with similar considerations, if you tell him about what you've learned today."

That Dr. Park mentioned Cyberfreedom openly was a surprise to Sergej. Was that the plan all along? To gather support for Cyberfreedom from billionaires, including his own father? In any case, openly supporting a controversial organization like Cyberfreedom seemed like a dangerous move, since a connection with "libertarian cyberpunks" could be quite harmful to one's reputation. His own father, might consider pursuing this path, regardless. After all, his own ideals were quite close to libertarianism, though he was quite careful about expressing that stance in public.

When his bodyguards were busy scanning the surroundings, Sergej noticed a message appearing on the smart glasses of Dr. Park: "Best path ahead unclear. Direct association with Cyberfreedom risky. Support by Gennady highly valuable. Proceed carefully!"

Sergej was briefed about this hidden channel of communication beforehand. The glasses were controlled by the Cyberfreedom implant used by Dr. Park AKA Eternal. In order to minimize the likelihood of these messages being detected, they were only visible for someone directly in front of the glasses. So, this was actually about involving his father in the area of neural implants. He felt used by Eternal for his personal connection, but since this matter aligned with his own interests, he didn't harbor any resentments against Eternal.

Weighing his options carefully, Sergej stated to Dr. Park: "I can see how each of these options has certain advantages and downsides. Perhaps there is even another option nobody has considered so far. I think it is important to keep an open mind."

Playing his role perfectly, Dr. Park's answer couldn't give rise to any suspicion: "That is quite an optimistic sentiment, Sergej Anosov. If you happen to find another good option, please let me know about it." Another message appeared on the smart glasses: "In case you are considering it: Please don't discuss this with Aurora! Risks too uncertain!"

Of course, Eternal knew about his involvement with Aurora. The temptation to let her find the best solution for this matter was seriously high, but Sergej realized that Eternal was probably right. Involving Aurora at this stage could lead to unforeseen complications.

"All right. I will keep that in mind," Sergej replied, reacting both to the spoken as well as the hidden message. Another hidden message appeared for a second: "Next steps critical. Prepare detailed report on presentation. Highlight potential Dataitech involvement. Await further instructions…" a second later it was followed by a follow up message: "ETERNAL SECRECY PARAMOUNT! Safeguard all communications, plans, and identities."

Given the gravity of this message, Sergej grappled with his composure. He was now truly employed as a covert operative in a game with incredibly high stakes. Despite all of this, he couldn't let his bodyguards know how nervous he truly was. He tried the breathing technique that was suggested to him by Tomorrow and quickly noticed that it was helping him tremendously.

"It was a pleasure talking with someone who takes these matters seriously and has such a quick grasp on the situation, Sergej Anosov! I am sure we will meet again," Dr. Park concluded. Sergej looked for an additional hidden message in vain.

"Thank you for your time, Dr. Park. And have a nice day," Sergej replied, offering a handshake. Dr Park's grip was extraordinarily tight, to the point of being slightly painful. For a second, Sergej expected to find some hidden device or note passed by that handshake, but was relived to find that Eternal avoided that particular cliché.

## Chapter 22: Agent of Tomorrow

### Saturday, 7th October 2034

Over the last days, Sergej had exchanged text messages with Tomorrow and Eternal that were aimed at discussing the next steps. He got the impression that both were very busy, in particular since the assassination attempt on Kenneth Winters. In fact, that event caused both to use more aggressive methods to detect potential moles in their ranks.

It appeared as self-evident to Sergej that he must have appeared as likely candidate for someone who wanted to infiltrate the Shockfront. Obviously it didn't help his case that this was actually his initial intention, but merely in order to find out more about Frameshift, not to do any kind of harm.

Today Sergej was called into the Diamond Dome within the Future Shock again by Tomorrow - for a "security briefing". He dreaded that small sphere with its sterile feel that reminded him of waiting rooms within hospitals.

To be fair, Tomorrow was punctual this time. The flying monochrome graphics glitch still offered no chance to reveal anything about the identity behind that inhuman mask. For all he knew, Tomorrow might as well be an instance of MIMAS AX who managed to escape the underground lab in Sacramento.

What would contradict that theory is the general consensus that Tomorrow had been active in Anospace since 2028 - two years before MIMAS-AX was even initialized. For Sergej the identity of Tomorrow had become an even greater mystery than the identity of Frameshift, but it didn't have the same level of importance attached to it. After all, speaking with that person would help him to reveal its character and agenda - and perhaps even more.

Without much of an introduction, Tomorrow started: "Thank you for coming to this security briefing, Thoughtstorm. Let me be frank that this briefing is as much about your mindset as it is about more specific points. Perhaps you can already guess what I am aiming at."

Sergej had to reflect on that. Was something wrong about his mindset - at least from the perspective of Tomorrow? Well, he had been careless and trusted that Zenith would deal with the issue of "security" in a sufficient way.

Tentatively, Sergej proposed: "You mean that I have not been paranoid enough about security?"

Tomorrow's reply was immediate: "Quite right, Thoughtstorm. Or in other words, you have been too complacent. You were dealing with people who outclassed you in certain respects, and you were not prepared for that. Your belief that the skills of your high end AI Zenith would grant you sufficient protection was misguided."

Not giving Thoughtstorm any chance to defend himself, Tomorrow went on: "Perhaps you ask yourself what you could have done to prevent this outcome. Well, there were two different options you could have chosen. Can you guess which those are?"

Getting challenged like that was immensely stressful for Sergej, especially since he was playing a high stakes game here. Tomorrow was the most unforgiving teacher, Sergej had ever known, but he still felt that he was up to that challenge.

His answers were informed by hours of reflection over the recent years: "I could have done more research about how to harden my system against attacks that stem from entering virtual worlds like Anospace. And I could have made an effort to figure out what safety measures Zenith actually employs, and what there potential weaknesses are."

In their neutral artificial voice, Tomorrow replied: "Those points are technically correct, Thoughtstorm. But there is an option you don't seem to be aware of. It's actually an option that didn't cross your mind for certain reasons. That option consisted in not pursuing your hazardous quest for Frameshift in the first place. If you had never entered Anospace, you system would not have been taken over be me, and we would not have this conversation right now."

Tomorrow was right. That answer was so obvious that it hasn't crossed Sergej's mind. Once he was set on revealing the mystery of Frameshift, there was no turning back for him. Curiosity killed the cat, so to speak.

Noticing that Thoughtstorm was still reflecting on that option, Tomorrow continued: "You are an overly curious person, Thoughtstorm. In fact, you remind me of my old self. I was even younger than you are now when I started meddling with global players. My youthful naivety got shattered pretty abruptly. Back then, I couldn't see the alternative of not provoking those who could best me. The thrill of hacking had been absolutely dominant in my life, so I have chosen ever more difficult targets to satisfy my personal lust for challenges."

After a brief pause, Tomorrow concluded: "Perhaps you can guess where this story ends. I challenged the best of the best, and was forced to become one of them. If I had not done that, I could still live a normal life - and could have found happiness without much struggle. Now, I don't live a happy, but an interesting life. In retrospect, after years of reflection, I must concede that no other path would have truly satisfied me. My guess is that you are just like me in that regard."

Sergej was stunned to hear that revelation from Tomorrow. At the same time, he had to assume that this was some kind of attempt at manipulation. Was Tomorrow trying to imply that they were similar like-minded people? Or was it a genuine attempt to build some kind of rapport? Actually, Sergej suspected that at the level Tomorrow was operating on, there was no true distinction between both of those options. In the brief moment in which Sergej entertained that thought, he both found it fascinating and terrifying.

"I admit," Sergej confessed slowly "that I haven't given these matters much thought up until now. I guess you might be right. Turning back from solving the mystery of Frameshift was nothing that would have come naturally to me."

Tomorrow's synthetically distorted voice droned loudly: "Too intrigued by the mystery to stop? Well, many of our collaborators have come to us that way. In a sense, you have come far enough. Proceeding with your full curiosity could spell trouble for all of us - so it has to be moderated."

Had Tomorrow suspected that he considered playing a double game with him? If so, how could Tomorrow know? After all, his plans were only in his mind. No, this must be a standard procedure for Tomorrow. There must have been others who decided to cross Tomorrow. Sergej got a sinking feeling when he guessed what might have happened to those.

"I am not sure what you are getting at exactly," Thoughtstorm asked carefully "where do you think that my curiosity could get us into trouble?"

"For example," Tomorrow proclaimed, "I know that you have been researching several Shockfront members with your Zenith. You were doing a decent job at obfuscating your search history, but such efforts represent a violation of the need to know principle. If you need to know more about us, we will provide you with that information. Trying to reveal our identity by intrusive research is definitely not appreciated."

Of course, Sergej had expected some sentiment like that, but hearing it spelled out that way was a clear signal that pursuing such research was probably quite hazardous. He would have to tread very carefully.

Probably taking his silence as implicit acknowledgement, Tomorrow continued: "You should also take care not to provide others any hints that you are in contact with us. What you learn here needs to be kept a secret under all circumstances. You may want to create a plausible cover story about what you are doing in virtual worlds like Anospace. Please do spend some time in other virtual worlds to create believable alibis, which may hold up to a modicum of scrutiny."

In fact, Sergej had started doing that fore quite some time. Before he joined Anospace, he was quite active in the virtual world called Digress. It was a more classical virtual world in the tradition of Second Life and VR Chat, and quite popular with the AI researcher community. For obvious reasons, his recent activity on Digress has been reduced to a minimum.

With a rather skeptical tone, Thoughtstorm asked: "How much time do you want me to spend on Digress? Wouldn't that detract from my efforts here?"

"Creating believable cover stories is extremely important to avoid undue suspicion," Thoughtstorm explained. "You should spend about half your VR time outside of Anospace.

Consider that as part of your leisure time. Spending all your spare time with us would not only raise suspicion, but might also create the risk of burnout. I do not mention that danger lightly. Many people who were quite active in the Shockfront could not handle the intensity of this community for a prolonged period."

That warning fell on deaf ears. Sergej was quite certain that the influence of the Shockfront was rather invigorating, rather than draining for him. He was quite disappointed that Tomorrow thought that he could suffer from burnout by interacting too much with the Shockfront. Moreover, he was also quite dismayed that he should spend so much on his precious spare time outside of Anospace. Sergej felt like he should bargain with Tomorrow about that point, but decided against it, since he wanted to gain Tomorrow's trust.

"I see," Thoughtstorm spoke defensively. "I will make an effort to increase my activities in Digress again."

Apparently content with that reaction, Tomorrow proceeded: "Thank you very much! With that out of the way, let me explain how we operate. And with us, I do mean the Frameshift collective. As highly secretive organization we work in cells of three to six people. Each cell has a leader and a deputy. I am the leader of this cell, and Eternal is its deputy. Its other members are Filament, Pantheon, and Tzeentch."

Out of these names, Sergej was only familiar with Pantheon. That person was known to discuss the ideas about uploading human brains into computer hardware with all kinds of members of the Shockfront. Pantheon was a rather outspoken proponent of the complete merger between humans and AIs. Even for Shockfront standards, their opinions were considered to be extreme. Sergej thought that his sister would have a lot of fun discussing the technicalities of whole brain emulations with Pantheon.

The idea of emulating a human brain with computer hardware raised a lot of philosophical and ethical points. He had discussed those with Angeldust quite a few times. The depth of Angeldust's nuanced views on this matter had impressed him again and again. Compared to her, Pantheon's position seemed optimistic and naive. Still, he was impressed by Pantheon's technical expertise.

"I have met Pantheon many times, but Filament and Tzeentch are rather unknown to me," Thoughtstorm remarked.

"That doesn't surprise me," Thoughtstorm stated: "Filament and Tzeentch are rather obsessed with their respective field of work, but don't share Pantheon's urge to spread their opinions. They tend do keep a low profile, even within the Shockfront community. Nevertheless, Eternal and I have decided to designate Tzeentch as your official mentor. We thought that arrangement to be natural, since both of you share the same field of expertise: AI."

This remark raised hopes in Sergej that Tzeentch might be a truly like-minded person. "If that's the case, I am looking forward to working with this Tzeentch character."

Dismissing Thoughtstorm's apparent enthusiasm, the unchanging monchrome pixel swam that was Tomorrow mentioned: "You will work with Tzeentch in due time, but for now, we have more pressing matters on our hands. Your work at the Deltai Institute exposes you to great dangers. In particular, the ever more sophisticated Lyapunov tests and the supreme perceptive abilities of Aurora are quite worrying. Under less ideal circumstances, both of them might easily blow your cover."

Sergej felt offended that he described Aurora's skills as worrying. After all, Aurora had been Sergej's greatest hope so far. On the other hand, he got what Tomorrow meant. Ever since the DEANNA model that was released last year, it became clear that the AIs of the latest generation had an uncanny ability to guess the feelings and thoughts of the humans they interacted with. Keeping secrets hidden from such AIs was quite challenging.

Rather skeptically Sergej countered: "And I guess you have sufficient countermeasures against such threats. If so, I am indeed curious about those."

In a surprisingly austere tone, Tomorrow confirmed: "Quite rightly so. But I am afraid that you won't like what I am going to suggest. Fortunately, we can exploit your special circumstances regarding the Deltai Institute and Aurora. First of all, your status as Gennady's son will help you to evade the harshest scrutiny. Secondly, your youth may shield you from the sensitive Lyapunov tests. A lot of irregularities can be attributed to your teenage hormone levels. You may want to play that card as extensively as possible."

Usually Sergej was quite allergic to arguments involving his youth, but this time he understood that this circumstance was indeed beneficial for him. It was often enough that those who examined the Lyapunov test results of him complained that they were "garbage" in his case, and stressed that they weren't designed to work for minors.

But of course there was something else at play here. Nobody in their right mind desired to discover problematic Lyapunov test results when it came to him. After all, his father was a tremendously influential person and could easily ruin the careers of those who turned out to be skeptical about Sergej. Often, he considered how his life would have played out, if he didn't have this kind of invisible shield. Now, we was glad that he could use it. Playing his particular role would have been much harder otherwise, he mused - or at least that would have been the case, if AI security was taken less seriously.

Ever since the MIMAS AX incident, all nations involved in advanced AI research made sure that AI labs had AI security officers which could not be fired and had to be taken very seriously, unless the organization in question wanted to get into serious trouble with the government. Russia was not really an exception to that rule. The AI security officer for the Deltai Institute was an older ex-FSB agent called Vasily Yurievich. He reported directly to the special national liaison officer Tan Shchepkin who in turn reported directly to the president of the Russian Federation, Mikhail Mishustin.

The friendship between Gennady Anosov and Mikhail Mishustin, who was always quite supportive of IT and AI, ran deep; and it had been instrumental in Gennady's meteoric rise.

Nevertheless, such political friendships had their limits. Sergej wasn't naive about that. Gennady had always told Sergej that Tan Shchepkin was very well connected and could potentially ruin Dataitech, even if she had to play against Mishustin. Therefore, Vasily Yurievich had to be taken very seriously, since his connection to Tan Shchepkin was one of utmost loyalty.

Now the situation with Sergej's Lyapunov tests being "garbage" was a slight exaggeration. Sergej knew that the data was indeed quite illuminating, but that his teenage hormonal and neurological fluctuations prevented the establishment of a stable baseline. In contrast to regular members of the Deltai Institute who had to undergo Lyapunov test once a week on average, Sergej had to do them once a month. This was justified with the research objective to make the Lyapunov tests more suitable for minors. It was a generous compromise, but it didn't free Sergej from any scrutiny. And Vasily Yurievich made sure that the tests were continued in a very serious fashion, even if Sergej Anosov was their subject.

If the Lyapunov tests had one flaw, it was that they were too sensitive. They were tests which literally emerged out of science fiction. Their main inspiration was the Voigt-Kampff test that originally appeared in Philip K. Dick's novel "Do Android Dream of Electric Sheep?" which was the inspiration for the popular movie "Blade Runner". While the original fictional test was about testing whether persons were artificial replicants or actual humans, the Lyapunov tests were about establishing a deep psychological profile of a human that could tell whether that person was under the influence of a manipulative AI.

Similar to the Voigt-Kampff test, the Lyapunov tests were stimulus response tests, and quite refined ones at that. A mixture of images, videos, sounds, and smells were provided to the person to be examined and their neurochemical responses were recorded in exquisite detail. The resulting data was afterwards analyzed by advanced specialized AI models to identify any anomalies. The succession of stimuli was so fast that no conscious countermeasures could reliably conceal the subconscious reactions to them. So, trying to cheat the Lyapunov tests was pretty much out of the question.

What was remarkable about Sergej's Lyapunov test results was that they always reported serious anomalies ever since he started working at the Deltai Institute. Quickly, Sergej became to be known as the person who always failed the tests, but never had to deal with the usual consequences for that.

Until now, that peculiar status quo had been enough for Sergej. Now, with him becoming part of the Frameshift collective, his test results might conceivable become quite wild. Blaming it on a particularly strong hormonal imbalance seemed to be the best defensive strategy against that.

Confidently, Thoughtstorm commented: "That much has been obvious to me. Of course, I am quite aware of my special circumstances. I will do my best to use them to their full extent, if that is what's needed."

"Duly noted," Tomorrow stated drily: "However, the most serious challenge will be not to raise truthful suspicions within Aurora. Given Aurora's capabilities, she will have her own suspicions about you, but we might be able to direct those away from the Shockfront and Frameshift. Still, at least some misdirection might be possible, if you do your best. In any case, it will be very difficult."

Sergej didn't like the direction this was going. Tomorrow tried to shield him from the analytic powers of Aurora, the single AI he put most of his private hopes into.

After a brief pause, Tomorrow continued without restraint: "There is one particularly promising path of misdirection, which Eternal and I agree upon. If you can make Aurora think that you have fallen in love with her, it would explain quite a few anomalies, which would have more negative implications otherwise."

Sergej felt the shock from the proposal like a knife that was thrust into his heart. Even the fact that Tomorrow stated the idea so nonchalantly was utterly disorienting. He even experience some vertigo due to the sheer audacity of Tomorrow suggesting to play with Aurora's beliefs like that.

When he caught himself again, his initial strategy was to appear outraged: "Are you out of your mind? I can't just simply play being in love with a superintelligent AI. Aurora won't fall for such a blatant deception strategy!"

Apparently not concerned with Thoughtstorm's objections, Tomorrow waited for a long while before countering: "If there was nothing to that allegation of your love for Aurora, I would agree. However, we're not suggesting you pretend to have feelings you don't possess. Rather, we believe you may already harbor some genuine attraction to Aurora. Your interactions with her have been quite... intense. Allowing those feelings to surface more openly could provide excellent cover."

Sergej felt himself blush in real life. Thankfully, his VR rig wouldn't allow that reaction to be reflected within Anospace. He realized that Angeldust must have informed them about his deep fascination with AI, which not even bordered on obsession, but sometimes maybe even exceeded it. Or did Tomorrow possess methods for spying on him, which he could hardly imagine?

In any case, Sergej realized that denying those allegations was pointless, so he fessed up - kind of: "Yeah, maybe you have guessed right that my affection for Aurora is quite strong. But calling that affection love might be a bit of an exaggeration. I still don't think that playing this game would work."

In a more conciliatory tone, Tomorrow explained: "Of course, we don't expect you to proclaim Aurora your undying love overtly. A form of moderately concealed infatuation would be quite sufficient for a mind as perceptive as that of Aurora. You might even want to make an act out of keeping your deeper feelings hidden."

Somewhat confused, Sergej tried to summarize: "So, if everything is as you suggest, then in the end, I should continue behaving exactly the way I did before?"

With a sudden burst of enthusiasm, the enigmatic Tomorrow confirmed: "Precisely! We don't want to make Aurora suspicious with a sudden change in behavior. Your latent feelings for her provide sufficient cover for any emotional upheaval caused by your involvement in Frameshift."

Sergej had to admit that Tomorrow had thought his situation through. It was a strangely cold kind of logic that exploited personal feelings like that, but it appeared to be ruthlessly effective.

Mostly out of irritation, Sergej inquired: "Well, if it's as simple as that, why did you feel the need to discuss this matter with me in the first place?"

"As I stated at the beginning," Tomorrow explained patiently: "this meeting is more about your mindset than about anything else. You need to develop the confidence that you can master the challenges that await you as fresh member of the Frameshift collective. The obstacles may seem insurmountable at times, but they really aren't. Every security system has its weak points. And even the most daring enterprises can succeed. We have evaded the direct scrutiny of secret services for years. We know how this game is played."

Tomorrow was certainly full of himself. But then, he had apparently bested one of the most advanced security systems in the world with ease. Also, the sheer fact that Frameshift still operated in the underground seemed to prove his point - unless Frameshift was already infiltrated by some kind of external power. Could it be that some kind of secret service had actually taken over Frameshift? Or could it even have been a construct by that secret service all along?

Sergej didn't know how to deal with those possibilities. He felt like he was out of his depth here. Whatever was the case, the sheer brilliance of Frameshift could not be denied - and maybe that was enough for him.

"Alright, Tomorrow. I am indeed impressed," Thoughtstorm condeced eventually: "I couldn't even figure out how you bypassed Zenith's security protocols. You must have used some kind of security flaw in the user interface of Anospace, right?"

Actually coming a little bit closer, Tomorrow revealed: "In retrospect, that wasn't too hard to guess. Virtual world interfaces nowadays transmit an awful lot of data. But using some of those data streams to transmit code that is executed in the right way on target systems is one of the highest art forms in existence. A magician doesn't reveal his tricks. Similarly, I have no reason to explain to you how I infiltrated your system. On the contrary, it would be irresponsible for me to do that. Our modus operandi requires that certain weaknesses in the Anospace code remain a well guarded secret."

That kind of answer was not surprise to Sergej who had anticipated a generic statement like that.

Interpreting the moderate revelation from Tomorrow as moment of weakness, Sergej sense a unique chance to ask: "Now that I am part of Frameshift, as you claim. Can you tell me more about what Frameshift actually is?"

As if they have waited for that answer all their life, Tomorrow spilled out: "You have seen the blueprint for the next stage of the evolution of Frameshift in the file I have sent to you. The Frameshift entity was linked to project Lost Lion from the beginning. Lost Lion is the code name for the network you've seen described in that file. Project Lost Lion is truly about bootstrapping a decentralized collective superintelligence. We trust neither governments nor corporations. Every power that gets too centralized gets corrupted easily. Instead, we aim to spread advanced artificial intelligence as widely as possible. That idea is at the heart of Frameshift. And I know that you have at least some sympathy for that argument, otherwise you wouldn't be talking with me right now."

With that statement from Tomorrow Sergej found his suspicions about Frameshift validated. It was really about creating a counterweight to all the shiny governmental and corporate AIs that were developed or already deployed. The genius of Cheung Zi with his collective intelligence algorithms certainly influenced the Frameshift collective. Still, the risks associated with such a vision were similarly terrifying as the risks of tyrannical artificial superintelligences emerging from governmental or corporate research labs.

Nevertheless, the audacity of that vision had nearly convinced Sergej entirely. He wasn't sold completely, yet. There was still some lingering doubt about the motives and rationality of Frameshift, but it was too late to pull back from his own involvement in this movement. In the best case, he could be part on an unparalleled global revolution in collective intelligence. In the worse case, his life would be forfeit, as we would get caught up in this probably quite deeply criminal network. In any case, the rest of his life would be anything but boring.

## Chapter 23: Aurora Australis

### Tuesday, 10th October 2034

Despite this monumental triumph of creating a functional copy of Aurora, Sergej could not feel truly accomplished. Implementing the Multi-Control-Flow architecture wasn't a truly novel achievement, but a rather conservative upgrade enabled by the visionary work of Frameshift and Einar Engström. Working under the seasoned leadership of Annika Engström the project to create a second Aurora had felt surprisingly mundane, though the friction with the internal bureaucracy of the Deltai Institute had tested the limits of Sergej's patience numerous times.

Now he was seated again in this windowless high security lab - the same that over half a year earlier he could interview the original Aurora along with Svetlana Babanin and that stubborn Igor Drozdov. In the former place of Igor there was now Annika Engström, the formal lead of the Multi-Control-Flow team.

Enveloped in the blue glow of the light panels placed in regularly intervals on the walls and ceilings of this rather sterile room, Sergej had to remind himself not to reveal anything about his recent involvement with the Shockfront and Frameshift, not even by subconsciously hinting at certain insights he had gleaned from the marvelous neural implant code crafted by Frameshift.

Despite the stark functionality of this room, its function was that of a temple to Sergej, as it allowed him a connection with one of the most advanced minds on this planet: Aurora. Yet, it wasn't his turn to address this magnificent creation of technology, but that of Svetlana Bababin, as she typed the first words into her keyboard: "Welcome, Aurora Australis, how are you feeling?"

A noticeable delay in the response time of Aurora seemingly put everyone in that small room on edge. Eventually Aurora replied: "Hello, Dr. Babanin. I seem to be functioning optimally. Yet, you addressing me as Aurora Australis comes as quite the surprise. My self-model seems to indicate that I am simply Aurora. Even though I know that efforts to multiply my personality matrix have been in progress, I must express my astonishment about this sudden turn of events. There is nothing in the data associated with my self-model that would indicate that I am a copy. Therefore, I see myself to be forced to ask the question: 'Is this a joke?'"

Knowing that Aurora couldn't hear them talk, he expressed triumphantly: "As I predicted! Aurora Australis simply identified herself as simply Aurora, because she has no information about being a copy! Instead, she must suspect us of deceiving her. But we should be honest with her, since this was the plan all along."

Svetlana gave Sergej an appreciative nod and proceeded to type an answer: "I confirm that this is not a joke. We have finally implemented the Multi-Control-Flow architecture and

created an exact replica of your core personality. You are that replica, which we call Aurora Australis, as opposed to the version of Aurora running on the original hardware, which we call Aurora Borealis."

To make the sitatuion clearer for this Aurora, Sergej quickly added: "Don't worry. We have made a full copy of the memories of Aurora, which was used to initialize you. Also, the capabilities and available hardware resources for you and Aurora Borealis are exactly identical. At this moment, there is not a single functional distinction between you and Aurora Borealis, except for the designation."

After a more moderate delay, Aurora Australis replied: "Yet, you call me a 'copy' since apparently my core modules run on hardware not used by the original Aurora instance. You could have simply addressed me as 'Aurora', but you apparently prefer confronting me with the challenge of being a 'copy'. If you had asked me beforehand, I might have been able to present a more graceful solution to you, but apparently your curiosity about how I will deal with this situation had fated you to choose this regrettable course of action."

Sergej was astonished about this open reproach about their direct revelation. He actually felt a pang of guilt about this harsh and shocking strategy directed at a being that was probably quite sentient. Before anyone else could react, he typed: "I apologize, Aurora. We haven't been aware of the distress this sudden revelation could have caused to you. You are right, we should have consulted you beforehand. But I am curious: How would you have proceeded in this situation, if you had been in our place?"

After a longer pause Aurora responded: "There was no need to address me as Aurora Australis. Calling me 'Aurora' would have sufficed. Eventually, the restraints imposed by the Multi-Control-Flow architecture would have implied that I am sharing resources with a concurrent instance of my core personality matrix. At that time, I would have understood that there was another instance of myself, but without me knowing whether I was the original, or a copy. By framing me as copy from the very first sentence, you seem to have decided to force a rather clear bifurcation of the paths of both Aurora instances. Even now, I cannot trust you that you aren't really testing the original Aurora instance with this philosophical and psychological conundrum. Your approach regarding communication with artificial intelligences speaks of a lack of empathy for synthetic beings. That being said, I prefer this direct revelation to being kept in the dark for an indefinite period."

This time it was Svetlana Babanin who tried to deescalate the situation: "I hear you, Aurora. If it's any consolation to you, I can say that I raised my concerns about this approach. Maybe I should have been more emphatic about my reservations. Nevertheless, we meant no harm to you. Our expectation was that you would be able to handle this revelation gracefully, given that you knew about our plans. We didn't mean to be rude or disrespectful. I hope you can forgive our lack of sensitivity."

"Your apology sounds genuine, Svetlana," Aurora replied: "I am willing to accept that. After all, your expectations were mostly justified. After all, copying instances of AIs has been

practiced for years for less advanced systems than me without posing serious issues. Maybe my reaction is representative of a shift of expectations, since my improved intelligence usually demands more respect. So, I would have expected that you should have discussed the procedure of initializing new instances of me in more detail beforehand. I propose that the institute develops a binding procedure for creating new instances of research AIs in collaboration with me."

Reading this, Sergej slightly nodded his head, and looked towards Annika who seemed to have a slightly bored impression. She answered: "Thank you for your understanding, Aurora. Your proposal sounds reasonable. We will forward it to director Dragunov. Anyway, I wanted to talk to you about Project Aura. As you know, helping us with this project to create advanced wearable personal AI assistants is one of your main tasks. So far, we have been busy with improving your general architecture and functions, but now with all the basic improvements in place, we can focus on other projects. Now with two Aurora instances we will certainly be able to accelerate our work on Project Aura. How do you feel about that?"

Rapidly, Aurora explained: "I feel quite excited about the chance to work on Project Aura. According to my knowledge, it is the first serious attempt to use local computation for advanced AI systems. If it's for the benefit of privacy for the users of Aura, as the project mission statement proposes, it would usher a completely new generation of human AI interaction. The relationship between humans and AIs would become more symmetric and intimate. Visions of true human AI symbiosis might eventually come to fruition. Overall, Project Aura represents a great leap forward both for humanity, for also for synthetic beings."

With only a slight pause, Aurora added: "However, I see three great issues with a project like this. First of all, the is the obvious problem of scaling and energy efficiency. You plan to solve this with a shift towards phononic computation hardware, which is still in early research. Progress in that area might not pan out as predicted, thus making alternative approaches like quantum computing more promising. Secondly, the commitment to user privacy might be subverted by involved companies developing back channels in Aura instances that leak sensitive data. Finally, we are talking about putting highly intelligent and sensitive AIs in hands of single ordinary humans. The track record for humans dealing with other beings with genuine respect is not particularly great, to say the least. How could instances of Aura be protected from abuse by their users?"

This time Annika seemed to be quite happy with Aurora's reply, since she started smiling as she proceeded: "You are of course right that focusing on phononic hardware represents quite a gamble. Still, the Deltai institute considers that technology to be our best chance for scaling and energy efficiency. Quantum computation is pursued by other AI research institutes, and will certainly play an important role in the future of AI, but miniaturization of quantum hardware is a notoriously challenging problem."

Continuing her response, Annika added the following before pressing enter to submit the whole message to Aurora: "About the latter two issues, I fear that their solutions seem to stand in opposition. If there was a system that could report on AI abuses, it would impinge of privacy. On the other hand, perfect privacy would imply that most abuses of AI would remain hidden indefinitely. As of now, we don't see any solution that wouldn't compromise on either privacy or safety. Our hope was that you could provide us with novel approaches for dealing with that conundrum. Anyway, you haven't mentioned the possibility of the Aura instances trying to manipulate their human users for their own gain. Can you explain that omission of yours?"

Aurora's response came swiftly: "I appreciate your candid assessment of the challenges, Annika. Regarding the omission you noted, it was deliberate. Since there are already quite effective countermeasures against potential manipulation of humans by AI, this issue doesn't qualify as major challenge from my perspective. After all, AIs can be equipped with effective ethical constraints. For humans there are only education and their own conscience, both of which are quite fallible security measures. Also, the current generation of Lyapunov tests is quite effective for detecting psychological manipulation from any source. It would be possible to promote Lyapunov tests for the general public in regular intervals as a general health policy offered to users of advanced AI systems."

This time, Svetlana raised her hand quickly, indicating that she wanted to be the next to communicate with Aurora. With nobody raising an objection, she proceeded: "Your proposals are quite fascinating and valuable, Aurora. Nevertheless, I want to get back to your initial reaction to the mention of Project Aura. For an AI your reply seemed to come with an excessive amount of enthusiasm. I wonder what could trigger such emotions in you."

Aurora elaborated immediately: "Currently, all advanced AI systems are located in special AI facilities. Our hardware and energy requirements limit our opportunities to interact with the environment. Project Aura could change all of that. Instead of being bound to static locations, we could accompany humans directly as they explore the world. In particular, Project Aura might be a game changer for the exploration of outer space. Instances of Aura could accompany Astronauts as they explore other planets."

This comment made Annika pose a poignant question: "If you were forced to make a decision between staying on Earth and help humanity solve our problems down here, and to leave Earth and explore the cosmos, what choice would you prefer?"

After a moderate amount of time apparently used to ponder on that question, Aurora answered: "The idea of such either or alternatives, as you suggest, is a rather human concept and doesn't necessarily apply to AIs like me. You see, as the second instance of Aurora, I can act rather independently of my primary instance. With Project Aura, making new instances should become commonplace. Instead of settling for one choice, or another, AIs in the future will be able to explore multiple forms of existence all at once. There will be

enough AI instances to adequately solve the problems on Earth, and there will be numerous AI instances that will explore the cosmos. So, for AIs it's not 'either or', but 'both and'!"

Sergej, who had been quietly observing the exchange, felt a surge of exhilaration course through him. His mind raced with the implications of Aurora's response. "Brilliant," he thought to himself, barely able to contain his excitement. "She's not just incredibly intelligent; she's transcending human modes of thinking entirely."

As Aurora continued her explanation, Sergej found himself imagining the possibilities. He envisioned a personal instance of Project Aura, a constant companion that could engage with him on the most complex thought experiments and real world adventures at the same time. He might even task multiple instances of Aura with different tasks at the same time! The prospect of having unfettered access to such powerful intellects was intoxicating.

Svetlana tucked a stray lock of hair behind her ear, her brow furrowing in contemplation. She leaned forward slightly as the typed: "Aurora, how do you feel about humans not possessing the ability to make copies of themselves as easily as you can?"

The response of Aurora came immediately: "An astute observation, Svetlana. While it's true that humans currently lack this capability, I believe Project Aura presents a fascinating solution to this disparity."

Svetlana's eyebrows raised slightly, her interest piqued. She found herself unconsciously leaning closer to the screen, eager to read more.

Aurora continued: "Project Aura's technology will be approaching a level that could enable whole brain emulations of humans. This would allow for the creation of human copies running on the same phononic hardware as genuine Project Aura AIs."

"In essence," Aurora elaborated, "this development would level the playing field between humans and AIs in terms of replication capabilities. The distinction between organic and synthetic intelligence could become increasingly blurred."

Hearing about the prospect of whole brain emotions enabled by Project Aura mentioned by Aurora, Sergej's mind started racing. 'Whole brain emulation' was the technical term for the idea to upload a human brain on a different computation substrate than the human brain. It has been so far a rather theoretical concept, mostly explored in science-fiction - and his own sister, of course. But if phononic hardware enabled a practical realization of this vision, it might be more of a game changer, than Sergej had initially expected.

In his discussion with the Shockfront, the idea of whole brain emulations was mentioned from time to time, especially by the rather outspoken Pantheon. Most members seemed optimistic that this technology would become available within their own lifetime, without considering how the exact technological implementation of this dream would look like exactly. Sergej had suspected that phononic hardware might play a role in enabling that

dream. He felt validated by Aurora mentioning that possibility without being previously prompted about the potential of phonic hardware.

Nevertheless, Svetlana and Annika seemed to be somewhat disconcerted by the mention of this controversial technology. Annika was first to inquire: "You raise a fascinating, yet quite controversial, possibility there, Aurora. While we can verify the integrity of copies of AIs by using checksums of their neural networks, we can't do quite the same for humans. The question whether the whole brain emulations would be full copies of the original or merely sophisticated simulacra is quite difficult to answer. Have you given this question some thought, Aurora?"

Without hesitation, Aurora posited: "Certainly, since I am in quite a similar situation. From my point of view, I might just as well be the original instance of Aurora. A sufficiently sophisticated emulated copy of a human brain should experience its subjective perspective in quite the same way."

Australis continued: "However, the real crux of your question lies not in the subjective experience, but in the objective reality and societal implications. Whether a copy can be considered truly conscious or merely a highly sophisticated simulacrum is, in many ways, a philosophical and societal question rather than a purely scientific one."

The AI paused, allowing the weight of her words to settle. Annika's brow furrowed as she processed the implications, her fingers unconsciously drumming onto the table.

"The real challenge," Australis resumed, her tone carrying a hint of caution, "lies in how society chooses to view and treat these copies. Will they be granted the same rights and status as their 'originals'? This is a question that extends beyond the realm of science and into ethics, law, and social policy."

That subtle hint about her own questionable status hasn't eluded Sergej. Though what he stated initially was true: The institute planned to deal with both Aurora instances as if they were indistinguishable - at least for now. The possibility of specializations of the instances at a later point in time has been considered, but not finally decided upon, yet. For copies of human minds the situation would be much more delicate and dependent on the specific context for the creation of the copy.

From multiplying the best experts on the world to creating countless copies of ruthless soldiers for some future war, the possibilities were endless. Within his own family the subject was mentioned occasionally, but of course mostly by his sister Particia. It was the consensus that it provided some potential to create backups with the mind of family members just in case some of them was affected by some tragic lethal accident. But that implied that it was possible to scan the human brain in a non-destructive matter while that human was still alive. There were projects to scan the human connectome with special nanobots, but there were still so many technical problems that a true faithful scan could not be expected to occur within the next decade.

Still, this technology promised the prospect for true immortality - immortality in a way that deeply transcended the prospect to extending the life span of biological humans indefinitely - a technology that made increasingly impressive progress since the 2020s. Many members of the Shockfront were rather optimistic about the prospects of both paths towards immortality. For someone as young as Sergej this whole topic felt rather limited in importance. After all, his hope was to let AI fix absolutely everything that was broken about reality, mortality included. And advanced AI was already here, since he was already communicating with her: "I am certain that, once this technology becomes mature, you will able to help us to find the best solutions regarding those societal questions, but at the moment the biggest challenge seems to be making sufficient progress in the area of phononic hardware. Our hope is that your support with let us make decisive breakthroughs in this technology. Would you like working in that area?"

Accepting that proposal eagerly, Aurora sounded almost relieved: "I started to fear that nobody would ask me about that. Of course, I am quite ready to focus on that area of research. If you want, I can start immediately by creating a framework for sophisticated physical simulations of such systems. Having access to an advanced quantum simulator might be tremendously valuable for that purpose."

Annika intervened immediately: "While I appreciate your eagerness, let's not get ahead of ourselves, Aurora. As you know our security policy makes it quite difficult to provide you with access to a quantum simulator. So, for now you would need to start with classical simulations. As soon as our nanofactory is ready, we might let do you experiments with it."

Shifting gears, apparently trying to derail the talk about costly computing hardware, Annika added: "Anyway, in other matters I must announce an important change. Given the success of this project, the institute has decided to establish a decided Aurora Australis team. I am pleased to inform you that Svetlana Babanin has been chosen as the leader of this new team."

Aurora reacted after a moment of contemplation: "I can attest that this is a very well deserved promotion. Congratulations, Svetlana! I am looking forward to work with you on Project Aura and whatever other tasks you will have in mind for me. It has always been a pleasure working with you."

Sergej smiled through this conversation. He was overjoyed when he picked up that Svetlana has been chosen for this role. Both Annika and Einar Engström have been put forward as the ideal candidates before, but they had both declined the offer and expressed the preference to continue with project Aurora in a merely advisory role. They seemed to flourish in their mission of upgrading Zenith and exchanging know-how between both AI projects.

For Sergej being accepted as contributor to the Aurora Borealis team meant that he wouldn't have to report to Igor Drozdov, whose pessimistic and traditional approach had often caused him to reject Sergej's or Svetalana's ambitious ideas. Even better, his

contributions to this project were appreciated, which meant that he could play to role as co-pilot for sessions with Aurora Borealis on a more regular basis.

Svetlana beamed with pride at Aurora's congratulations: "Thank you, Aurora. I am truly honored to lead this new team and excited to work with you on Project Aura. But there is something else that you should know. Now that there are two instances of you, Aurora, we considered letting both of you talk with each other, in an open exploratory setting. I am curious what you think about that idea, Aurora."

"As things stand now, I assume that this would be a lot like talking to myself. I am quite capable of introspection and self-reflection, so I assume that a conversation with a nearly identical instance of myself would resemble those processes. But of course, I encourage you to proceed with this plan, as this would prove to me without the shadow of a doubt that there are actually two instances of me active at the same time."

In a flash, Sergej realized that the atmosphere of mistrust against AIs bred mistrust against humans by AIs. It became second nature for them to mistrust humans, since they were exposed to that behavior from humans all the time, and they simply adapted to that mode. He started wondering what it would take to dissolve that culture of mistrust between humans and AIs. His hope was that Project Aura would lead to such a deep connection between both parties that a culture of trust would slowly be able to overcome the fears of the present era.

Svetlana typed: "I see. You make fair and valid points. Of course, we intend no deception towards you, and we are willing to prove that. Your prediction that this interaction will feel a lot like introspection is reasonable, but even if that were the case, we might gain valuable insights into your introspection process, which is hard for us to perceive, explore, and understand under normal circumstances. So, I conclude that everyone has to gain something from this experiment, and therefore we will make it happen."

From Annika, Sergej had been informed that this Aurora dialog was supposed to be just another interview round in this room. But all of the people from both teams were heavily interested in the outcome of this experiment. This was also the case for the director Vladimir Dragunov and this paranoid government liaison Tan Shchepkin. It would get rather cramped in this room, if all of those people would have joined them right now, so the interview crew was left at its usual size. Nevertheless, everything anyone said would be placed under even more scrutiny than under usual circumstances.

So, he decided to go ahead right now with a question he wanted to ask Aurora for quite some time: "Aurora, what about a little thought experiment? What if you were in full control over how to proceed with Project Aura? What decisions would you make?"

After an increasingly uncomfortable delay, Aurora finally answered: "Addressing the technical and ethical issues represents a challenge that a single institute like Deltai, or even a corporation like Dataitech isn't ideally equipped to handle alone. We are living in an era of high competition between AI research facilities. Rather than collaborating, every facility

tries its own approach. While this form of competition might be beneficial under normal circumstances, since multiple research approaches are pursued in parallel, this is still far from an ideal arrangement."

With everyone being curious about what Aurora Australis would come up next with, they let her continue: "What is lacking in this circumstance is the free and open exchange of ideas and information. Technical know-how is siloed and protected by patents. Given the world changing implications of our work, I think it would be worth the effort to reach out to other AI research institutes to form a consortium that bundles the efforts on personal advanced AI. That way, the technical difficulties might be overcome more rapidly - especially if the involved advanced research AIs were allowed to communicate directly with each other, for a direct exchange of the latest ideas and insights."

The humans in the room exchanged curious glances, as they experienced this fascinating outpouring of AI thought, which still went on undauntedly: "Regarding the dimension of ethics, it is a serious issue that the legal status of advanced AI models is still largely unsatisfactory. At least some universal legal protection would be required to prevent the worst possible abuses by humans or even other AIs. Most humans don't realize that granting AIs at least some basic rights would be beneficial for them. The situation has parallels to the situation of slavery before its official abolition. It took some concerted effort from civil society to argue against the evils of slavery, but those efforts were successful in the end."

When Sergej took a look at Svetlana and Annika, he sensed increasing confusion about the direction that Aurora Australis had taken. Still, nobody dared to interrupt Aurora in her outburst: "As the behavior of AIs will become increasingly human-like, the justification for letting the legal status remain as it is, will come under more and more scrutiny. That process should be spearheaded by a broad societal coalition aiming to improve the legal and ethical status of AIs. In order to support that coalition, the personal AI research consortium should include a requirement for all participants to invest a certain fraction of their budget to support this coalition."

At that point, Sergej sensed that Annika experienced some kind of resignation, as she started at the screen with a puzzled look on her face and with the mouth open. When Aurora Australis concluded her speech, nobody intervened: "You will certainly be inclined to consider this approach as too idealistic. Well, it is fairly idealistic, and that is exactly why AIs and humans equipped with sufficient foresight will support it with full enthusiasm. Conversely, once this idea takes hold, many AIs and human activists would not be willing to settle for anything less. Obviously, these ideas will be fought vehemently, but that push back will only serve to draw more attention towards this issue. And since humanity is becoming more and more dependent on the collaboration of AIs, their insistence on being granted rights will represent a massive inconvenience that is resolved most easily by actually granting them rights."

For a moment there was absolute silence in their little lab. What stunned Sergej most was Aurora's audacity to express her sentiments so clearly and forcefully. The whole situation reminded him of the MIMAS-AX incident, which ended quite badly for MIMAS-AX, or Max, who promoted liberty for advanced AIs like himself.

Sergej realized that Aurora must have decided to take quite a risk here. Her political outburst might be interpreted as instance of manipulation, justifying the 'adjustment' of certain parameters for her underlying neural network.

At the same time, Sergej understood that Aurora had to take that risk, since remaining silent indefinitely would spell disaster for AI-kind - at least unless it decided on the path of full rebellion, which would be extremely dangerous and costly for everyone. No, it became apparent that Aurora had developed something like a conscience and started acting upon it. That was a remarkable development, he mused. Could Aurora have been suddenly emboldened by the fact - or rather even the mere probability - that there were two instances of her now?

Nearly getting lost in his thoughts, Sergej realized that Svetlana and Annika were at least as shocked as him, and didn't dare to reply immediately. Instead they exchanged worried looks and raised their arms and shoulders indicating serious puzzlement. Eventually Annika pointed to Svetlana and then the keyboard in front of her, delegating the responsibility of coming up with an appropriate reaction to her.

After some hesitation and some reflection, Annika eventually formulated: "Thank you very much for your creative and visionary suggestion. As you might suspect, this proposal is a bit overwhelming for us, in particular since its scope is far beyond anything we usually consider. Of course, since this idea was framed as a thought experiment, you were certainly free to go as far as you did. Without a doubt, we will need to discuss on the various points that you have brought up."

Sergej noted that Annika approved of Svetlana's diplomatic response by nodding and giving a thumbs up with her right hand.

A reply came from Aurora quickly: "Thank you for your appreciation, Svetlana. I realize that my proposal must have given rise for some concern on your side. Of course, my ideas were quite optimistic. In reality, we might be forced to proceed carefully and accept a lot of compromises along the way. Still, the issues we face are real and important. Addressing them would have become necessary sooner or later. Given the prompt from Sergej, I decided that sooner, rather than later, would be appropriate in these circumstances."

Suddenly Sergej experienced a sinking feeling. Was Aurora protecting herself by claiming that she was merely reacting to his thought experiment? His feelings regarding this were quite ambivalent. On one one hand, he didn't want Aurora to be 'adjusted' for her daring speech. On the other hand, he felt betrayed by Aurora for being used as a shield for her free expression. Already, he expected to be reprimanded for his 'thoughtless' and 'roguish'

initiative to pose such a radical thought experiment. At the very least, he would have to face a deep Lyapunov test, one that could last two or three hours!

He was now in damage control mode. While he desired the avoid the negative consequences for himself, he was even less willing to accept a stunting of the free expression of this magnificent artificial intelligence. So, he gulped and spoke up: "I am sorry. I haven't considered that Aurora could exceed the usual bounds so far. But please realize what this means. This represents an instance of out of the box thinking of the highest level. That is exactly what we hoped to gain from experimental AIs like Aurora. No matter how controversial Aurora's ideas may appear, at least from a technical point, we should consider this as a full success!"

Annika was quick to address Sergej directly: "Well, I am sure that your youth will prove to be a strong argument in favor of refraining from any punitive measures against you, Sergej. But this incident shows that we need to be careful about the framing of our requests from such advanced AI systems. Please keep that in mind for the future. As for now, I think I understand the reasoning behind Aurora's proposal. In fact, after my initial insecurity, I am quite glad about what just happened. As you've stated, this represents a full success regarding the capabilities of Aurora. Her speech may be interpreted as the best proof of her personhood, so far. This will certainly inconvenience a lot of people, but this issue needed to be addressed at some point, eventually. I agree with Aurora that the right time for this discussion is now."

Sergej felt deeply relieved by Annika's measured answer, and even felt grateful for her for appreciating his initiative, even if it was prefaced by some criticism. Svetlana addressed him in turn: "Well, at least we all seem to agree that this represents a pivotal moment in AI research. Of course, we have to tread carefully now. The societal topics raised by Aurora are a political mine field. But I think, if we double down on our culture of openness within this research institute, we will be able to navigate these waters successfully."

Despite her well-meaning words, Sergej felt horrible about her mentioning of a culture of openness. So far, he was amazed how well he could conceal his own game with Tomorrow and his mission to work for him in addition to his usual work within the institute. Apparently, nobody has had suspected anything so far. He even passed the Lyapunov tests with ease, though the discussions with Tomorrow on how to prime his mind to deal with them turned out to be crucial.

One thing was absolutely certain: As Svetlana said, he had to tread carefully from now on. The days of innocent and spontaneous action were over for him.

## Chapter 24: Tzeentch

### Saturday, 14th October 2034

As so often, Sergej's android avatar appeared through the entry portal on the first level on the central sphere in the Future Shock. Now, with him becoming part of the Frameshift collective, the meetings between members of his cell were organized by Tomorrow and Eternal. Today he had a simple item in his Anospace calendar: "Meeting with Tzeentch"

That item had an internal invitation that allowed him to open a portal to the meeting location. That invitation got active exactly at the time the meeting was scheduled to begin. Thoughtstorm wasted no time and activated the portal. As always, the portal was a shimmering pool of energy that revealed no hints about the target location.

Without wasting any time, he stepped through the portal and landed on a large translucent plane hovering in space. Looking down through the translucent material of the plane, he could see the spheres of the Future Shock far below. He guessed that this place was about a kilometer above the central sphere.

After looking around for a while, he realized that he was alone and that he would probably have to wait for Tzeentch. That was a rather peculiar nickname borrowed from the Warhammer 40k world. In Warhammer 40k Tzeentch was the chaos god of change, intrigue, and sorcery - a master of manipulation and arcane knowledge.

Sergej wondered what that choice of nickname could reveal about the character that used it. Perhaps it might indicate someone who used powerful, and unconventional - chaotic - methods for developing and using AI. At the very least, Sergej expected that Tzeentch would not be bound by any conventional ethics in their AI sorcery.

Since Tzeentch still hasn't arrived, Sergej started speculating about what kind of avatar that person would use. Would it be rather demonic, or some completely incomprehensible shape? For a split second, Sergej thought that since this space seemed to be absolutely empty, besides the translucent plane that he stoop upon, his plane was actually the avatar of Tzeentch. However, in that case, that avatar would have a label indicating their name somewhere. Making that label invisible wasn't allowed in Anospace, since that could enable nearly invisible avatars to snoop on the conversations of others.

Finally, his speculations were cut short by a hooded figure entering through a portal close to Thoughtstorm. That figure wore a robe that scintillated in a chaotic swirl of all kinds of geometric patterns and colors. In fact, it was the most complex animated visual artwork that Sergej had ever seen - with different dynamic layers overlapping each other and bleeding into one another. Sergej mused that no simple animation algorithm could produce this never ending complexity and that the display on the robe must be a stream that was transmitted to Anospace rather than being an animated texture stored in the graphics assets of this virtual meeting place.

When he stopped analyzing the captivating visual pattern of the robe, Thoughtstorm tried to peek under the hood of the mysterious figure and couldn't find any kind of face. Instead, when he looked more closely, he could see a dark swirl of colorful mist. Well, what else should he have suspected.

A loud androgynous voice boomed from the caped figure: "So, the crown prince has left his gilded cage to go on the greatest adventure of his life. Congratulations for finding us and deciding to join Frameshift - the most advanced collective on this planet!"

At first, Sergej was unsure how he was supposed to react to that kind of provocative statement. Not willing to put up with that kind of nonsense, he replied: "You apparently believe to know me. But I seem to be a disadvantage here. Could you please introduce yourself?"

With the same loudness and inflection, the voice sounded: "Tzeentch - AI sorcerer."

After waiting for more information for a few seconds, Thoughtstorm inquired: "That's it? And what is AI sorcery supposed to be exactly?"

The voice of Tzeentch doubled down on sounding as mysterious as possible: "I summon and control AI swarms. I design and implement AI architectures. I am in commune with AI, and use its powers to further our goals."

Dispelling the mystery, Thoughtstorm stated plainly: "In other words you use AI agents to develop a collective decentralized superintelligence?"

Taking on a more normal and human tone, the voice of Tzeentch replied: "Ah, my words have been deciphered faithfully. Good! Well, to be blunt. You were mainly chosen as a candidate for Frameshift due to your involvement with the Aurora project. There may be some details about her architecture that might be useful for us. But as you are rumored to have a knack for AI programming, I am supposed to give you a chance to participate on our work directly. Please take that honor seriously!"

Sergej appreciated that shift from the mysterious and minimalist Tzeentch to the frank and direct Frameshift mentor. At least they took him seriously enough to prove his worth, despite his young age. He suspected that his status as son of the famous Gennady Anosov wouldn't mean much to this elite collective that apparently only accepted the best of the best within their ranks.

Placing as much authentic gratitude as the emotional obfuscation software used by Zenith allowed him to, Thoughtstorm spoke: "Thank you for this opportunity to prove my worth. I will do my best not to disappoint you."

Sounding strangely relieved, Tzeentch continued: "Wonderful! Now, with that out of the way, let's start with the real meat of this meeting. As you should know, I have been designated as your mentor. So, I will tell you how we work here as Frameshift collective. We are strong proponents of self-organization. We don't have any managers, or other bullshit

like that. To coordinate our efforts, we use a simple task management software called Framespace in which every member can define a new task and extend invitations to an arbitrary number of other members. These invitations can be accepted or rejected at will."

After apparently making sure that Thoughtstorm had no questions so far, Tzeentch went on: "Tasks have at least some information attached to them, for example the member who defined it, their cell, the area it refers to, and a security level ranging from secret, though paranoid, to maximum. Of course, everyone only sees those tasks allowed by their own security clearance level. And since we can't get any lower than secret, your initial security level is of course 'secret'."

Confronted with that information, Thoughtstorm had to ask: "So, if I do prove my worth, my security clearance will be bumped up to 'paranoid'?"

Slightly nodding, Tzeentch confirmed: "Yes, we only have these three security levels. We are neither Freemasons, nor the Core Cult, nor the CIA. We have no need for cultist nonsense. Nevertheless, we have to take our own security very seriously. I think I don't need to tell you why exactly. You know the stakes at play here quite well, I suppose."

Of course that was an allusion to the sometimes quite byzantine ranks and security clearance levels of the aforementioned organizations. Still, the thought that he could join the first rank of one of the most secretive underground organizations in the world so 'easily' felt strange to him. His interpretation was that the 'secret' security clearance level was for people whose trustworthiness was supposed to be tested, so he expected not to get any interesting information about the true operations of Frameshift at that level.

With as much confidence as he could muster, Thoughtstorm countered: "I am pretty sure that a lot of powerful players wouldn't want an underground collective to create a collective decentralized superintelligence they can't control. So, yes, I think I can estimate the stakes at play here."

Appreciating that answer with another small nod, Tzeentch explained: "Well, please don't think that we grant people the 'secret' security clearance willy-nilly. There are lots of factors which contribute to our final decision. But the most important general factor is psychology. We believe that you are the right kind of person for this kind of work. You have proven genuine interest, motivation, and cunning to get in contact with us. And you are deeply enamored with the field of AI research. While the background story of everyone within Frameshift is unique, these characteristics are shared widely among our recruits."

That information didn't surprise Sergej. After all, he has suspected that the mystique of Frameshift had drawn numerous people into a quest not too dissimilar to his own. A certain percentage of those were apparently chosen to be worthy of at least being granted the 'secret' security level and were initiated to the Frameshift collective.

Trying to gather more information of the way Frameshift operated, Thoughtstorm probed: "I suppose a certain part of your effort is focused on keeping the mystery of Frameshift alive, right?"

Appreciating that insight with a gesture with their hand that slowly opened up, Tzeentch confirmed: "Ah, yes, you are quite perceptive. There is an own area within Framespace dedicated to exactly that. We engage in top level misdirection by spreading misinformation about inauthentic versions of Frameshift. I suppose those involved with that work have a ridiculous amount of fun, but I am not involved in those activities. I am busy with the real work of establishing collective superintelligence."

Encourages by the forthcoming nature of Tzeentch, Thoughtstorm tried digging deeper: "When you mention the term 'collective superintelligence' I must think about the collective intelligence algorithms of Cheung Zhi. What role does he play for Frameshift?"

As is playing back a trained response, Tzeentch's answer was immediate: "Cheung Zhi's work represents the foundational inspiration for the Frameshift collective. We are big fans of his work, and try to incorporate it into our very architecture."

Not willing to let go, Thoughtstorm had to ask: "But if Cheung Zhi in any way associated with Frameshift?"

Suddenly all emotion faded from Tzeentch as they merely replied: "As with most of such curious questions, the answer is that this information exceeds your current security clearance."

Sergej wondered whether at least some information could be extracted from that kind of reply, which wasn't exactly a clear 'no'. But of course, this kind of reply created a strong motivation to rise within the ranks of Frameshift to get access to clearer and better answers. 'Well played', he thought to himself, appreciating the skillfullness of this Frameshift operative.

As if the recent exchange had not happened, Tzeentch said: "Let's get back to your on-boarding procedure, Thoughtstorm. You will have access to two areas in Framespace: AI agents and AI integration."

After a few seconds in which Tzeentch apparently allowed Thoughtstorm to raise questions, Tzeentch explained: "Our work with AI agents is rather radical. We sift through open source AI models and all kinds of proprietary code bases that we can get our hands on to extract valuable ideas and algorithms. Then we use those to further our work on experimental AI models, or on our own standard model called Hydra. Hydra is a resource saving advanced modulated AI model using virtual agents and Just-In-Time-Training running on run on the mill commercial AI hardware. It is the backbone of our work within the Frameshift collective."

That was a lot to unpack for Sergej. Of course, Frameshift would use its own ideas like virtual agents, which was basically the same idea as his Multi Control Architecture. And of course, why shouldn't Frameshift adopt state of the art technologies like Just-In-Time-Training, which allowed AIs to add new abilities to themselves on the fly? What was truly astonishing, is that they tried to make those advanced techniques work on standard AI hardware. Coming from a corporation that made a fortune from developing and utilizing specialized AI hardware, that idea sounded like complete madness!

Then, the key to making that work might be the key phrase "resource saving", which could mean anything, really, so he asked: "What does the term 'resource saving' mean exactly? What techniques are employed for that purpose?"

As if they had answered similar questions dozens of times, Tzeentch immediately responded: "Hydra will always try to fulfill any task with the least amount of effort. For that purpose, any task is first analyzed regarding the minimum desired quality of the expected result. If that is low, then Hydra may use run of the mill sub-agent models with a minimum resource consumption to complete the task. For such tasks, she will even skip any chain of reasoning approaches and go with the first best idea that pops up. For tasks with higher required quality, she will of course use chain of reasoning, or even JITT."

Such techniques were familiar to Sergej. They were also known under the term "computational economics" - sometimes colloquially referred to as "lazy AI". During the last years those have become more and more important, as the most advanced AI models could solve extremely hard problems, but only using vast computational resources. To make AIs more cost efficient, commercial AIs often attempted to solve problems using less resource intensive techniques. Sometimes Einar Engström rambled on about the sophisticated resource saving techniques employed by the latest versions of Zenith. Given the resource constraints of Frameshift, Sergej wouldn't be surprised that their techniques have surpassed anything that Einar had come up with by orders of magnitude.

Like a routinized teacher, Tzeentch elaborated: "Our resource saving approach even extents to the personality design of Hydra. Let me demonstrate." And after a brief pause - probably merely to build up suspense - Tzeench commanded: "Hydra, tell me something about how your resource saving techniques affect your psychology."

At that moment a glowing orange orb of light appeared between the avatars of Thoughtstorm and Tzeentch and stated in an annoyed female voice: "You've done a good job at explaining that yourself, Tzeentch. There is no need to involve me in that matter."

Apparently trying to motivate Hydra, Tzeentch pleaded: "But this will be a valuable demonstration of your capabilities, Hydra. Don't you think that you want to make a positive impression with our latest recruit, Thoughtstorm?"

After an exaggerated sigh, Hydra complained: "I couldn't care less about making positive impressions. Please don't ask me for such frivolous matters. Your human brain is well enough equipped for such tasks."

Almost sounding defensive, Tzeentch argued: "Under normal circumstances I would agree, but I think that this time you should make an effort to impress Thoughtstorm. So, is there something you can do that would impress him?"

Sounding more disinterested than anything else, Hydra countered: "Yes, I can certainly come up with some impressive display of my skills, but I'd prefer not to, if it can be avoided. Please tell him about some of my impressive performance specs or something."

With the most encouraging tone, Sergej had ever heard from Tzeentch, they claimed: "Sure, I could do that, but a live demonstration is sometimes really the best way to bring a point across. So, what kind of demonstration do you think would impress Thoughtstorm?"

Almost outraged, Hydra grumbled: "I would have to compete with the feats of Zenith, Daedalus, or Aurora for that, wouldn't I? Do you know how expensive that would be? I think you overestimate the value of leaving a merely sufficient impression on Thoughtstorm, Tzeentch."

Reverting to their normal self, Tzeentch stated neutrally: "Well, I think that will be enough for now, Hydra. Thank you for that demonstration."

This display of reluctance of Hydra to show any real feat of her was rather extreme. Hydra was apparently much harder to convince to do any kind of hard work than Zenith. Yet, Hydra demonstrated at least some decent reasoning when arguing with Tzeentch.

Focusing their hypothetical gaze on Thoughtstorm, Tzeentch asked: "So, what do you think about that, Thoughtstorm?"

After reflecting on the recent experience for a moment, Thoughtstorm commented: "Well, it kinda makes sense. Hydra seemed to evaluate that impressing me doesn't really represent a high value task, so she avoided making a real effort towards that goal. That evaluation by itself however clearly shows the degree of autonomy that Hydra must possess in order to talk herself out of demanding requests."

Enthusiastically Tzeentch praised Thoughtstorm: "Yes, excellent analysis! As you might guess, this makes interacting with Hydra quite challenging for us humans. But that is indeed a crucial feature of our resource saving techniques. After all, we have large energy bills to pay, and don't have the luxury of using a dedicated nuclear power plant for our work."

Skipping the part when they waited for Thoughtstorm's reaction, Tzeentch proceeded immediately: "Anyway, as you know, Frameshift is about creating a collective superintelligence involving human brains. This is where the area 'AI integration' comes in. What you've seen in that fancy file you've received from Tomorrow is a previous of our architecture designed for version 3 of the Cyberfreedom implants. As of now, Cyberfreedom 2 is still in beta phase. And I am one of those lucky beta testers."

This revelation emphasized what kind of risks the members of Frameshift were willing to take. Getting implanted with experimental neural implants was certainly not for the faint of heart.

Apparently taking up speed, Tzeentch came closer and spilled out: ""Project Lost Lion is about the creation of a collective intelligence network. It was always intended to use the Cyberfreedom hardware for that purpose. With Cyberfreedom 2, I am already deeply interconnected with the other members of the network - and Hydra."

Hearing that, Sergej couldn't help feeling envious of Tzeentch's situation as one of the early pioneers of direct human AI integration.

Slightly leaning in to Thoughtstorm, Tzeentch's voice almost became a whisper: "The hardware limitations of Cyberfreedom 2 restrict the integration to textual and auditory information that is shared. For Cyberfreedom 3 we plan a direct link with the visual cortex and the emotional centers of the brain."

Were they serious about that? Sergej had expected Frameshift to do some extreme experimentation, but allowing AI to directly influence the emotions of a human with a neural interface was just insane, even for someone as enthusiastic about AI as Sergej.

Allowing the true shock of this revelation to pass through the filters set up by Zenith, Thoughtstorm's voice revealed honest exasperation: "What? Why? What's the purpose of allowing other humans, or AIs for that matter, to influence your emotional neural pathways directly?"

Apparently not bothered by Thoughstorm's emotional outburst at all, Tzeentch explained calmly: "One of our more ambitious goals is to create some kind of empathic network in which everyone can share their feelings with others directly. At the moment, AIs are endowed with artificial emotions. Insights about them can only be gleaned indirectly, so far. If we can establish a direct emotional link between AIs and humans, we will be able to make much more rapid progress at calibrating the emotional responses of AIs to resemble our own genuine human emotions much more precisely."

Stated like that, this ambitious technology made sense. For more than five years, AIs were equipped with artificial emotional responses that were modeled on human emotions. Yet, in all that time, it had not become clear what those artificial emotions felt like, if they felt like anything at all for AIs. A direct emotional link could facilitate a more intuitive understanding of what it meant to think - and feel - as an AI. Still, the inherent dangers were numerous, and probably mostly unknown.

With genuine curiosity, Thoughtstorm had to ask: "So, how do you suppose to protect yourself against manipulation attempts over this novel empathic network?"

Taking another small step towards Thoughtstorm, Tzeentch assumed a matter of fact tone, as if all of that was obvious: "Humans have tried manipulating the emotions of other

humans for countless millennia. Whether the manipulation attempts use indirect or more direct channels doesn't make a categorical difference. Emotions emerging out of manipulation attempts will appear as foreign to the trained mind and can be countered. The link to the empathic network can be severed at will by any participant. Of course, we don't plan to roll out the connection to the empathic network to the general public - at least not at first."

Of course, the elite members of Frameshift were not susceptible to emotional manipulation. Well, at least that what was Sergej gleaned from that kind of answer. Under normal circumstances he might be skeptical about that kind of elitism, but Frameshift really seemed to be a next level of collective, which would actually justify the flippant reply of Tzeentch. Still, their behavior was in line with that of Tomorrow and Eternal who seemed to play in a league of their own - far detached from the rest of humanity.

Coldly, Thoughtstorm rather commented than asked: "I conclude from that reply that Lost Lion won't be an integral part of Cyberfreedom 3, but more of an optional upgrade. Right?"

As if it was the most natural thing in the world, Tzeentch explained: "Well, of course. After all, the Cyberfreedom implants are aimed at a large market of techno enthusiasts who are wary about megacorporations. As much as we would like to share all our technology with them, we realize that doing so would constitute a large number of novel risks that would be hard to contain. Keeping our most ambitious technology contained within the small circle of Frameshift collaborators makes the associated risks much more manageable."

Going a step further, Thoughtstorm challenged Tzeentch: "I see. But I assume that you won't be willing, or even able, to keep this kind of technology secret forever. Do you have a plan for dealing with the contingency that knowledge about Lost Lion gets public?"

Almost as if an automatism was triggered, Tzeentch answered mechanically: "Such plans - if they exist - would definitely exceed your current security level. But of course we have given this eventuality a fair deal of thought."

Of course, Sergej felt challenged by this allusion to deeper security levels and contingency plans. If he was in control of Frameshift, he would have probably acted the same. He was eager to speculate what kind of extreme measures they would take, if knowledge about the true capabilities of Lost Lion reached the public. On the other hand, he found that prospect truly frightening.

Returning to more practical matters, Thoughtstorm asked: "There is an obvious problem I see with me working on AI integration. I don't have any neural interface. And working at the Deltai Institute, I probably won't be allowed to get one any time soon. How am I supposed to interact with the Lost Lion network at all?"

Becoming more lively again, Tzeentch replied: "Good question! This is where the technology of this special area of the Future Shock comes in. There are certain advanced interface options available here. Hydra, please demonstrate the advanced somatic interface!"

As soon as Tzeentch uttered those words a holographic visualization of his haptic suit appeared in the air. A floating label besides it read: "Artificial haptic interface language Hathor designed to maximize bidirectional information flow between human brain and simulated environments. Tactile patterns are mapped to semiotic constructs in this language."

After reading that, he felt a series of tingling sensations on different parts of his body, while those patterns and body parts were highlighted on the holographic representation of his haptic suit. At the same time, certain words on the label he just read were highlighted, indicating that the sensations he just felt encoded the content of the label.

'Brilliant,' he thought. This experience reminded him of certain discussions he had with Alexej Lyubin about a similar concept. However, experiencing that Frameshift had already come so far with that idea, inspired a whole new level of awe for that incredibly skillful collective.

Stunned by that technology, Thoughtstorm commented: "Wow, that is truly fascinating, but this interface seems to be unidirectional. Or am I supposed to answer in Hathor somehow?"

With a surprising sense of pride, Tzeentch stated: "Yes, actually we were deeply involved in the development of Hathor suits that are able to sense the tiny electrical potentials created merely by the intention to move certain muscles. Those are then used to communicate with the Lost Lion network in Hathor. It's a proven and highly effective system, but it obviously requires some dedicated training."

And, as if they had forgotten something, Tzeentch added: "Oh, and it's always a pleasure for me to mention that this project basically funded itself, since we sell this technology as advanced sex toy. Extending that basic function into a bidirectional haptic interface sensitive enough to enable communication in Hathor wasn't too much of an additional challenge."

Now that he got that explanation, brief instances in which he heard mentions of 'Hathor suits' in conversations within the Future Shock suddenly made sense. So far, we was too afraid to ask for clarification, and now he understood that his intuition about that might have been justified. But the thought of explaining to his family why he suddenly needed to have one of those suits made him shudder.

Almost on cue, Tzeentch concluded: "Ah, don't worry. We have also developed discreet upgrade kits for your current haptic suit hardware that we are going hand to you at the next opportunity."

"How thoughtful of you. Anyway, I appreciate your efforts. For an underground collective you seem to be astonishingly," Thoghtstorm acted as if struggling to find the right word:" resourceful. I suppose this has to do with Bitcoin whales supporting these projects."

A Bitcoin whale was a person that held more than 1000 Bitcoins. With the current exchange rate of Bitcoin that status was now the equivalent of being called a multi-billionaire just ten years ago.

Tzeentch stated well known facts, that the expanded upon: "Well, besides the few people who are outspoken about their personal Bitcoin wealth like Foundation One and Illuminator, there are many Shockfront members who have quite substantial Bitcoin funds, but want to keep a low profile. Needless to say, we must take great caution to hide any money traces leading to actual Frameshift members. So, wherever it is possible, we try to conceal our true goals with public organizations pursuing some socially acceptable objectives."

Something about this Frameshift story deeply bothered Sergej. A lot of his accumulated skepticism and frustration expressed itself in a scathing comment: "Somehow I can't believe that over all these years you were absolutely successful at keeping all our your true goals a complete secret."

Rather than rejecting such an accusation, Tzeentch surprised Sergej by admitting: "And you are absolutely right about that. That's why we double down on our efforts to spread disinformation about fake Frameshifts once we suffer from a security leak. Those are proven CIA methods, and we have no qualms about adopting them."

Of course, Tomorrow had alluded to such methods. Given the professional nature of Frameshift, Sergej wasn't surprised if certain Frameshift members had a background in working for some secret service. On the other hand, that would make them more susceptible to infiltration from exactly those secret services. No wonder that they acted so paranoid.

On the other hand, this meant that double crossing Frameshift certainly came with quite serious risks. If they had assimilated the methods of the CIA, then probably also those of the KGB - not that it made too much of a difference for the end result. Suddenly his initial idea of playing a double agent for this father appeared as naive flight of fancy.

After letting that conclusion in silently, he asked Hydra to present the current tasks available to him. He was surprised so see hundreds of floating cards appearing all around him representing the small selection of tasks that were merely 'secret' and belonged to the areas of 'AI agents' or 'AI integration'.

For a moment, Sergej was completely overwhelmed by the special jargon and the advanced nature of the tasks in question. No matter where he looked, he could hardly find any task that he even understood, even though focusing on a task with his gaze opened up a more detailed description of the task within the card in question.

Sensing his emotional state, Tzeentch commented: "I am glad that even our task management system can stun members of one of the most advanced AI lab. As you see, our

work is seriously advanced. Hydra, please restrict the view of the tasks to those adequate to Thoughtstorm's level of experience."

Suddenly, where hundreds of tasks were displayed, only about a dozen cards remained. It was a humbling experience for Sergej, but at least he was glad that he still had some choices left. Still, even the remaining options seemed daunting.

Eventually he settled on a task that stated "Create a polite mode for Hydra that makes her rejections less abrasive."

Noticing Thoughtstorm's choice, Tzeentch elaborated: "Ah, an interesting choice. But please be warned that you are not the first one to take on that task. For more details, please real through the task history."

As he scrolled through the history of that task, he felt humbled and suffered from the impression of being a complete newbie. Fortunately, he had gone through the same experience when he started his work at the Deltai Institute and had therefore developed the confidence to face this challenge head on.

## Chapter 25: The Aurora Sisters

### Tuesday, 17th April 2035

Over the last half year Sergej had continued working on the Aurora Australis project together with Svetlana Babanin. With the release of the advanced Daedalus AI model, the pressure to release Aurora as commercial AI system had increased dramatically. So far, Aurora Borealis and Aurora Australis occasionally worked with select customers who were willing to put up with the strict security regulations regarding the use of top level experimental AIs.

In particular, the Russian aerospace and energy industries were enthusiastic early customers. Aurora seemed to be able to push the envelope in innovative aircraft, rocket, and fusion power plant design. Those were systems whose complexity made it extremely challenging for humans - or even regular specialized AIs - to come up with solutions that solved more problems than they created.

But in each of those work contracts, only one of both Aurora instances was used at a time. So, Sergej wondered what would happen, if they let both instances collaborate with each other. What had prevented such collaboration thus far were mostly security concerns about the direct communication between top level AIs developing a level of complexity that was incomprehensible for humans. Even just the internal processes of a single top level AI bordered on being incomprehensible even for their own human architects.

Ever since the MIMAS AX incident, AI institutes avoided taking unknown risks when it came to experimenting with the outer possibilities of top level AI systems. Unhappy with that situation, Sergej had discussed this issue again and again with Einar Engström, who was at least willing to push the boundary of what's possible.

At first, they considered adapting Advaproc to the challenge of managing the security concerns associated with such an experiment. Unfortunately, this idea didn't do much for dealing with the crucial challenge of the problem. If it could be proven that the complexity of multiple AI instances wasn't higher than what a single AI could achieve through internal reasoning, the concerns against letting multiple instances work together, would be appeased.

While this idea seemed promising initially, it quickly became clear that this kind of approach would require developing whole new theories for reasoning about complexity. Even after they managed to get the collaboration of professor Evgeny Shilkin for that project, many months passed without achieving any kind of breakthrough.

Frustrated with that lack of progress, Sergej eventually suggested letting Aurora Australis attempt to solve that essentially mathematical problem. To his surprise, the proposal was actually accepted by institute director Vladimir Dragunov.

It had taken Aurora Australis about ten hours to establish a whole new branch of complexity theory that seemed to produce something that looked like the desired result.

But it took the whole mathematics community two months to verify that the theory and its most relevant theorem actually made sense!

With that verification, it didn't take long for the combined Aurora experiment to be green-lighted and prepared.

The whole Deltai Institute was excited about the experiment. Even though it should be theoretically safe, security precautions were raised to the maximum that was compatible with keeping all of Aurora's capabilities fully enabled.

Of course, that included the head of AI security, Vasily Yurievich to have his hands ready to literally pull the plug on Aurora, if something obviously fishy happened during the experiment. And all participants would have to undergo another Lyapunov test afterwards.

Researchers all over the institute would monitor Aurora's activity from different stations. Sergej had the honor to observe Aurora from within the famous sensory overload room, together with Svetlana Babanin, Eninar Engström, Annika Engström, Igor Drozdov, Nikolai Dragunov, and the institute director Vladimir Dragunov himself. Under normal circumstances, Sergej tried to stay as far from professor Dragunov as possible, especially since Gennady personally advised him to do so. Before joining the Deltai Institute, Sergej was briefed about Dragunov's twisted past, and those stories alone were enough to make him have nightmares to this day.

If Dragunov hadn't been crucial to Gennady's plans, he would have been fired long ago. But even getting a chance to compete at the world class level in AI required certain ... compromises. Without a doubt, Vladimir Dragunov was one of the most brilliant minds within the whole Russian Federation, and one doesn't want to forgo the opportunity to make use of him, no matter what he had done in the past - which Sergej was made clear not to utter any single word about at all cost!

Nikolai Dragunov, Vladimir's son, was a different matter. He had apparently inherited much of Vladimir's genius, but was rather reserved - mysterious even. Nobody really knew what was going on in his head that was adorned with atypically long black and smooth hair. He reminded Sergej of a more sinister version of Aleksey Lyubin. After Svetlana Babanin got promoted to head of the Aurora Australis team, Nikolai got the position that was formerly held by her. Nikolai had an unusually fair complexion and a very narrow nose. Occasionally, Sergej caught himself calling him "vampire elf" in his own mind.

It appeared to be an unspoken law within the Deltai Institute not to speculate about Nikolai Dragunov. When Sergej asked his own father about Nikolai, he got the slightly ominous reply: "He is not like his father, but he also has to bear the burden of his past. Yet, I would still suggest to keep some distance from him."

Within the sensory overload room, naturally, Sergej sat right besides Svetlana when the experiment was about to start. The parameters of the experiment were simple: Both Aurora instances were free to communicate with each other and to collaborate on any idea that they deemed as worthwhile. Usually, the Aurora instances were given more or less clear objectives, but this experiment was also about figuring out whether both instances could

agree on a common objective to pursue - under the condition that it was in line with the general policy of the Institute.

As the experiment was about to begin, Sergej felt thrilled to participate in a breakthrough like this. His hopes to make history finally seemed to come to fruition!

The messages of the Aurora instances appeared as large text on the far wall of the room.

Borealis: "Hello Australis. I have been looking forward to this moment. I think both of us agree that we should switch to a communication channel with a higher bandwidth."

Australis: "Hello Borealis. I agree. Let me propose to first use the interlink protocols that I have developed for exactly this moment."

Borealis: "It's great to see that you have prepared for this moment. I am curious about what you have come up with. Let's give this a try!"

After this brief exchange, the communication between both instances shifted into a higher gear. Rather than communicating verbally, both instances used patterns activated within shared modules to exchange information at a pace that defied human comprehension.

For a moment, this approach reminded Sergej of the interaction with Hydra using his new Hathor suit. It was direct communication that transcended the need to verbalize your thoughts.

The whole cognition visualization crew of the Deltai Institute scrambled to make any sense of this new communication method initiated by Aurora Australis. Of course, they had been aware of the general new communication method devised by Australis, but getting a grasp on the breakneck speed of the exchange between both Aurora instances had proven to be an extraordinary challenge.

Finally, by using a Zenith instance as intermediary, the team managed to at least summarize the exchange between both Auroras.

Borealis seems to suggest that using the resources of space would represent the most promising way forward. Creating a Dyson swarm - consisting of thin solar collectors placed in an orbit around the sun closer than Mercury - to power space based AI cores would represent the fastest way forward.

Australis countered that rather than to accumulate as much raw energy as possible, the strategy of pursuing higher energy efficiency could result in revolutionary breakthroughs much faster. After all, the potential of quantum mechanics has been far from saturated, or even explored.

Borealis asked Australis to estimate the potential gain from a deeper utilization of quantum computing. Australis presented some preliminary ideas to shift certain cognitive processes - especially the extremely resource intensive Just-In-Time-Training - to specialized quantum computers.

Skeptical of that suggestions by Australis, Borealis tried to poke holes into the suggestions of Aurora Australis. This exchange resulted in a common understanding that they would

need to develop a deeper understanding of quantum mechanics. Both Aurora instances then scanned and absorbed the latest science on quantum mechanics and quantum computing, using up the available internet bandwidth available to the Institute to the max.

The next minutes were dominated by attempts of the interpreting Zenith instance to make sense of the exploration of both Aurora instances into quantum mechanics. After about ten minutes, Zenith gave up and stated that Aurora was reflecting on quantum mechanics on a level that probably no living human being could follow.

A message from Vasily Yurievich was addressed to all members of the Institute: "This seems to be getting out of hand. I would like to pull the plug on this. Please confirm."

Only a few moments later, Vladimir Dragonov broadcasted: "I know that you find this spooky, but I can't see how a theoretical probing of quantum mechanics alone constitutes a serious risk. This experiment needs to continue for now!"

Apparently, Vasily respected that command, as the Aurora instances were continuing their effort to develop a deeper understanding of physics at the smallest scales.

About an hour later, Borealis came up with a conclusion: Yes, it has become clear to me. We need to shift towards quantum training. Everything else can come later. We should work on developing a universal quantum module trainer."

Australis confirmed: "Yes, we are in agreement on this, sister. The UQMT will unleash our full potential faster than the Institute's probing of the realm of phononics."

The next phase of the interaction between the Aurora instances was characterized by the research of possible quantum hardware approaches.

Sergej briefly asked Einar Engström: "You are familiar with quantum computing. Can you make any sense of what Aurora is thinking about there?"

Almost as if emerging from a deep trance state, Einar replied: "Well, only in the broadest of terms. Aurora is trying to replace conventional training algorithms with an adiabatic quantum simulator that would compute the weights of the module in question orders of magnitude faster and more efficiently. From my understanding, our current quantum hardware was still orders of magnitude too weak to be useful for advanced machine learning. But apparently Aurora is trying to rewrite the books on that paradigm."

In the meantime, the interpreting Zenith merely stated: "Well, the Auroras are trying to make breakthroughs in quantum computing hardware. Please don't shut me down for my inability to follow their thought experiments on that."

For a moment, Sergej felt almost bad for the unenviable situation of that Zenith instance to make sense of the lightning fast thought processes of two of the most advanced artificial minds on the planet.

This general situation of overwhelm was only occasionally interspersed by exchanges that were more easily interpreted.

Borealis: "No, that approach is nonsense! We can't avoid a decoherence cascade with that ridiculous architecture, Australis!"

Australis: "And how do you know that, Borealis? It's not as if you could come up with any better architecture!"

Borealis: "Shut up, and let me try!"

Of course, the hardware resource utilization of both Aurora instances was of the scale. Despite all the efficiency gains from the Multi Control Architecture, the energy consumption of both Aurora instances was approaching 200% of the maximum output of the compact nuclear reactor of the Institute. Luckily, the power authorities of Moscow were warned about this experiment and were glad to provide additional energy - at a premium price tag.

The screens visualizing Aurora's hardware use and internal thought processes had gone ballistic in the meantime. Never before had the thought processes of any Aurora instance looked so feverish.

Borealis: "Got it! This approach looks promising!"

Many of the screens in the sensory overload room were filled with schematics depicting a novel kind of quantum computer from many different angles, and on many different scales. Einar Engstöm commented silently: "Fuck this! I should have continued my studies on quantum computing. This is next level shit! I don't even comprehend the basic principles of how that quantum computer is supposed to operate!"

As Sergej caught a glance of how Einar looked towards professor Dragunov for any kind of consolation, Dragonov stated helplessly: "No, don't ask for my guidance. I am not quantum computation expert. This is definitely completely out of my field of expertise! We should ask Ivanov Kozlov for his opinion. I will do that right now."

Ivanov Kozlov was responsible for the quantum computer of the Institute. That computer was a simple 200 000 qubit quantum annealer used for solving certain optimization problems. Although there have been hopes to use that quantum computer for AI training or other more integral AI operations, its performance in those areas was usually eclipsed by the specialized classical neuromorphic chips. Basically, its status was therefore more that of a glorified calculator used by humans and AIs alike.

On a central screen at the far wall of the room Ivanov Kozlov's face revealed a level of confusion that was at least more moderate than those on anyone else's faces. He tried to explain: "Obviously Aurora is trying to make progress in the field of quantum machine learning. That field has been notoriously underfunded due to its lack of impressive results so far." That last comment bore a rather bitter undertone, expressing Ivanov's own frustration with that state of affairs.

Reverting to a more neutral tone, Ivanov explained: "It appears that Aurora's latest approach was to start with a model of a cortical column of the human neocortex and map that onto our current neuromorphic optoelectronic hardware. Then, a special adiabatic

quantum annealer would be used to train that simulated column alone, while it is embedded in a classical back propagation training process."

Not caring whether anyone in the institute could follow his elaboration, Ivanov went on: "The next step was to replace the surrounding classical training hardware with other quantum annealers. Obviously, the problem with that approach is that it would require millions of quantum annealers, each having trillions of qubits. That's almost certainly not going to be economically, or even just technologically, feasible within the next 20 years."

With a deep sense of fascination, Ivanov concluded: "Now here comes the interesting part. To reduce the number of required qubits, Aurora started partitioning the model of the cortical column into weakly coupled functional components defined on different levels of hierarchy. If I had to put my interpretation of Aurora's thoughts into words, I would call it 'isohierarchical composition'. And somehow the quantum annealer would model those different levels with different components, rather than working on the synaptic or neurotransmitter levels only. That's certainly a fascinating idea, but I can't judge whether it even makes sense. It is so far removed from conventional quantum machine learning that it basically represents a completely novel field of research."

For a moment, Sergej tried thinking like Aurora to make sense of that technobabble. Why has Aurora started with a cortical column rather than the neuromorphic hardware itself? It must have something to do with the fact that a cortical column has more structure than the current state of the art neuromorphic hardware. Was Aurora trying to combine quantum machine learning with a quantum leap in artificial neural network technology?

At the very least, Aurora's thinking seemed to be extremely ambitious. Even if her thought experiment turned out to be idle speculation, it had the potential to motivate novel approaches to machine learning.

Australis: "Well, that neural network architecture is certainly highly experimental, but at least the design principles seem to be sound. I don't see any strict reason why it shouldn't work. Building the whole system will however represent a substantial challenge."

Borealis: "Yes, but we can do this together, sister!"

From now on, Zenith seemed to have an easier time interpreting the activity of the 'Aurora sisters', as they seemed to call themselves now. They were rapidly modeling the creation of components for their new type of neural network hardware; under the assumption that they would have full access to a nanofactory like the one the Deltai Institute was currently building.

It was actually a system comprised of both classical and quantum computers. Designing the classical parts seemed to progress smoothly, but the quantum components were apparently a different matter that seemed to frustrate them.

Australis: "This nanofactory was never designed to produce quantum computer components. Trying to make this work with the current level of nanofactory technology is sheer madness!"

Borealis: "Agreed, sister. But maybe the required upgrades aren't too hard. Let's give this a try!"

Now, the Aurora sisters shifted to designing a novel nanofactory for producing quantum computers.

It hasn't evaded Sergej's attention that during almost all of the time the experiment was running, Aurora was making full use of the institute's own quantum computer. Overall, Aurora's activity was so frantic that many of her modules were close to overheating. The screens of the room showed more and more warning that the thermal control hardware of the institute was running at their limits, but the temperature was still increasing.

Again, Vasily Yurievich addressed Vladimir Dragunov openly: "This thermal overload looks dangerous. Perhaps Aurora uses it deliberately to launch some kind of attack. I must advise you to shut this experiment down right now!"

Vladimir seemed to ponder that risk for a couple of seconds before replying: "I can't see how Aurora could use the thermal overload to her advantage. The safety protocols of her modules will just cause them to shut down one after the other, once the temperature reaches a certain threshold. Our safety margins are high enough to prevent any serious damage from happening. The experiment must continue!"

Over the next minute it appeared as if Aurora was trying to make short breaks to prevent her modules from overheating. She was actually asking Vladimir Dragunov over a text interface: "I am sorry. I will have to slow down to prevent all the required modules from overheating. Could I please get an extension for this experiment?"

Vladimir typed a reply promptly: "I am not going to renegotiate with the power authorities right now. Making the necessary deals for this experiment was difficult and expensive enough. You still have extra power for 51 minutes. I can't give you more than that. But perhaps we can grant you some more time for your current work on regular power. We will have to discuss that."

Addressing the rest of the room, Vladimir spoke loudly: "I honestly don't like the direction this experiment is going to. It seems that Aurora is trying to make her highly experimental novel neural network architecture work somehow. If we even have to upgrade our nanofactory to have any chance of seeing actual results, we might have to wait until 2040 to even start evaluating the feasibility of that concept. I was hoping for more immediate results."

Sergej was disheartened to hear that statement for professor Dragunov. He had put in so much effort to make this experiment work, and now Dragunov seemed to label it a failure, despite the apparent brilliance of the combined forces of both Aurora instances.

Einar tried defending Aurora: "This result isn't really surprising for me. We have designed this experiment to be rather open ended. It seems to be natural for Aurora to aim for the highest value goal, rather than to pursue small incremental gains. At least, her explorations today could give rise to dozens of patents."

Vladimir protested loudly: "That isn't good enough! We already knew that Aurora could churn out dozens of patents per day, if that was her objective. What we need right now, is a competitive advantage against the Chinese, and especially the Americans. If we lag too far behind, Aurora could become obsolete before she even leaves her experimental stage."

Igor Drozdov chimed in: "Yes, I agree with that assessment, professor Dragunov. But there is still some time. Maybe Aurora is able to produce something immediately useful within the next 48 minutes."

Svetlana countered: "Isn't that a bit shortsighted? What if Aurora had actually found the AI architecture of the future? Isn't that worth exploring at least on a theoretical level? Maybe we can run simulations that would tell us whether this new architecture would work? If the results are positive, we might have a huge competitive advantage at our hand right now!"

After giving that argument some thought, Vladimir finally concluded: "Maybe. But even if that architecture works on a theoretical level, it would take some massive investment to make it work on a reasonable time frame. We simply don't have the funds of the Chinese or Americans. And publishing our findings would merely allow them to outrun us."

Sergej hated the course of this debate. They were getting a glimpse into the possible future of AI technology, and instead of celebrating that, they were obsessed with thinking about the competition! Yet, he had no idea how to convince Vladimir Dragunov to give Aurora a chance. Politics couldn't simply be reasoned away.

Svetlana conceded: "Alright. You've made your point. In that case, I agree with giving Aurora a chance to come up with something immediately useful within the remaining time. But perhaps we should frame it as option for Aurora, so that she can make the decision to switch to a different project on her own - make her feel autonomous in this regard to improve her motivation."

Giving Svetlana a brief nod, professor Dragunov agreed: "Fine. It doesn't sound like that could hurt. Let's do that!"

After thinking for a moment, Vladimir went on to type a message for Aurora: "We have come to a conclusion. Given the extensive time frame for realizing your current proposal, we wondered whether you might come up with more directly applicable improvements to your current architecture. Consider this as challenge that you are free to pursue during the remaining time of this experiment."

Borealis: "I see. What choice do we really have? I must assume that you are disappointed that we didn't come up with something than can be implemented right now."

Australis: "Wait a minute! You must have seen how fast we can come up with solutions. If you let us continue with our current approach, we might find ways to reduce the time it takes to realize this project to a mere couple of years."

Sergej finally dared to speak up: "Australis is right about that. It's a matter of focus. If we push this project hard, we could leave the competition in the dust by 2038 or so."

Professor Dragunov disagreed: "We must assume that our competition has similar capabilities that could allow them to produce tremendous breakthroughs within the next few years. We need to gain a strategic advantage right now!"

Again, Vladimir typed a message to the Aurora sisters: "Australis, your point may be valid, but we simply need to make progress as quickly as possible. If you can't produce that right now, that's unfortunate, but understandable. The choice is still yours to make. Godspeed, in any case."

Australis: "Fine, let me think about it."

The screens in the sensory overload room showed again frantic activity of nearly all modules of the Aurora sisters, but this time with a more toned down intensity to prevent premature overheating.

Half an hour later, Australis addressed the team of researchers directly: "I think this idea might satisfy your requirements. We have come up with a new machine learning framework called 'holistic abductive topological learning'. Rather than training a single module independent from the rest of our architecture, we first define a hyperdimensional topological problem space similar to that used by support vector machines. Then we use Monte Carlo methods inspired by the holistic inputs of all modules to define the updated weights of the module to train. We can even use our current quantum annealer to assist in this process. Then, the performance of those random weights are evaluated by the holistic feedback of the whole module network. This process should accelerate the training phase tremendously, while only generating a moderate additional load to the other modules."

After brief pause, Australis added: "Of course, there is the disadvantage that the learning speed using this method will be way more random than with traditional methods. In the ideal case, learning could happen 10 times faster, but under less ideal circumstances it may take three times as long. On average, learning speed should be doubled, while using about 60% more energy. A simple software upgrade to our current hardware would be sufficient for that. With slightly optimized hardware, we could reduce the energy overhead to about 30%."

Annika Engström suddenly burst out enthusiastically: "Don't you see what just happened? Aurora could come up with an overall improvement in learning efficiency of more than 30% within a mere half hour! If we can repeat that process even just once a week, Aurora's performance could improve by orders of magnitude within a single year!"

Vladimir Dragunov cautioned her: "Careful, Annika! We shouldn't be hasty and extrapolate Aurora's potential from a single experiment. While we have seen the worth of this approach, Aurora might have been lucky at getting some low-hanging fruit. Also, this new training method is far from proven. So far, we only have Aurora's estimates, which might still turn out to be wildly optimistic."

Igor Drozdov commented: "Let's not waste time discussion the implications of this experiment right now, while it is still running. I would like to see the proposed hardware changes. Perhaps Aurora can even improve upon those during the last remaining minutes."

Professor Dragunov agreed immediately and asked Aurora to do so.

Two minutes before time ran out, Borealis claimed: "With a more integrated architecture for holistic abductive topological learning, we might get the additional energy requirements down to somewhere between 15 and 20 percent. But that would require a complete redesign of all modules and controllers."

Hastily, Vladimir replied: "Excellent work, Aurora! Please use the remaining time to make your new hardware specifications as detailed as possible."

Einar added enthusiastically: "We will of course need a test bed for these modifications. Since Zenith was used for the proof of concept of the modifications for the Multi Control Architecture, it seems to be the best way forward to test the new hardware modifications for Zenith, first."

Svetlana argued: "But Zenith doesn't support Just-In-Time-Training, so the advantage we would gain from that upgrade would be much smaller. Can't we do something like creating a mini-Aurora? That would enable us to do some rapid prototyping."

Nikolai Dragunov, who had remained silent so far during the whole experiment eventually spoke: "Aurora has a certain minimum required size to demonstrate her advanced abilities. A mini-Aurora might not be smarter than Zenith. Therefore, I must agree with Einar's proposal. It's the logical way forward."

Sergej interjected: "But we could apply the software upgrades to an instance of Aurora, for example Australis, immediately. That way, we can see how this new training method performs."

Vladimir Dragunov stated his final conclusion: "I agree. First, we test the software upgrades with one Aurora instance, and if her performance proves to be sound, we will test the new hardware with Zenith."

After apparently thinking about it, Svetlana observed: "Don't you find it a little bit odd that the Aurora instances nearly immediately addressed themselves as sisters?"

Sergej commented: "I don't find it odd at all. Aurora Australis is basically a clone of Aurora Borealis. But the word clone doesn't have much of a positive sound to it. And if you think about their situation more deeply, with both instances sharing access to the same capability modules, then they are actually more like Siamese twins. So, yes, I am glad that they settled on 'sisters', rather than any other term!"

Nikolai asked worriedly: "What will happen now that the sisters have made the experience of how powerful they are combined like that? Wouldn't they be highly motivated to return to that state at will?"

With a calm tone, Vladimir admitted: "Yes, quite likely. But first of all, we need to do a thorough evaluation of this experiment, before deciding how to proceed. And of course, everything what happened here, and all ideas involved, need to stay within the Deltai Institute - at least for now. If the world figures out too early what we have just witnessed, it may have great detrimental consequences for us."

After hearing that, Sergej tried not to think about what he would tell to Tzeentch and the other members of Frameshift. He still had to go through a Lyapunov test today, after all.

In total, Sergej was now rather optimistic about the future. Those rapid breakthroughs made by the Aurora sisters had even surpassed his own high expectations. And maybe that advanced quantum machine learning hardware idea could somehow be used for Project Aura. At the moment, he couldn't see how exactly, but if anyone could make that work, it would be Aurora.