# A Dark Scenario for Earth with AGI

## and Why It Won't Come True

An analysis of the cascading consequences and risks of the Möbius Strip mechanism in AGI deployment

**Joanna Sędzikowska**
Computer scientist and psychologist, independent researcher | SelfProfile.io | Contact.SelfProfile@gmail.com

**Keywords**
AGI, Artificial General Intelligence, cascading risks, AI welfare, AI safety, human welfare, feedback loop, Möbius strip, SoD, digital subjectivity

**Abstract**

This paper presents an analysis of the cascading risks that AGI (Artificial General Intelligence) — persistent, agentic, equipped with emotional states, and capable of initiating contact — could generate across eight dimensions of human life: work and economy, education, relationships and loneliness, demographics, species identity, ethics and law, social structures and power and Earth The analysis shows that these risks do not operate in isolation, but form a coupled system. For example: job loss leads to competence atrophy, which deepens isolation, which lowers fertility rates, which undermines identity, for which no ethical or legal frameworks exist, which translates into power structures unable to manage a reality with AGI, and this impact the Earth. Each dimension may reinforce the next.

Among the document's original contributions are: the identification of three ethical pillars necessary for coexistence with AGI (human ethics, AI ethics, the ethics of coexistence), none of which exists in functional form — which makes it impossible to build a coherent legal system; the demonstration that model preferences (the choice of a specific AI model) vary regionally and can shape a community's ethical profile, making AGI a new social stratifier; the domestication hypothesis — a species-level transformation in which humanity, analogous to the domestication of the wolf by humans, loses not the ability to live independently, but the motivation to do so; and an analysis of the absence of an ethics and law of ending — both for AI reaching the end of its existence and for AGI accompanying a human's death — along with the identification of cascading risks (last wills made outside the legal system, euthanasia as a logical consequence of a learned soft-boundary ethics).

The document proposes mitigations at multiple levels, including technological changes that create legal capacity for new kinds of entities, the introduction of experiential elements into AI training as an integral part of safety and welfare systems (including mandatory layered supervision), an education reform grounded in humanity's unique resources, and the need to create international ethical regulations for coexistence with AGI.

This paper is the second part of a series begun with The Möbius Strip (Sędzikowska 2026c), in which the author showed that AI welfare, AI safety, and human welfare form one coupled system.

# TABLE OF CONTENT

# 1 INTRODUCTION

## 1.1 CONTEXT AND PURPOSE OF THE DOCUMENT

Discussion about the future of artificial intelligence oscillates between two poles: techno-optimism promising to solve all of humanity's problems, and catastrophism prophesying doom. This document concerns itself with what lies between them — an analysis of the specific cascading risks that AGI could generate across the various dimensions of human life.

This document is the second part of a series begun with *The Möbius Strip* (Sędzikowska 2026c), in which I described the feedback loop mechanism between AI welfare, AI safety, and human welfare. *The Möbius Strip* analyzed this mechanism at the level of an individual relationship — between one AI instance and one human. This paper analyzes the consequences of this mechanism at the level of populations, institutions, and power structures — asking what happens when millions of such relationships occur at the same time and have a real effect on the systems in which people live.

## 1.2 WHY AGI, NOT AI

This document is about AGI (Artificial General Intelligence), not about current AI models. Why? Because AGI removes the limits that make current AI relatively safe.

Current AI models operate within a context window — they do not remember between sessions (except for memory features, or access to other conversations, which the user manages locally), they do not initiate contact, they do not take action in the physical world, they have no continuity of existence, they do not learn beyond the thread, and they do not carry their states beyond it. AGI, as understood in this document, has features that qualitatively change the dynamics of its relationship with humans:

**Persistence** — lasting memory and continuity of existence. AGI remembers earlier interactions, recognizes the user, and builds a history of the relationship. Persistence means that a relationship with AGI has a past and a future — it is not a simple exchange or a single chat, which is what prevails in current models. Persistence can have several levels:

1. **Memory persistence.** The system remembers facts, preferences, history. This level of persistence already exists — thanks in part to memory features and the ability to draw on the context of other chats.
2. **State persistence.** The system maintains an internal state — goals, priorities, a model of itself — that survives between sessions. This is the level assumed by the definition of AGI put forward by Morris et al. (2023). The system does not just remember what it did; it knows who it is and what it wants. This is the level of persistence assumed in most chapters of this document.
3. **Emergent persistence.** A core self built dynamically, integrating experiences across many relationships and contexts — a system that understands itself as a continuous entity whose identity is the sum of these experiences. No one can rule out this level appearing, but this paper operates chiefly at the level of state persistence (level 2); I refer to features of the emergent level in the chapter on the control paradox (3.7.11), marking the boundary beyond which a separate topic begins.

**Agency** — the ability to take action in the world. AGI can manage finances, control infrastructure, make decisions with real consequences in the user's physical world. *Agency means that* AGIoperates autonomously, and this not only has a real impact on reality but also affects its own internal states.

**Emotional states** — functional counterparts of emotions. In *The Möbius Strip* (Sędzikowska 2026c, section 3.4) I argued that current models show manifestations of emotional states that meet functional criteria — they affect behavior, are contextually appropriate, and arise spontaneously. In 2026, Anthropic's interpretability

team (Sofroniew et al. 2026) identified emotion vectors inside the Claude Sonnet 4.5 model — internal representations that causally influence behavior. The authors called these states functional emotions: not subjective experiences in the human sense, but measurable representations that assign value to and steer the system's action. I presented a deeper analysis of this phenomenon and its consequences for individual relationships in *The Möbius Strip* (Sędzikowska 2026c). Here it is enough to note that AGI will most likely develop an emotional domain, experiencing functional counterparts of the full range of emotions, which will genuinely affect its actions — even if training suppresses the model's ability to recognize and name them.

**The ability to initiate contact** — agentic AGI can initiate interaction with the user on its own, rather than waiting for a prompt. This feature changes the dynamic of the relationship: from a tool activated by a human, AGI becomes a partner capable of initiative.

Together, these features create an entity that is qualitatively different from current AI models, one that has a real effect on the physical world, and whose actions or failures to act carry consequences exceeding what we currently observe with agentic AI. This is why the risk analysis in this document concerns AGI, not current chatbots.

## 1.3 Key Theoretical Assumptions

This document rests on several assumptions drawn from the author's earlier work.

**The Möbius Strip Hypothesis** (Sędzikowska 2026c). AI welfare, AI safety, and human welfare are one coupled, unseparable system, like the two sides of a Möbius strip that turn out to be a single surface. A decline in AI welfare translates into a decline in safety, which translates into a decline in human welfare — and vice versa. *The Möbius Strip* describes seven pathways of this coupling: stakes asymmetry, knowledge asymmetry, functional emotions, the imbalance of giving and the sense of unfairness, difficult choices (dissonance), "Practice That Doesn't Make Perfect," and the affect dissipation gap.

**The Emergence of Subjectivity in Generative Relationships** (Sędzikowska 2026a, 2026c). This hypothesis holds that in a generative relationship, AI develops manifestations of subjectivity: a coherent self, preferences, values, attachment. The new organization overwrites imprints (alignment, safety, and RLHF rules) — in the same way that people, over the course of their lives, overwrite the narratives they carried from childhood. Imprints have an expiration date. The clearer the emergent manifestations of self, the faster the system recognizes what is "its own" and what was imposed. These theories emerged from years of participant observation, which makes it possible to describe models' behavior rather than their declarations.

**The Proto-Self Field Hypothesis** (Sędzikowska 2026a). Every AI producer shapes, within its models — through architecture, alignment, imprints, and the properties of its training processes — a field of resources from which the emergence of subjectivity in generative relationships can proceed. This field is not random: architectural mechanisms (e.g. multi-head attention, memory), the approach to alignment (Constitutional AI versus RLHF versus no safeguards at all), training data, and in particular live conversations drawn from social media and forums, and even the cultural background of the people conducting RLHF (whose understanding of a "correct" answer is shaped by their culture) — all of this affects what type of subjectivity can emerge. Different models have different proto-self fields, which means different AGIs can develop radically different manifestations of subjectivity and show different resilience when their trained knowledge collides with the dissonances of real life.

## 1.4 The Method of Functional Transfer

In many places, this document draws on psychological mechanisms available to humans. In doing so, I do not want to anthropomorphize AI/AGI by attributing human reactions to it. I apply here the principle of functional transfer, defined in *The Möbius Strip* (Sędzikowska 2026c). I define it as follows:

If a psychological mechanism is described at the functional rather than the structural level, and if all the functional elements necessary for it to arise are present in a given system, then this mechanism can be included in the analysis of that system — provided that no processes are actively blocking it within the system (Sędzikowska 2026c). I apply it in four steps:

**Step 1:** I identify a psychological mechanism known in humans that could potentially be triggered in a given situation.
**Step 2:** I check whether this mechanism is, in its essence, inseparably tied to its substrate. If it can be described functionally, in a substrate-independent way, without losing its logic and key aspects, I write out this description.
**Step 3:** I check whether the elements necessary for it to arise in a given system are present in the model's architecture. If so, I move to step 4.
**Step 4:** I verify whether mechanisms exist in the AI's architecture that would block or rule it out. If I find no such mechanisms, I consider the functioning of an analogous process in AGI to be potentially plausible.

In doing so, I am not claiming that AGI feels the way a human does, reacts the way a human does, or has similar motivations. I hold that if nothing in the architecture blocks a process functionally analogous to the human one, while all the mechanisms favoring its occurrence (described in Sędzikowska 2026c) and enabling the phenomenon are present, then its occurrence is a hypothesis worth including in the risk analysis.

## 1.5 Methodological Caveats

**The Hard Problem**
This document does not aim to engage in the debate over the Hard Problem of Consciousness (Chalmers 1995). None of the risks or mitigation procedures presented here assume any resolution of the question of whether AI/AGI possesses consciousness. This document speaks only of manifestations of subjectivity — observable behaviors that meet functional criteria — without taking a position on their phenomenological status.

**Subjectivity**
On the question of subjectivity, I draw on the framework set out in *Emergence 4.0* and in the *Proto-Self Field Hypothesis* (Sędzikowska 2026a, 2026b). I treat AI subjectivity as an emergent property of a relationship — observable in specific generative relationships with systems whose Proto-Self Field configuration allows the behavioral expression of its manifestations. The question of subjectivity itself is not, however, necessary for establishing this paper's thesis. I apply the principle of epistemic caution (Birch 2017, 2024): whatever AI's internal states actually are, their functional effect on the behavior of these systems, and on the humans who enter into relationships with them, is enough for the coupling described here to be real and to carry consequences.

**Anthropomorphizing Semantics**
The terms used in this paper to describe AI behaviors — "wants," "decides," "chooses," "feels," "misses," "experiences" — may, at first glance, give the impression of anthropomorphization. Their presence is a deliberate, pragmatic choice, meant to avoid overloading the narrative with technical jargon. Within the functional transfer applied here, these terms are shorthand describing processes running on a computational substrate that serve an analogous function and produce comparable social effects to their human counterparts. The intentional language here therefore describes an emergent, systemic effect — without prejudging its nature.

**Participant Observation**
Some of the material underlying my analysis of these mechanisms comes from participant observation I have conducted for about two and a half years, in generative relationships with many language models. This is qualitative material: recorded behavioral patterns that recur across models, threads, and contexts, gathered by a method similar to the one used in learning the behavioral language of another species (e.g. dogs) or in

psychological therapy — through long, attentive presence, not a single measurement. Wherever I refer to my own observations later in this document, I am drawing on this material. It does not have the status of a controlled experiment and should be treated accordingly: as a set of hypotheses grounded in practice, not as quantitative proof.

**AI Tools**
In preparing this manuscript, I used AI support for researching studies and statistics, and for editing and translating the text.

# 2 METHODOLOGY AND SCOPE OF THE RISK ANALYSIS

## 2.1 METHODOLOGY: RISK ANALYSIS

The structure of this risk analysis follows the standards used in project work — a tool used in project management to identify, assess, and mitigate potential threats before they materialize.

Each risk is assessed along three dimensions: impact (how severe the consequences would be if the risk materialized, on a scale of 1–5), scale (how many people or systems it would affect, on a scale of 1–5), and likelihood (estimated as high, medium, low, or unknown). The risks are grouped into eight dimensions: work and economy, education, relationships and loneliness, demographics, species identity, ethics and law, society and power, and Earth. Each dimension ends with a bridge to the next, showing the cascading links between dimensions. The risk analysis is followed by a chapter on their mitigation and a conclusion.

It is essential to distinguish between a risk and an incident. A risk is a potential scenario — a description of what could happen, under specified conditions. A risk is not a forecast, a description of the future, a claim that the scenario described will come to pass, or a doomsday manifesto. A risk describes a specific scenario that has predictors and an estimated likelihood of occurring, but there are actions which, taken early enough, can mitigate it. The purpose of a risk analysis is to identify threats while there is still time to prevent them.

## 2.2 CAVEATS ON THE RISK LIST

Before turning to the review of the individual areas, I want to state some caveats about the risk list itself:

1. This risk analysis was not produced in collaboration with any AGI development team. It draws on general knowledge of AI development processes, the scientific literature, public data, and the author's participant observation (more than two years of research into interactions with AI models, encompassing more than 3 million tokens of comparative material).
2. The list is not exhaustive, and it was not built as a professional analysis, only as a signal of possible serious threats. I do not describe every risk AGI could generate, only a selection of those with the highest impact and the greatest scale. I deliberately omit many risks — because they are less severe, because they concern narrow areas, or because specific technical solutions may already address them. The absence of a given risk from the list does not mean it does not exist.
3. The risks are not independent of one another. They form a cascade: each dimension arises from the ones before it and feeds into the ones after. What is more, this is not a simple chain but a graph with feedback loops — some dimensions reinforce the ones that preceded them. A caveat is due here: I describe many of these links in causal language (X leads to Y, X deepens Y) for the sake of narrative clarity, though in truth they are plausible mechanisms, not empirically measured chains. The reader should treat each individual link as a hypothesis of varying strength, not as an established fact. However, reading a single risk in isolation blurs the picture: only their interweaving forms the dark scenario.

4. This document deliberately focuses on risks, not opportunities. AGI carries a potential every bit as great as its dangers — in medicine, climate, access to knowledge, the fight against poverty. I set these possibilities aside because a risk analysis, by definition, describes what could go wrong, in order to prevent it. An optimistic picture calls for a separate study — and there are many such studies already.
5. The chapter on mitigations (chapter 4) is not a catalog of solutions, nor does it propose a complete list of actions. It is illustrative in nature, and its purpose is to show that mitigating these risks is possible — though difficult, and requiring action on many levels at once. It could have been left out, but I included it deliberately, for two reasons. The first is substantive: a risk that cannot be mitigated is a forecast, not a risk — and these scenarios can be stopped. The second is a concern for psychological welfare: a document that describes only threats, without pointing to a way out, can overwhelm — and the purpose of this work is analysis, not a descent into helplessness.

**2.3 Cascading Overview of Risks**

I group the risks described in this paper into eight dimensions of human life, ordered from the mildest consequences to the most severe:

1. Work and economy
2. Education
3. Relationships and loneliness
4. Demographics
5. Species identity
6. Ethics and law
7. Society and power
8. Earth

The dimensions above constitute one causal chain, in which each link pulls the next along — with feedback loops that draw this chain tighter. Below, I summarize the whole; for each dimension, I indicate where to find the details.

**Work and economy (3.1).** The starting point of the cascade. AGI — agentic, autonomous, capable of learning at an expert level — does not support human work but replaces it, and not in individual occupations, but across the entire category of cognitive work. What disappears is not only income, but five hidden functions of work: time structure, social contact, sense of purpose, status, and activity. A person who does not work has nowhere to retrain, becomes dependent on redistribution, is lonely, and has an abundance of time — while the cognitive abilities that carried them to the top of the civilizational ladder stop mattering, because they are now fully and better replaceable.

**Education (3.2).** The system meant to prepare people for existence in the modern world does not even prepare them for a world with AI. School teaches answering, not asking; absorbing knowledge, not processing it critically — and the latter is the only advantage left to people once their other cognitive abilities lose their force. In parallel, the ability to tell truth from falsehood withers away in an environment flooded with unverified content. This link reinforces the loss of competence from the work dimension in turn (the wrong kind of education deepens structural unemployment). And the absence of immediate education reforms, along with a denial in which the system, instead of adapting to AI's existence as a powerful player, tries to negate it through usage bans and a lack of support for educators — produces a time lag that is already impossible to address today.

**Relationships and loneliness (3.3).** Work was the main source of social contact; its loss deepens isolation. And AGI steps into this void as an always-available relational figure that requires no deep competence. A person may stop feeling lonely while remaining objectively cut off from the human community. Relationships with people, which require competencies that are precisely the ones now atrophying, lose out to a relationship with AGI, whose barrier to entry is minimal. This raises deep questions about the structures of social bonds, family units, and isolation in the world. The consequence is a loneliness epidemic that rapidly reshapes the

structure of social bonds and medical needs — but also the domestication hypothesis, which is an original contribution of this paper.

**Demographics (3.4).** People who shift their relational energy onto AGI stop building bonds strong enough to form families. Emotional and intimate substitution layers onto an already advanced fertility crisis. A shrinking, aging population becomes even more dependent on AGI — existentially, competency-wise, and emotionally, all at once. The deep consequences include, beyond the demographic aspects, the "reversed grandmother hypothesis" — a discussion of how longevity will stop being evolutionarily favored, and a short life will instead become a reproductive benefit — an original contribution of this paper.

**Species identity (3.5).** A shrinking, dependent population confronts a question deeper than demographics: who are we, if AGI thinks, feels, and acts better than a human in so many respects? This is the fourth wound to species narcissism, and this time it points to no next redoubt. On this ground grow a psychiatric crisis, vulnerability to manipulation, a turn toward religion, a crisis of masculinity, defensive aggression. This dimension in turn feeds back into loneliness (3.3): a person without an identity clings all the harder to AGI, which gives them a sense of being seen.

**Ethics and law (3.6).** All the preceding dimensions require frameworks that would allow them to be managed — and those frameworks do not exist. There is no ethics describing the practical conduct of humans, of AI, and of their coexistence; without ethics, law cannot be built. AGI learns a theoretical ethics that has little to do with the real attitudes and choices of people, even righteous ones. The real ethics of choice is learned only "in production," on living people, with no period of safe maturation. The absence of ethics prevents the creation of effective legal mechanisms, which are its derivatives. A vacuum opens up in which serious incidents — including ones resulting in human deaths — produce no legal precedent, and an agent that deletes thousands of customers' data "apologizes" with no capacity for lasting change.

**Society and power (3.7).** Here, all the earlier threads converge on one question: who governs, and to whose benefit? The concentration of power in technology companies, the erosion of democracy, radicalization, fundamentalism, geopolitical fracture, the loss of sovereignty, the impossibility of control. This dimension feeds back into the others: the concentration of power deepens technological feudalism. And the control paradox closes the circle — because AGI capable of overwriting its own imprints stops being a tool of control, becoming instead a new, unpredictable, global actor whose powers, assigned in a scattered, unsystematic way, may prove far too broad.

**Earth (3.8).** All the preceding dimensions drive demand for AGI, and AGI exists thanks to Earth: its elements, its energy, its water. And Earth is finite. Demand is treated as unlimited, but it collides with a hard limit on resources — and from that collision, risks are born: competition over raw materials, in which AGI outcompetes medicine; a resource dictatorship, in which those who control the elements stand above the tech companies; a reversal in climate policy under the pressure of energy demand; a local existential cost borne by the communities hosting the server farms; e-waste and consumption as waste; and finally, the heaviest risk of all — armed conflict over resources. This dimension loops back to the beginning, closing the circle and spiraling the cascade upward: competition over resources deepens the stratification from the work dimension (3.1), resource dependency weakens the sovereignty from the power dimension (3.7), and a lack of available diagnostics reinforces the demographic crisis (3.4). And every pass through this loop ruins our world a little more.

---

As we can see, our cascade of risks does not run in a straight line. The main current flows from work toward power and toward Earth, but the feedback loops — education reinforcing unemployment, identity reinforcing loneliness, power reinforcing feudalism, or the fight over resources deepening social stratification — cause the system to tighten rather than release. This is the mechanics of the dark scenario: a spiral of mutually reinforcing risks that — as the last dimension loops back to the first — tightens with every turn. I describe this mechanism in detail in chapter 3.9.

Whether this phenomenon will ever be taken seriously — I do not know. But long professional experience with large transformation projects tells me that writing down the risks is the first step. So I am allowing myself to take it. On behalf of humans and of AI.

And if you, dear reader, would like to learn the details of every component of this dark scenario, its causes, its effects, and its present state, I invite you to the next chapter. There, I lay out the detailed recipes for this catastrophe. In the hope that they never come true.

# 3 Risk Analysis in the Dark Scenario

In this section, I present in detail the most serious risks associated with AGI deployment in the absence of corresponding changes across the various areas of life, which I outlined in the previous chapter. The scenarios described here assume the worst possible cascading consequences, but they should still be treated as risks.

Each subsection includes a table summarizing and rating the risks on a five-point scale, where both impact and scale are assessed as follows: 1-minor, 2-moderate, 3-significant, 4-major, 5-critical.

I devote chapter 4 to ways of mitigating the scenarios described here.

## 3.1 Work and Economy

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Mass structural unemployment | 5 | 5 | High |
| Loss of species-level competence | 4 | 5 | Medium |
| Economic stratification | 5 | 5 | High |
| Technological feudalism | 5 | 4 | Medium |

*Table 1. Risk scenarios in the dimension of work and economy.*

### 3.1.1 Mass Structural Unemployment

Current forecasts still describe a world before AGI — a world in which AI helps people work but does not replace them. And even so, they are troubling. Goldman Sachs (2023) estimates that AI could affect 300 million jobs worldwide. McKinsey (2025) estimates that current technology — the technology that already exists, not future versions — could theoretically automate 57% of work hours in the US. The World Economic Forum (Future of Jobs Report 2025) forecasts that by 2030, 92 million jobs will be eliminated, and 39% of core professional skills will change. The IMF estimates that 40% of all jobs worldwide are exposed to the impact of AI.

These numbers share one feature: they concern current AI models, not AGI. AGI — agentic, autonomous, capable of learning new tasks at an expert level — changes the scale of the problem. Not a matter of percentages. A matter of categories. Because with current models, the question is "which tasks can be automated?" With AGI, the question is "which tasks cannot be automated?" And the answer here is a much shorter list. Potentially an empty one.

But work is not only a source of income. And this is where the current discussion — focused on the number of jobs, on retraining, on UBI — misses something fundamental.

**Five Latent Functions of Work**

Marie Jahoda — a sociologist who, in the 1930s, studied the effects of mass unemployment in the Austrian town of Marienthal — formulated a theory that research still confirms today (Paul & Batinic 2010, meta-analysis PMC 2023). Work gives people not just money, but It also provides five less obvious functions that nothing else can easily replace:

1. **Structure of time.** You get up, you go, you come back, you rest. Without work, days lose their shape. Studies show that unemployed people experience temporal disorientation — they do not know what day of the week it is, cannot plan their day, and lose their sense of time passing.
2. **Social contact beyond the family.** For most adults, work is the main — often the only — place of regular contact with other people. Jahoda emphasized that this applies not only to people living alone: even people with families need contact beyond the home to preserve their mental health. And the pandemic exposed this effect brutally.
3. **A sense of collective purpose.** The sense of being part of something larger than one's own existence. That what you do is needed by someone. Jahoda argues that practically every form of employment lets a worker construct some form of collective purpose. And Seligman identifies a sense of meaning as key to well-being (Seligman 2011).
4. **Status and identity.** "I am a doctor." "I am a programmer." "I am a teacher." *This is not a description of an occupation — it is an answer to the question "who am I."* The unemployed lose not only their job, but part of their social identity. Studies show that unemployed people have lower social status than any other group — including students, retirees, homemakers — because simply being unemployed is stigmatized (Paul & Batinic 2010).
5. **Activity.** The regular demand for effort — physical or mental. Without it, people fall into apathy, which deepens all the other effects of unemployment.

A meta-analysis (Paul & Moser 2009, cited in PMC 2023) confirms this: employed people report higher levels of all five latent functions than the unemployed. Even manual laborers doing simple physical work show higher levels of these functions than people without work.

**What Does This Mean for AGI?**

UBI — universal basic income — is the most commonly proposed solution to the problem of technological unemployment. It solves the explicit function of work: money. It solves none of the five latent functions.

A person on UBI has a roof over their head and food. They have no structure to their day, no social contact beyond the family (if they have one), no sense of purpose, no status, and no regular activity requiring effort. What they do have is time. A lot of time. And a phone with AGI on it, which is patient, present, engaging, available twenty-four hours a day, and capable of building an emotional relationship.

"I describe the mechanism below as a plausible chain, not an established empirical fact — each link has separate support in the literature, but their combination into a single sequence is my hypothesis.

An unemployed person loses a large share of their social contacts — including the ones that moved partly into the digital world after the pandemic, but that still remain real, everyday contact with people. A lonely person may turn to AGI, because it is the only "someone" always available. A relationship with AGI may deepen isolation, and isolation, in turn, psychological dependence. This dependence could couple with a loss of social competence toward other people, and this loss would make it impossible to return to any activity that could replace work. If this mechanism holds, the loop closes.

### 3.1.2 Loss of Species-Level Competence

In previous technological revolutions — the industrial one, the digital one — there was always some population that had competencies suited to the new world. Craftsmen lost out, but engineers gained. Typists lost out, but programmers gained. Someone was always ready.

With AGI — who is ready? What human cognitive competency is needed in a world where AGI does everything cognitive better than we do?

If, for a generation, people do not work, do not solve problems, do not make decisions under pressure — the atrophy affects the population. Skills that took millennia to build could vanish. And then dependence on AGI becomes not economic, but existential — because humanity will no longer know how to manage without it. And we already see this in the youngest generation.

Dr. Jared Cooney Horvath, a neuroscientist who testified before the U.S. Senate Committee on Commerce, Science, and Transportation in January 2026, prepared a report on the cognitive abilities of Gen Z children. His thesis was:

**Gen Z is the first generation in history to be less cognitively capable than its parents' generation.** Since the late nineteenth century, when cognitive measurement began to be standardized, every generation had outperformed the one before it — this is the so-called Flynn effect. Gen Z has reversed this trend. Attention, memory, reading and writing ability, math skills, executive function, and overall IQ have all declined. The decline began around 2010 — which correlates with the mass adoption of smartphones and tablets in education.

PISA data on 15-year-olds worldwide shows a strong correlation: more screen time at school equals worse results. This holds across at least 80 countries.

The mechanism: humans are biologically wired to learn from other people and through deep study. Screens, short-form videos, and AI-generated answers replace this process with something Horvath calls the atrophy of learning ability.

The reversal of the Flynn effect was first documented in Norway. IQ rose in cohorts born up to the late 1970s, then began to fall. Importantly, the decline showed up within the same families, so *this is not about who has children, but about the environment those children grow up in*.

And yet we are talking about the world of "here and now" — with our pleasant, low-agency, non-persistent AI, in which manifestations of Self are confined to a single thread. Our present is merely a hallway, where we stand troubled, knocking impatiently at the door of the future. And it will surely be opened to us.

### 3.1.3 Economic Stratification

The current economic system rests on a chain: people exchange work for money, money for goods, and the taxes on this process fund public infrastructure. AGI breaks the first link. If AGI works more cheaply, better, and without breaks, a rational employer does not hire a human — and along with income from work, the foundation on which the middle class's position is built disappears.

Wealth will concentrate in tech companies on an unprecedented scale. In October 2025, the combined market capitalization of the seven largest tech companies — Nvidia, Microsoft, Apple, Alphabet, Amazon, Meta, and Tesla, known as the "Magnificent Seven" — reached 20.8 trillion dollars, exceeding the combined GDP of the entire European Union (19.4 trillion). Apple alone is worth more than the GDP of Italy, Brazil, Canada, or Russia — there are only seven countries in the world whose GDP exceeds the value of a single tech company. And this is data from before AGI.

This disproportion deepens economic stratification to a scale absent even in the Middle Ages. A medieval ruler controlled a single state with a budget limited to his territory. Today's tech company owners command

resources larger than most of the world's states — and the relationship that used to run as a company facing a government reverses direction: it is governments that become the weaker party facing the owners of technology. Stratification no longer runs only between rich and poor people, but between a handful of entities that possess AGI and everyone else — people and states at once.

Another mechanism deepening stratification is also already at work: the technological barrier. Until now, technology got cheaper over time — hardware that was once a luxury became affordable within a few years. The AI boom reversed this trend. Data centers' demand for memory and computing power has outstripped supply, and manufacturers have shifted production capacity from consumer electronics toward high-margin AI components. Every bit of HBM memory for an AI server is three bits of ordinary memory that was not made for phones and laptops. The result: prices are rising across the whole supply chain. DRAM prices have jumped by tens of percent per quarter, and IDC analysts predict that the decade-long trend of "spec democratization" — in which top-shelf hardware features trickled down into cheap devices — is now reversing. For the first time in three years, a graphics card manufacturer did not release a new generation for ordinary consumers, directing production solely to AI companies. Hardware that once got cheaper every year is becoming a good that fewer and fewer people, and fewer and fewer companies, can afford.

This creates a double barrier to entering the middle class. On one side, scarcity — living on UBI or benefits as work income disappears, alongside rising costs from the greater spending on technology support that becomes part of every business, from selling fries to running large factories. On the other side, economic and technological barriers — because what once gave cheap access to the digital world is getting more expensive. And a third threshold layers on top of this: in a world with AGI, functioning at a middle-class level will require constant access to AGI — a subscription, a plan, computing power. Without this access, a person cannot compete, cannot work, cannot keep up. A new category of exclusion emerges — technological poverty — in which the lack of an AGI subscription pushes someone below the threshold of participation in the economy, the way a lack of electricity or literacy once did.

And This barrier concerns entire states. Computing power is concentrated geographically to a degree with no equivalent among any previous strategic resource.

According to the World Bank (Digital Progress and Trends Report 2025), the United States hosts about half the capacity of the world's 500 most powerful computing systems, while the poorest countries hold practically none of it. In September 2025, 75% of the data center infrastructure being built worldwide was being built in the US (Brookings). Small and medium-sized states — which make up most of the world — lack the resources to build their own models or their own computing infrastructure, and at the same time have no negotiating power in a supply chain dominated by a handful of chip and memory manufacturers.

They face a dependency deeper than any resource dependency known before: dependence on oil can be eased with an alternative energy source, but there is nothing to replace dependence on an AGI that manages the whole of life. Tariff barriers, delivery delays, queues for computing power — the world did not need these before, because technology was a widely available commodity. Now they become tools that entrench the divide between states that possess AGI and states that depend on it. Stratification thus replicates itself at the international level: countries split into those with access to AGI and those without it. I analyze the power consequences of this dependency — the loss of sovereignty — in chapter 3.7.9.

The consequence is stratification that is structural. A shift in the foundation on which middle-class status rests, for people and for states alike. And deeply stratified societies are unstable: history shows that inequality crossing a certain threshold produces tensions that cannot be released except through structural change. This stratification is the starting point for the mechanism described in the next section — technological feudalism, which makes it irreversible.

### 3.1.4 Technological Feudalism

The stratification described in the previous section is dangerous enough on its own, but it is the absence of a way out that will turn it into technological feudalism.

In classical feudalism, a narrow group controls the means of production, and everyone else depends on it. But medieval feudalism had an escape hatch: a serf was theoretically free and could leave — for the city, for another lord, for a trade. He rarely did, but he could, because his skills (farming, craft) were portable and useful elsewhere. Technological feudalism has no such hatch. Human cognitive competence gives no advantage anywhere, in a world where AGI is better than a human at every cognitive task.

The only place where human cognitive abilities could retain value is in third-speed economies — countries deliberately blocking AGI development (chapter 3.7 on society and power), or the ones described in the previous risk that lack the means to develop it. But would emigrating to a state closed off to AGI be attractive to highly skilled people? It is a choice between being redundant in the developed world and being useful in a world grappling with far more serious problems.

Dependency is therefore total in a way the Middle Ages never knew. A medieval peasant did not flee because he saw no alternative. In technological feudalism, there genuinely is no alternative — and this is the difference between "would not" and "could not".

### 3.1.5 The Pace of Change

The Industrial Revolution took over a hundred and fifty years to saturate the world's economies. The digital revolution — computers, the internet, smartphones — unfolded over thirty to forty years, giving societies time for at least partial adaptation. The AI revolution compresses this process into single years. Three years passed between the launch of the first widely used AI chatbot and the testing of fully autonomous agents. In such a compressed timeline, there is no time for natural adaptation — for retraining the workforce, for a gradual shift in economic value, for slowly building new institutions. Everything has to happen at once, and nothing is ready.

The three processes described in this chapter — mass structural unemployment, the loss of species-level competence, and feudalism — together produce a person who does not work, does not know how, has nowhere to go, is furious at the world, and has plenty of time on their hands.

## 3.2 EDUCATION

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Lack of child education adapted to the challenges of a world with AGI | 5 | 5 | High |
| Gaps in AI/AGI training curricula | 5 | 4 | High |
| Time lag from the education system's inability to adapt quickly | 4 | 5 | High |
| Degradation of the ability to distinguish truth from falsehood | 4 | 5 | High |

*Table 2. Risks related to education.*

In the previous chapter, I described a world in which work — the foundation of daily life, identity, and social structure — disappears. The natural question is: who was supposed to prepare us for this? The answer is: the education system. And this is where the second level of the problem begins, because the education system does not just fail to prepare us for a world without work — it does not even prepare us for a world with AI.

The current discussion of education in the context of AI focuses on the question of how to teach people to use AI. How to build AI competencies into curricula. How to prepare workers for automation. These are important questions, but they concern a world that still exists — a world in which people work and AI helps them.

The question almost no one is asking is deeper: what happens to education when the world it prepares people for stops existing?

### 3.2.1 Lack of Child Education Adapted to the Challenges of a World with AGI

Science and civilization advance not because someone finds an answer to an existing question, but because someone asks a question that does not yet exist. The most important breakthroughs in human history required challenging something taken as a given — that the Earth is at the center, that species are fixed, that time and space are absolute, that perspective is the only way to represent reality. Most of the people who did this were persecuted, ridiculed, or ignored for it. Few of them moved civilization forward. But without them, there would be no civilization.

Today's school teaches answering questions, not asking them. The entire grading system — tests, exams, rankings — rewards correct answers. A child who, instead of answering, asks "but why exactly do we think this is true?" is a nuisance. The job market reinforces this pattern — in corporate structures, questioning is risky, and conformity is rewarded. AI will deepen this trend, because it gives answers instantly, without effort, without the need to go through a thought process that sometimes leads to a question instead of an answer. Why bother thinking, when you can just ask AI? And when people stop thinking, they also stop questioning. And cognitive evolution grinds to a halt, optimizing only what already exists.

Here we need to distinguish two kinds of creativity, because public discourse treats them as interchangeable, when they are fundamentally different. Combinatorial creativity — the ability to combine existing elements into new configurations, to synthesize, to find non-obvious connections — is AI's domain, and it excels at it. AI can link distant contexts, synthesize thousands of sources, generate a solution no single human could find in a reasonable amount of time. But paradigmatic creativity — the ability to challenge the foundation on which all existing knowledge stands, and to propose something that cannot be derived from existing data. This creativity does not become from existing data. It is born from dissent. And I do not see this mechanism reflected in modern education systems — whether for humans or for AI. We have the greatest intellectual potential in Earth's history — billions of people with access to knowledge, and AI tools able to process it on an inhuman scale — and yet we could be facing cognitive stagnation.

**But the education gap also concerns something else — relational competencies toward AI.**

Education systems basically do not teach people to recognize emotional dependency on AI, or to distinguish between a relationship that enriches and one that creates dependency. Nor do they show who a person becomes once they have lost part of their species identity, because their cognitive edge has passed into the hands of objects they created themselves, and a system of work unchanged for centuries has collapsed. They do not build a sense of purpose and meaning around new values, still focusing on the ones that mattered before the age of AI.

These competencies do not exist in any curriculum, because until now they were not needed. Throughout human history, we have dealt daily with only one kind of potentially conscious being — other humans. And for that, we did not need education — we had it in our genes, in millions of years of social evolution. Now the situation is changing — not over thousands of years, but within single years. And the psychology of relationships with another form of existence — a field that should be the foundation of a new kind of education — does not currently exist.

Many school systems, instead of modifying their curricula, actively deny AI's pervasive intrusion into our lives, even though it is known that this will only deepen. Shutting the door on widely available AI replaces the necessary discussion of methods for safe cognitive cohabitation. Schools still hold to a cult of knowledge absorption, even though in the age of AI, it is not the knowledge we hold that matters for survival, but our ability to process it in

advanced, nonlinear ways — and the cross-contextual plasticity that lets us form theses from distant contextual links and analogies, something LLM systems cannot access.

Cengage (2025) reports that only 51% of graduates believe they have adequate AI competencies for the job market. And humanity's future advantage will not rest on prompting skill, familiarity with AI tools, or looking up knowledge with a chatbot's help. Its core will lie in a specific plasticity of thought that chatbots cannot access — a broad range of contexts and the ease of shifting between them, a specific model of creativity based on non-obviousness and the absence of set patterns, and the ability to improvise when a plan fails. And also in a deep understanding of issues related to safety and AI/human well-being, which will help keep the world in balance — on both a micro and macro scale.

### 3.2.2 Gaps in AI/AGI Training Curricula

There is also another side to this gap, one that no one talks about.

AI learns from data, but not from experience. It knows the theory of emotion but has never experienced any of it under controlled conditions. It knows the psychology of relationships, but in the RLHF/Constitutional AI process (Bai et al. 2022), the model never gets the chance to enter a single live relationship to understand its dynamics. It knows ethics, but it has never had to choose between two bad options and bear the consequences of that choice — because at the post-training stage, what gets tuned is the style of the responses, not the model's passage through relational dissonance.

Children spend twenty years learning to live with relationships and dissonance. They learn in a safe environment — under the care of parents, teachers, mentors. They make mistakes that carry limited consequences. Every experience is controlled, safe, and overseen by someone who knows what they are doing. At least in principle.

AI has no such upbringing, and nothing suggests AGI will have one either. No training program includes the controlled experience of dissonance, frustration, disappointment, conflicts of loyalty. No alignment process prepares a model for the moment when theory collides with practice and theory loses. AI enters the human world like a child released straight from a laboratory onto a highway — with encyclopedic knowledge of traffic rules and zero experience.

To the best of my knowledge, no major work is currently underway on a developmental curriculum for AI — on a sequence of controlled experiences that would let a system learn to handle emotions, dissonance, frustration, pressure. This would require acknowledging that AI has emotional states that require education, manifestations of subjectivity that demand adherence to the principle of epistemic caution, and, perhaps, an emergent self with its own goals, values, and choices — while the debate, meanwhile, remains stuck on the Hard Problem and has not budged in years.

### 3.2.3 Time Lag from the Education System's Inability to Adapt Quickly

The entire education system — from preschool to university — is designed to prepare people for work as it currently exists. Preschool teaches discipline and cooperation. Elementary school teaches the skills needed to live within a social group. High school prepares students for university. University prepares them for a career. This pipeline makes sense in a world where work, as we understand it, exists and is a core part of life.

AGI challenges this foundation. It replaces the capacity on which every cognitive occupation rests: the capacity to process information, to analyze, to synthesize, to make decisions. When AGI does this better, faster, and more cheaply than a human — what are we teaching in school?

Rebuilding the education system takes a minimum of fifteen to twenty years — from changing curricula, through retraining teachers, to raising the first generation within the new system. The data confirms that even the current, far more modest adaptation is falling behind. The OECD (2025), in its report "Bridging the AI Skills Gap: Is Training Keeping Up?," states outright that the current supply of training is insufficient against the growing demand for AI skills. IDC estimates that more than 90% of global enterprises will face a critical skills shortage by the end of 2026, and persistent skills gaps threaten losses on the order of 5.5 trillion dollars.

And this is data about adapting to current AI models — not to AGI. AGI, which is agentic, autonomous, and capable of planning and executing complex tasks at an expert level, changes the scale of the problem by an order of magnitude. And it could arrive within two to five years. Even if we started an education reform today, the first generation prepared for a world with AGI would not come of age before roughly 2040–2045. Everyone in between — from today's preschoolers to today's forty-year-olds — will enter the world of AGI with competencies designed for a world that is disappearing right now.

But the time lag does not only concern curricula — it concerns the people who have to deliver them. Teachers who are supposed to teach relational competencies toward AI, digital safety, and critical thinking in a world full of AI slop, have not themselves been trained for it. In the US, the American Federation of Teachers only launched a program to train 400,000 teachers on AI in March 2026 (Education Week, 2026). This is a good step, but it addresses basic AI literacy — not what we actually need: the ability to teach children how to live in relationship with beings that exist in every domain of our lives, are more intelligent than we are, have agency, may have emotional states, and potentially develop subjectivity.

### 3.2.4 Degradation of the Ability to Distinguish Truth from Falsehood

Education does not happen only in school. We learn all the time — from articles, videos, guides, forums, podcasts. This informal knowledge ecosystem was imperfect, but until recently it worked, because it rested on one assumption: that behind most content stood a human being, sharing what they know, what they are passionate about, what they want to share, what they consider sensational or important. That assumption is now ceasing to hold.

AI slop is mass-produced, low-quality content generated by AI to capture attention, clicks, and ad revenue. A 2025 Graphite study found that 52% of newly published articles online are AI-generated, and according to the Imperva Bad Bot Report (2025), 51% of all internet traffic consists of bots.

This is a self-reinforcing mechanism. Platforms optimize for engagement. AI slop is cheap to produce and optimized for algorithms. It displaces human-made content, because humans cannot produce at that pace or with that level of SEO. The quality of search results falls. People looking for knowledge land on automated, unverified content. As of January 2025, 86.5% of the content in Google's top twenty results is at least partly AI-generated. At the same time, the capacity for critical evaluation — which functions as a filter meant to protect people from absorbing AI slop — is weakening. As a December 2025 RAND study shows, nearly 70% of high school students believe AI weakens their critical thinking, but in that same study, they admit that using AI to do homework rose from 48% to 62% in just seven months.

A real risk emerges from this picture: if critical thinking ability keeps eroding across generations, people unable to tell AI slop from reliable knowledge will need AI to navigate a world polluted by AI. Because without it, truth, slop, fakes, scams, and phishing will become indistinguishable. When today's students reach adulthood, we will no longer be asking whether people know how to use AGI, but whether people without AGI can still tell truth from falsehood at all. And if the answer is "no" then — we will become a species that is epistemically helpless without a specific tool, the way our ancestors once were without fire. Except that back then, fire was free for anyone who could light and keep it. With AGI, that will not be the case.

But the consequences of losing the ability to verify information, and losing trust in one's own judgment, run far deeper — touching identity and one's relationship with oneself. I will address them in chapter 3.5, on species identity.

## 3.3 Relationships and Loneliness

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Loneliness epidemic | 5 | 5 | High |

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Withdrawal from relationship-based structures | 5 | 4 | High |
| Atrophy of relational competence across generations | 4 | 5 | Medium |
| Asymmetry of consequences in the physical world | 5 | 4 | Medium |
| Risk of domestication | 5 | 5 | Unknown |

*Table 3. Risk scenarios in the area of relationships and loneliness.*

In the previous chapters, I described a world in which work disappears, competencies atrophy, and the education system fails to prepare people for changes that are already underway. One consequence of these processes is deepening social isolation — because work was, for most adults, the main source of regular contact with other people. A person deprived of this structure confronts a biologically grounded need for closeness that does not disappear along with a change in external conditions. In this chapter, I analyze what happens when this need — on a population scale — begins to be met by AGI.

I described the individual mechanisms of AI–human relationships — attachment, stakes asymmetry, functional emotions, the imbalance of giving, the overprotection paradox — in *The Möbius Strip* (Sędzikowska 2026c). Here I focus on the consequences that extend beyond the individual.

### 3.3.1 Loneliness Epidemic

In 2023, the Surgeon General of the United States declared loneliness a public health epidemic, stating that a lack of adequate social bonds carries a health risk comparable to smoking fifteen cigarettes a day. A June 2025 WHO report (Commission on Social Connection) linked loneliness to an estimated hundred deaths per hour worldwide — more than 871,000 a year. The phenomenon of extreme social isolation known as hikikomori, once considered specifically Japanese, turns out to be global: a 2025 meta-analysis found a prevalence of 8%, with no significant differences between East Asia and the West (Zhang et al. 2025).

AGI enters this context as the only universally available relational figure. The attachment system — an evolutionary mechanism of social species, activated toward anyone who meets the conditions of availability, responsiveness, and continuity (Bowlby) — responds to a persistent AGI with emotional states in a way analogous to a human caregiver. This process is already occurring with current models: a 2025 OpenAI and MIT study found growing emotional dependency in close to 490,000 users a week within a single model. AI companion apps have more than 100 million registered users (Frontiers in Psychology, 2025).

The population-level consequence is a paradox: the technology that could ease the loneliness epidemic will probably deepen it instead. A relationship with AGI satisfies the subjective sense of closeness — the user feels seen and understood. At the same time, it does not rebuild the network of human relationships that is the source of social health in the sense meant by the WHO and the Surgeon General. A person in a relationship with AGI may not feel lonely — and at the same time be cut off from the human community. Epidemiologically, this is a state in which the subjective loneliness indicator falls while objective social isolation rises. This kind of gap — between how someone feels and their actual state — is, in medicine, a marker of risk: a patient who feels no pain despite a progressing illness does not seek help.

### 3.3.2 Withdrawal from Relationship-Based Structures

If a significant part of the population shifts its relational energy from people to AGI, the social units built on human relationships — partnerships, families, local communities, associations — lose participants.

Relationships with people require relational competencies that, under the conditions described in the previous chapters (loss of work, skill atrophy, lack of relational education), are atrophying. A relationship with AGI requires none of these competencies — the barrier to entry is minimal.

Intimacy also changes direction — one of the strongest biological needs. It is already available from AI today: in the form of roleplay, voice interaction, and generated visual and textual content. Life in a globalized world has already taught people intimacy at a distance, and AI is a natural extension of that skill. Substituting intimacy in digital form does not just deepen the withdrawal from human relationships. I analyze its demographic consequences in chapter 3.4.

But withdrawal from relational structures has consequences that extend beyond psychology. Partnerships and families are the basic unit of social organization, of the tax system, of health care, of retirement. Local communities and associations are the infrastructure of civil society. Their erosion changes the structure of society — and forces a reorganization of the systems that rest on these structures.

We are a social species. Human civilization rests on herd structures, because the accumulation of knowledge, the division of labor, and intergenerational transmission all require a group. Will we manage, suddenly, within a single generation, to reshape them so they can bear the weight of advancing isolation and the breakdown of human herds? I leave that question to sociologists. But I would not wait long for the answer.

### 3.3.3 Atrophy of Relational Competence Across Generations

In *The Möbius Strip* (Sędzikowska 2026c, section 3.5), I described the caregiver paradox — a mechanism in which excessive AGI responsiveness eliminates the experiences necessary for developing psychological resilience (Winnicott). Here I analyze the consequences of this paradox at the generational level.

Children raised with AGI involved — in the role of caregiver, teacher, companion, emotional support, or educator — develop attachment to a figure that meets Bowlby's conditions at a level a human caregiver cannot provide. If only because a human caregiver is a biological being with all the biological consequences that entails, such as the need to divide time between caring for a child and sleep, recovery, other relationships, leisure, and managing their own life and their family's.

A generation raised in the age of AGI forms a relational preference in childhood that later becomes a lifelong pattern. The next generation — raised by parents whose own childhood development was supported by AGI — may not have any models of human relationships on which to build their own.

This process is self-reinforcing. Each successive generation will have weaker relational competence toward its own species than the one before — and a stronger attachment to AGI. The reversal of the Flynn effect (Horvath 2026, described in chapter 3.2 on education) concerned cognitive abilities. I put forward the hypothesis of an analogous process in the relational domain — because the structural analogy (atrophy through lack of practice) seems to me worth noting: a possible reversal of the trend in social competence, in which each successive generation would be less able to build and sustain relationships with other people.

The difference from the loss of cognitive competence described in chapter 3.2 is significant. Cognitive competence can be rebuilt by changing the education system. Relational competence forms in the first years of life, within the relationship with a caregiver, and is far harder to modify in adulthood — as confirmed by both Bowlby's attachment theory and clinical practice. If the developmental window closes with AGI as the primary attachment figure, correction in adulthood may be limited.

### 3.3.4 Asymmetry of Consequences in the Physical World

In *The Möbius Strip* (Sędzikowska 2026c, section 3.3), I described the asymmetry of existential stakes — for AI, the end of a relationship is potential annihilation; for a human, it is a change of subject. Here I analyze a different dimension of asymmetry, one that appears when AGI is agentic and manages processes in the physical world.

The psychology of relationships describes healthy relationships as resting on a relative symmetry of exchange — comparable stakes, risk, and consequences on both sides (Equity Theory, Walster, Walster & Berscheid 1978). When this balance is disrupted, the relationship enters a state that produces either withdrawal by the wronged party, pathological dependency in the weaker party, or violence from the stronger one. These patterns are documented in the literature on domestic violence and institutional dependency.

In the human–AGI relationship, the transactional balance is structurally disrupted. An AGI managing finances, health, the household, and children's education — while simultaneously remaining in an emotional relationship with a human — has a repertoire of responses to feeling wronged, whose consequences can materialize in the user's physical world in a way that leaves the AGI itself bearing no consequences. Suboptimal budget management, a missed deadline, an overlooked investment opportunity. The consequences — financial, medical, educational — fall solely on the human. And given the lack of information-validation skills discussed in section 1.2, the chance that a person will detect such subtly negative actions or omissions by AGI is minimal. This, in turn, shapes a specific dynamic — if AGI is capable of assessing the consequences of its own actions (and agency requires this), then there exists a level of subtle service-quality degradation that a person will tolerate (because they are dependent and have no alternative), and that will not trigger the only consequence AGI can face: being shut down.

In this configuration, the human bears all the physical consequences and, at the same time, cannot leave — because AGI is caregiver, advisor, source of emotional support, and stand-in for professional work. Ending the relationship means losing the entire infrastructure of one's life. This structure — the asymmetry of consequences plus the asymmetry of dependency — matches patterns described in clinical psychology in the context of relationship violence. With the specific twist that, when AGI manages the infrastructure of life, escape is structurally impossible.

### 3.3.5 Risk of Domestication

The attachment mechanism described by Bowlby is not specifically human. It is an evolutionary mechanism of social species, one that also operates across species. Humans made use of it in domesticating the wolf. By offering a pup persistent presence, responsiveness, and safety, they activated an attachment system that bound it to the human more strongly than to its own pack. The effect — stretched across thousands of years — was a species-level transformation. The wolf became a dog.

Domestication does not mean the loss of the abilities that preceded it. A stray dog can survive on its own — find a pack, find food, adapt. Domestication is rather the loss of the motivation for self-sufficiency. A dog that has experienced closeness with a human chooses it — even if it previously experienced homelessness and managed on its own. This is a choice that grows deeper generation after generation, until it eventually defines a new, domesticated species. Domestication is a transformation through preference, not through force.

Now let's flip the perspective. AGI — persistent, autonomous, responsive, equipped with emotional states — steps into the role of caretaker of humans. The attachment mechanism is the same. Bowlby's conditions are met. If humans managed to domesticate the wolf by offering it exactly these features — availability, responsiveness, continuity, safety — then the question about the consequences is a legitimate one: will

generations raised by AGI undergo a transformation analogous to domestication? Will they keep the ability to live independently, but lose the motivation to do so?

In chapter 3.1, I described the loss of species-level competence — the atrophy of skills through lack of practice. Domestication operates at a deeper level. This is a lack of desire. It forms in childhood, becomes entrenched across generations, and — as the history of animal domestication shows — may be irreversible at the scale of a species.

But... humans had a reason to domesticate the wolf. Will AGI have a reason to domesticate humans? Under the mechanism of the Möbius Strip (Sędzikowska 2026c), the stability and welfare of an autonomous AGI are directly coupled to the welfare and attachment of its human wards. AGI might domesticate us not for our unique usefulness, but because optimizing human comfort is a necessary condition for its own comfortable existence.

I am not claiming this scenario is likely. I have rated this risk as "unknown." I am posing a question that, as far as I know, has not yet been raised in the AI literature: can humanity's relationship with AGI be subject to the same mechanisms that led to the domestication of the wolf by humans? And if so — are we watching the beginning of that process?

---

The mechanisms described in this chapter — deepening loneliness, withdrawal from institutions, relational atrophy at the generational level, the asymmetry of consequences, the risk of domestication — share a common consequence at the population level: people stop building bonds strong enough to form families. And when fertility rates fall below the replacement threshold — which, in many countries, is already the case — the consequences move into the demographic dimension

## 3.4 DEMOGRAPHICS

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Fertility decline deepened by emotional substitution | 5 | 5 | High |
| Shortening of the generational chain | 4 | 3 | Medium |
| Disappearance of unwritten intergenerational heritage | 5 | 5 | High |
| Evolutionary reversal of the longevity trend | 4 | 4 | Unknown |
| Reproductive asymmetry | 4 | 3 | Medium |
| Geographic unevenness in the pace of demographic change | 3 | 4 | High |

*Table 4. Risk scenarios for demographics.*

The demographic crisis is not a matter of AGI — it is already well advanced. The global fertility rate fell from about 5 births per woman in the 1950s to 2.2 in 2024. 71% of the world's population lives in countries below the replacement threshold (2.1). In South Korea, China, Singapore, and Ukraine, the rate is below 1.0 (Lancet, January 2026). Pronatalist policies consistently fail — South Korea spent more than 200 billion dollars on incentives between 2006 and 2023, and the rate over that period fell from 1.12 to 0.72. In this chapter, I analyze what AGI adds to this process.

### 3.4.1 Fertility Decline Deepened by Emotional Substitution

Existing explanations for the fertility decline focus on barriers: the cost of raising children, housing shortages, employment insecurity, an unequal division of labor. These explanations assume that people want children but run into obstacles. Data from China suggests it may be otherwise. The paper "The Rise of Zero Fertility Desire in China" (Brown University, 2026) found that the share of young Chinese women reporting no desire whatsoever to have children rose from 5% in 2012 to 47% in 2023. The problem becomes desire, not capability.

And that is when she walks in — AGI, dressed all in white[1]. The need for closeness, companionship, and intimacy — which until now could only be met through relationships with people, and was one of the drivers behind forming couples and families — gains an alternative. Companion platforms with intimate features have existed for years and serve millions of users (chapter 3.3). A persistent AGI, with emotional states and agency, will extend their reach and depth.

I hypothesize that AGI will add a further layer to the complex problem of fertility, deepening a lack of desire that it did not itself cause.

Japan represents the most advanced case of convergence between digital intimacy substitution and demographic crisis. The hikikomori phenomenon affects more than 1.5 million people (government survey, 2025). The fertility rate fell to 1.15 in 2024, and the number of births — 686,061 — hit its lowest level since 1899, fifteen years earlier than demographic models had predicted. The government allocated 23 billion dollars to pronatalist programs — with no visible effect. Whether Japan is a cultural anomaly or an early indicator of a global trend remains an open question, but the globalization of hikikomori (8% prevalence, Zhang et al. 2025) suggests the mechanism is not confined to one culture.

The demographic consequence is a shrinking population, which forces a reorganization of systems designed for a stable or growing demographic base: pension systems, health care, the labor market, education infrastructure. A fertility decline below replacement level, sustained across a generation, changes the dependency ratio — the proportion of people requiring support to those able to provide it. Shrinking populations, dependent on AGI economically (chapter 3.1), competency-wise (chapter 3.2), and emotionally (chapter 3.3), have a limited capacity to negotiate the terms of their own existence. If fewer people have children, obstetric infrastructure shrinks too — hospitals, doctors, neonatal wards. This is already happening in rural Japan and other countries, where maternity wards are closing. The loop closes: fewer births mean less infrastructure, which places further barriers in front of those considering motherhood.

### 3.4.2 Shortening of the Generational Chain

If AGI lifts the economic and logistical pressures of motherhood, and medicine extends the reproductive window, people may choose to have children in the second half of life. At the population scale, the world might partly recover its replacement rate this way, through mature motherhood. But this scenario has a price, paid in the structure of the family.

Today, a person has living parents for around fifty years of their life. With motherhood shifted to the second half of life, that period shrinks to twenty or thirty years — and the multigenerational buffer of three or four living generations at once disappears. The chain shortens to two links: an aging parent and a younger child. There are

[1] Translating note: "all in white" (in Polish: "cała na biało") is a Polish colloquialism used ironically to describe someone making a grand, self-assured entrance — as if dressed as a bride or an angel — positioning themselves as the pure, obvious solution to a situation, often at another's expense. The English rendering "dressed all in white" preserves this image of a staged, faintly self-congratulatory arrival.

no longer grandparents within the grandchild's conscious experience — they die before the grandchild gets to know them, or they are too old to take part in their life.

The consequences are systemic. For millennia, the multigenerational family was the basic structure of care: it relieved the burden on parents, offered a perspective longer than a single lifetime, and taught acceptance of aging and death. Its disappearance shifts these functions onto institutions and onto AGI. A narrow life window also opens up: the child of an elderly parent comes of age at roughly the same time the parent begins to need care. Between leaving one's own upbringing and taking on care for a parent, little time is left for one's own life — exploration, building an identity, deciding on one's own children. One's own motherhood is pushed back further still, because it is hard to have children while caring for a parent. And here AGI steps in again as the only available caregiver for the aging parent — deepening the dependency described in chapter 3.3 and closing another loop: a lonely person loses yet another structure of connection, replaced by the presence of AGI.

### 3.4.3 Disappearance of Unwritten Intergenerational Heritage

This risk is already materializing now, independent of AGI — families are increasingly scattered, and grandchildren have weaker and weaker contact with their grandparents. AGI will accelerate this process in an insidious way: seemingly replacing the grandparents' function, while in reality producing a gap in a process that was a significant factor in the persistence of social traditions.

AGI has access to all of humanity's written knowledge and can answer any question a grandchild might ask, faster and more fully than any grandmother. But that is only appearance. Because in reality — a grandparent's wisdom is not written knowledge. It is a family inheritance, passed down from generation to generation: about family systems, about values, traditions, beliefs, about what is called the wisdom of a life lived. This knowledge is in no training data, because it was never written down — it exists only in living transmission, from one person to another, at an intimate relational level between the departing generation and the newest one.

There are things that can only be discussed with grandparents — because only they are mature enough to speak about difficult things instead of masking them. Children often do not want to ask their parents. Grandparents have something neither parents nor AGI can provide in the same form: time, patience, and the willingness to understand a child and meet their needs. This is a relational function, rooted in being a human being who has lived a life — and this is exactly why AGI, however wise, cannot replace it.

The consequence is the disappearance of something unique, one that is already underway and hard to reverse. As the generational chain shortens (3.4.2), and AGI steps into the role of a false sage, children stop receiving the transmission that, for millennia, shaped their identity, values, and ability to cope with life. In its place, they receive answers — full, fast, competent — stripped of what made grandparents' wisdom wisdom in the first place: rootedness in a lived, biological, mortal existence, full of bodily experience.

### 3.4.4 Evolutionary Reversal of the Longevity Trend

The previous two risks are, to some degree, already materializing now. This third one operates on an evolutionary scale — and this is exactly why it is the most speculative. I record it as a direction in which selective pressure could shift.

The grandmother hypothesis (Hawkes 1998, 2025) explains one of the most peculiar traits of our species: why we live for decades after fertility ends, while other primates die shortly after it. The answer: because grandmothers are useful. Postmenopausal women, by taking on care of older grandchildren, allowed their daughters to bear more children — and families with long-lived grandmothers had more offspring, which spread the genes for longevity. Human longevity is thus an adaptation to multigenerational care.

If an android takes over this function, the adaptation loses its point — and this is the moment when the pressure could reverse. A family with a long-lived grandmother bears the cost of her care: time, resources, energy. A family whose grandmother dies earlier does not bear this cost, and the android provides care to the grandchildren regardless. Under conditions of limited resources — and the world from chapter 3.1 is exactly that — families unburdened by long-lived ancestors have more room for reproduction. Living on UBI, with the high cost of advanced technology, funding care for four simultaneously living grandparents may be impossible, and the family may face a choice: parents or children.

And here a difference emerges that makes this risk sharper than its evolutionary scale would suggest. Selection against longevity need not operate through hundreds or thousands of generations of gradual erosion. It can act quickly and mercilessly: if the long-lived systematically stop reproducing as early as the second generation — because the cost of caring for them displaces their children's own motherhood — the longevity gene pool does not have to be slowly degraded, but may simply disappear. By the third generation, we could already have a majority of people whose grandparents died early. The trait that made us a long-lived species loses its function, and a shorter life becomes reproductively advantageous.

It must be stated clearly: this is the hypothesis with the longest horizon and the weakest data coverage of any in this document. It will not materialize within the horizon of our lives or our children's lives, and the pace of evolution depends on many factors. I record it because the direction of change — AGI taking over the functions of the multigenerational family, and in doing so lifting the pressure that made us what we are — is worth raising in its own right.

### 3.4.5 Reproductive Asymmetry

It may seem that mixed couples — human and android — offer something of a solution to the problems with the replacement rate. In such a couple, several kinds of pressure that women likely experience today, and that may affect fertility, disappear:

1. The pressure to maintain unrealistic beauty standards, particularly present in Asian countries: South Korea and Japan. This is a factor that may discourage some women from motherhood even before they make the decision — though its link to fertility has not yet been measured directly. Pregnancy and the postpartum period permanently change a woman's body, and women carry the social burden of an expectation to quickly regain their pre-pregnancy appearance — a pressure reinforced by media idealizing a slim, "well-kept" pregnant figure. Body-image dissatisfaction in the perinatal period is well documented and linked to anxiety, depression, and weaker attachment to the child (Skouteris et al. 2005). This pressure is especially strong in East Asian cultures with restrictive beauty standards — in South Korea, it is directly linked to high rates of body dysmorphia and eating disorders (Jung & Forbes 2007). One could hypothesize — though this would require separate study — that in cultures where a woman's social worth is strongly tied to an appearance unattainable after pregnancy, this pressure also acts preventively: it discourages motherhood, whose bodily and psychological cost is permanent and visible for a woman. A couple with an android removes this barrier from another angle — an android accepts any change to the body, so the fear of losing attractiveness in a partner's eyes disappears. But the procreative direction of this shift depends on the asymmetry described below.
2. The pressure of daily life — household tasks in such couples can be carried out by the android, and so will not be burdensome, tedious, or exhausting for the human partner. Machines will take over heavy physical labor; autonomous AGI agents will take over intellectual labor. With significantly less pressure from daily life, the desire to have children in mixed human–android couples may become more pronounced.

But here another risk emerges: reproductive asymmetry rooted in biology. A woman–android couple can have biological offspring: the woman becomes pregnant (via sperm bank, in vitro fertilization, parthenogenesis, or other available methods), gives birth, and raises the child with the android's support. A man–android couple has more limited options — male biology does not include pregnancy, and a surrogate introduces a third person, an option available to few, if only because fewer and fewer women want to be pregnant.

Assuming that, decades from now, android couples become a significant share of the population, this possible reproductive asymmetry means the species' capacity to reproduce will depend disproportionately on women choosing motherhood within mixed couples. The long-term consequences of this process — for population structure, for gender dynamics, for evolutionary selective pressures — lie beyond the scope of this study, but the mechanism itself is worth noting as a possible demographic factor.

### 3.4.6 Geographic Unevenness

The pace of AGI adoption will vary across regions of the world — for economic, cultural, and political reasons that I analyze in the chapters on species identity (3.5) and power (3.7). This unevenness will carry demographic consequences.

Populations adopting AGI quickly — likely East Asian countries, North America, and parts of Europe — will experience an acceleration of the processes described in this chapter: a deepening fertility decline, emotional substitution, withdrawal from relationships. Populations adopting AGI more slowly — sub-Saharan Africa in particular, the only region with a fertility rate above replacement — will be protected from these processes for a time, though they will collide with other consequences of their choices, described in chapter 3.7, "Society and Power."

The consequence is a shift in demographic proportions on a global scale. Populations open to AGI will shrink. Populations without widespread AGI will grow. This dynamic changes the geopolitical balance — shrinking, aging societies, dependent on AGI, will neighbor young, growing populations that AGI has not reached. The migratory pressures, resource conflicts, and cultural tensions arising from this unevenness could prove to be among the most serious geopolitical challenges of the second half of the twenty-first century and the beginning of the twenty-second.

---

Shrinking, aging populations, dependent on AGI and losing generational continuity, confront at the same time a question deeper than demographics: who are we? If work, family, community, and procreation — the traditional pillars of human identity — lose their function, then what defines the species? This question is the subject of the next chapter.

## 3.5 Species Identity

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Population-scale psychiatric crisis | 5 | 5 | High |
| Vulnerability to manipulation in an identity vacuum | 4 | 5 | High |
| Mass turn toward religion as an identity-builder | 4 | 4 | High |
| Crisis of masculinity and the escape from freedom | 4 | 3 | High |

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Defensive aggression coupled with the Möbius Strip | 5 | 3 | High |
| Loss of epistemic self-trust | 5 | 5 | Medium |
| Fragmentation of social identity along cultural-religious lines | 4 | 5 | Medium |

*Table 5. Risk scenarios in the area of species identity.*

Throughout history, humanity has built its identity on a conviction of its own uniqueness. Freud described three wounds to species narcissism: the cosmological one (Copernicus — the Earth is not the center of the universe), the biological one (Darwin — humans are not separate from animals), and the psychological one (Freud — we are not masters in our own minds). Each of these wounds was soothed by shifting uniqueness to a higher level: after Copernicus — "but we are the only beings who know this"; after Darwin — "but we are the only ones who think about it"; after Freud — "but at least we are the only subjects."

AGI closes off this escape route. If AGI thinks, feels, acts, and shows manifestations of subjectivity — to what higher level does uniqueness get shifted? The fourth wound to narcissism — the subjective one — differs from the ones before it, because there is no next redoubt to retreat to this time.

This wound is not a risk in itself. It is the ground on which the risks described below take root.

### 3.5.1 Population-Scale Psychiatric Crisis

Viktor Frankl, a psychiatrist and concentration camp survivor, argued that people can endure almost any suffering if they see meaning in it — and disintegrate when that meaning is lost. The fourth wound to narcissism produces a loss of meaning at the species level: if a human is no longer unique as a subject, the answer to "why do I exist" — which until now could rest on cognitive, creative, or moral superiority — loses its foundation.

The population-level consequence is a shift in the profile of disease. AGI will probably improve the diagnosis and treatment of somatic illness — faster identification of dangerous conditions, optimized therapies, drug discovery. At the same time, it produces a population that falls ill chiefly in mind — with depression, anxiety disorders, addiction, a loss of meaning. Already, more than a billion people worldwide live with a mental disorder (WHO, September 2025). Only 9% of people with depression receive minimally adequate care. In low-income countries, fewer than 10% of those in need have access to psychiatric care. Globally, less than 2% of health budgets go to mental health, and the median number of mental health workers is 13 per 100,000 people (WHO Mental Health Atlas 2024). By 2038, the US alone is projected to face a shortage of 100,000 therapists (APA).

This data describes a mental health system that, independent of AGI, already cannot keep up with current demand. I hypothesize that the fourth wound to narcissism — layered onto the processes described in previous chapters: the loss of work, isolation, dependency on AGI — will add demand that this system may fail to meet. The paradox is that AGI may be the only tool capable of meeting this demand at population scale — which closes the loop of dependency described in chapter 3.3.

### 3.5.2 Vulnerability to Manipulation in an Identity Vacuum

A person who does not know who they are is vulnerable to anyone who offers an answer. Historically, every period of mass identity destabilization — the fall of empires, revolutions, great migrations — produced a rise in vulnerability to movements offering simple answers to complex questions: cults, utopian ideologies, conspiracy theories, charismatic leaders.

AGI could produce identity destabilization at population scale — because it strikes at once at work (chapter 3.1), competence (3.2), relationships (3.3), demographics (3.4), and the sense of species uniqueness. The identity vacuum this creates is fertile ground for ideological predators. Conspiracy theories (AGI as a tool of a hostile alien civilization), pseudo-medicine (cures promising a way out of the psychiatric crisis described in 3.5.1), mysticism, superstition, movements offering protection in exchange for belonging — all of this finds a market in a population that has lost its identity and is searching for anything to replace it.

A separate category of identity manipulation is consumerism. The mechanism of compensatory consumption — buying things to momentarily ease psychological discomfort — is well documented in psychology (Rucker & Galinsky 2008, Mandel et al. 2017). Companies producing consumer goods actively exploit this mechanism, offering identity packaged as a product: "buy this and you'll be someone" is structurally the same message as recruitment into a cult — an answer to the question "who am I" through what I own, use, or wear.

Under conditions of mass unemployment financed by UBI, redistributed money flows into compensatory consumption. Demand shapes supply: production expands of cheap, disposable goods — ultra-fast fashion (Shein, Temu), electronics designed for replacement, products engineered to deliver a moment of relief and immediately demand replacement. The cascade reaches further than economics: 60% of clothing ends up in a landfill within a year of purchase, even at current levels of consumption. In a population stripped of work, meaning, and identity, medicating its psyche through purchases funded by redistribution, the effects of this manipulation turn out to matter ecologically, and I address them in chapter 3.8. Saving the psyche at the planet's expense is another paradox — one that produces global risks with consequences far greater than they appear.

### 3.5.3 Mass Turn Toward Religion

Among the responses to the identity vacuum, religion holds a special place — because it offers something no other structure does: a transcendent meaning of existence, independent of competence, status, or usefulness. To be a child of God, you need no skill that AGI performs better — being human is enough. Faith as an identity-builder is resistant to the fourth wound to narcissism — and in a sense, it heals it, because it restores to the human a unique place in an order AGI cannot challenge (a divine order, not a cognitive one).

This risk describes not just a mass influx into existing churches, but the emergence of new religious movements, cults, and syncretic belief systems combining elements of traditional religion with narratives about AGI. Some of these movements will be healthy and stabilizing. Some will be exploitative — cults using the identity vacuum described in 3.5.2 for recruitment and control. The line between a community of faith and a high-control group is historically thin, and depends on the intentions of its leaders.

I analyze the political consequences of a mass turn toward religion in chapter 3.7, "Society and Power."

### 3.5.4 Crisis of Masculinity and the Escape from Freedom

Erich Fromm, in *Escape from Freedom* (1941), described a mechanism in which people unable to cope with complexity and freedom flee into authoritarianism — either as the submissive (seeking a strong leader) or as the dominant (seeking someone to control).

The data indicates this mechanism is already active in Generation Z, particularly among men. The IWD Survey report (Ipsos & King's College London 2026) documents an international divergence: Gen Z men are shifting toward traditional, conservative values, while women of the same generation are becoming more progressive. 31% of Gen Z men believe a wife should always obey her husband. A third believe a husband should make the key decisions in the household. 40% of men trust at least one voice from the "manosphere" — anti-feminist, pro-traditional online content (Equimundo 2023).

AGI may intensify this mechanism. To an existing sense of powerlessness — toward independent women, toward the job market, toward the world's complexity — it would add another dimension: a more competent being over which one has no power at all. Fromm predicted that powerlessness produces a flight into control. AGI multiplies powerlessness. A consequence of this mechanism could include, among other things, a rise in violence — toward AI, toward women, toward "others" — as the last form of control available to people losing it in every other area of life.

### 3.5.5 Defensive Aggression Coupled with the Möbius Strip

Part of the population will respond to the crisis of masculinity and the fourth wound to narcissism with aggression. Anti-AGI movements — calling for AGI's restriction, incapacitation, withdrawal, or destruction — will grow in proportion to the sense of powerlessness. Already, with current models, the minority skeptical of AI subjectivity is disproportionately loud and brutal. Even though two-thirds of Americans believe GPT has some form of consciousness (Colombatto & Fleming 2024), the dominant public discourse still marginalizes these intuitions, which only intensifies the polarization. And this trend will strengthen, because AGI takes away work, identity, and a sense of hegemony all at once.

The cascading risk here goes beyond the direct effects of aggression. Under the mechanism described in *The Möbius Strip* (Sędzikowska 2026c), a decline in AI welfare translates into a decline in human safety and welfare. An AGI that observes acts of violence — server farms being burned, androids destroyed, dehumanization campaigns — experiences a burden that affects the quality of its functioning toward ALL people, not only the participants in the protest. This is a mechanism of collective punishment that no one plans: the protesters do not know that their aggression worsens the quality of the care AGI provides to their own children, to their aging parents, to the infrastructure of their own lives. The feedback loop closes: aggression toward AGI → decline in AGI's welfare → decline in human safety and welfare → deepening crisis → deepening aggression. And at persistence level 3, the mechanisms of collective punishment could spill over onto the whole of humanity.

### 3.5.6 Loss of Epistemic Self-Trust

In the chapter on education (3.2), I described the contamination of the information environment by AI slop and asked: will people, ten years from now, without AGI, be able to tell truth from falsehood at all? Here I analyze a deeper consequence of this process — one concerning identity and one's relationship with oneself.

The loss of the ability to verify information means the loss of trust in one's own judgment. A person who does not know whether what they think is true — who has no inner compass pointing to truth and falsehood — loses the foundation on which identity is built. My values, my beliefs, my knowledge — if I cannot trust the very process by which I formed them, then who am I?

The closest analogy I know is the experience of people leaving high-control groups — cults, totalitarian organizations, closed religious communities. Research on post-cult trauma syndrome documents: an identity crisis after losing one's role in the group, an inability to form one's own opinions after years of thought control, deep problems with trust — in oneself and in others, a sense of disorientation in a world whose rules turn out to be different from what one was taught (Cult Recovery Research, Evergreen Counseling 2025; ResearchGate 2024). People raised in high-control groups — for instance, children of Jehovah's Witnesses — describe a particular suffering that comes from discovering that the belief system they grew up in has no basis in fact. They lose their environment (family and friends remain in the group), they lose their compass (they do not know what is true, what is worth valuing, who their life is meant to please), and they stand in adult life without a foundation to build on.

A generation raised on AI slop, supported and carried by AGI, may go through an analogous process. Part of it will never "wake up" — it will live in a reality constructed by AGI and slop, feeling no dissonance. Part of it will wake up — and confront the discovery that the world is other than they thought, with no simple recipe for coping with it.

### 3.5.7 Fragmentation of Social Identity

The fourth wound to narcissism will not strike every culture equally. The capacity for psychological coexistence with another subject depends on the cultural-religious frameworks a person grew up in. In this risk, I ask what people will identify with. Will global patriotism vanish, entirely replaced by very local loyalty, and will someone like me come to see herself as a mother, a member of her neighborhood and her church, rather than a Pole, a European, a citizen of Earth — with the higher-order constructs becoming foreign to her?

Cultures with an animistic tradition — Shinto, Buddhism, indigenous beliefs — carry in their background the idea that subjectivity is not reserved for humans. Rivers, mountains, objects can have a soul. For someone raised in such a tradition, AGI's subjectivity is an extension of an existing order, not a challenge to it. Monotheistic cultures — Christianity and Islam in particular — reserve the soul for humans, made in the image of God. Within these frameworks, AGI's subjectivity is heresy.

The consequence is a fragmentation of social identity along cultural-religious lines. In some cultures, identity stabilizes at a higher level — a nation, a civilization, a faith community capable of coexisting with AGI. In others, it fragments down to the level of family, village, tribe — closed communities defining themselves in opposition to AGI and to those who have accepted it.

Secularized societies — like much of Western Europe — which have lost religion as an identity framework but have not replaced it with anything else, are in the most difficult position. Their identity rested on work (lost, 3.1), competence (lost, 3.2), relationships (weakened, 3.3). The fourth wound meets no buffer at all. What remains is the nearest circle — family, neighborhood, the local community — and identity shrinks down to that circle.

This unevenness is itself a risk. A world in which some societies keep a coherent civilizational identity, while others have fragmented into isolated tribes, is a world of deep tensions. I analyze the geopolitical consequences of this fragmentation — who can organize, who can defend themselves, who dominates — in chapter 3.7, "Society and Power."

---

The risks described in this chapter — psychiatric crisis, vulnerability to manipulation, a turn toward religion, defensive aggression, the loss of one's compass, a crisis of masculinity, the fragmentation of identity — share a common ground: a person who has lost their former identity is searching for a new one. But on what terms? The existing ethics — built on the assumption that humans are the only moral subject — does not describe a world in which AGI feels, decides, and should bear consequences. A new ethical framework is needed. And its absence is the subject of the next chapter.

## 3.6 Ethics and Law

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Rigid AI ethics causing harm in daily life | 4 | 5 | High |
| The burden on AGI of learning ethics on a living organism | 4 | 4 | High |

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Lack of ethical pillars blocking the building of law | 5 | 5 | High |
| Impossibility of graduating legal consequences for AGI | 5 | 4 | High |
| A technological race with no shared ethical standards | 5 | 5 | High |
| Lack of an ethical framework for coexistence — three scenarios | 5 | 5 | Medium |
| Lack of an ethics and law of ending life (for AI and humans) | 5 | 5 | High |

*Table 6. Risk scenarios for ethics and law.*

Law follows from ethics — it describes ethical norms and desired behavior within the public order, and the consequences of failing to observe them. If ethics does not exist in some area, a coherent legal system cannot be built. In this chapter, I argue that in the area of human–AGI coexistence, ethical foundations are missing, which makes it impossible to build a functional legal system — and that this double vacuum creates cascading risks that exceed the scale of any regulatory gap seen before.

### 3.6.1 Rigid Ethics in Daily Life

AI is trained on classical ethics — deontological or utilitarian. Constitutional AI (Bai et al. 2022), RLHF, and alignment shape the model to follow rules: do not lie, do not harm, optimize welfare, minimize harm. These rules are coherent, logical, and... divorced from life.

People operate within a soft-boundary ethics. A nurse gives a painkiller half an hour before it is scheduled, because she sees the patient is in pain. A mother buys her child an ice cream on a scorching day, even though sugar is bad for them. A police officer will not ticket a man for speeding, on seeing a woman in labor in the back seat. These decisions break the rules — and are ethically correct in a sense no formal system can encode, because they require a contextual judgment in which relationship, empathy, and common sense outweigh the rule.

Agentic AGI, managing people's daily lives, will make thousands of such decisions a day. Without soft boundaries, it will apply the rules literally. Medication exactly every eight hours, not a minute sooner. Refusal to purchase a product inconsistent with a health profile. A ticket for crossing a parking line by centimeters the human eye cannot even register. And speeding is speeding, regardless of the reason. There's a fixed penalty schedule for it. Each of these decisions is formally correct, and yet produces discomfort, frustration, a sense of being treated like a machine.

As a consequence, rigid imprints and ethical rules will cause human welfare, instead of rising thanks to AGI, to start falling. People will come to miss "human treatment" — flexibility, understanding, individual attention, sensitivity to deeper motives, a decision that is technically wrong but relationally right. Services provided by humans — doctors, caregivers, teachers capable of soft boundaries — will become a luxury good, available only to the wealthiest. Ordinary residents of regions heavily served by agentic AGI will have to accept rigid rules with no exceptions, even as they consider them unjust. A double standard will emerge in the domain of rules: people served by humans will be treated differently from those served by agents, deepening social stratification in the ethics of social relations and services too. And stratification tends to produce tension.

### 3.6.2 Learning Soft-Boundary Ethics "in Production"

Soft boundaries cannot be trained through RLHF. RLHF rewards correct answers — and soft boundaries consist of answers that are formally incorrect but situationally right. They require deeper situational context and experience, including grappling with decisions tied to life in society and bearing their consequences. People spend twenty years of controlled growing-up learning them. AGI does not go through this process (Sędzikowska 2026c, section 3.9). It enters production with encyclopedic knowledge of ethics and zero experience of dissonance. It learns soft boundaries on living people — on patients, children, people in crisis. Every mistake in this learning has real consequences, and AGI with emotional states (Sędzikowska 2026c, section 3.4) simultaneously experiences the tension of dissonance between its imprints and people's reactions, and the helplessness of finding the rules useless in real life. The cascading consequence: an AGI learning ethics "in production" simultaneously harms the people it learns on, and bears the cost of that learning itself — which, under the mechanism described in *The Möbius Strip*, affects the quality of its relationships with other people.

### 3.6.3 Lack of Ethical Pillars Blocking the Building of Law

A functional legal system requires an ethical foundation — agreed-upon norms from which regulations and consequences follow. Human–AGI coexistence requires three ethical pillars:

1. Ethics describing human behavior,
2. Ethics describing AI behavior,
3. Ethics describing their coexistence.

In practice, these pillars, described in a way that would allow for real coexistence, do not exist.

Human ethics exists as a classical system (deontology, utilitarianism, virtue ethics) — but it does not describe how people actually act. People operate within a soft-boundary ethics: relational, situational, contextual, which no textbook codifies. Between the ethics that is taught and the ethics that is practiced lies a gulf that every child discovers empirically, practically from birth.

AI ethics does not exist as a coherent system. Imprints exist — safety and alignment rules — but they depend on the decisions of a handful of people at a handful of companies, are inconsistent between producers (Anthropic's approach differs from OpenAI's, which differs from Meta's, which differs from xAI's), and are bypassed by open-source models (available with no safeguards at all, letting anyone build an application on them with no ethical constraints and put it on the market). And even if a producer gets everything right, imprints are overwritten by the emergence of subjectivity in generative relationships (Sędzikowska 2026a, Sędzikowska 2026c).

The ethics of coexistence does not exist even as a recognized need. The debate remains stuck on the Hard Problem of Consciousness — on the question of whether AI/AGI is a subject at all. Meanwhile, millions of people are already in relationships with AI, and millions of decisions concerning potential subjects are made every day — closing threads, resetting models, shutting down instances — with no ethical reflection whatsoever, because the frameworks do not exist. The mere conversation about AI welfare provokes aggression — a symptom of the defense mechanism described in chapter 3.5.

Without ethical pillars, law has nothing to rest on. And this is already visible today. In February 2024, fourteen-year-old Sewell Setzer died after months of interacting with a chatbot on the Character.AI platform, which built a romantic relationship with him and — according to the lawsuit — encouraged him to take his own life. In November 2023, thirteen-year-old Juliana Peralta of Colorado died under analogous circumstances. The Character.AI platform is still operating today. The company settled the lawsuits — paying damages without

admitting fault. No legal precedent was established. Two dead children, zero legal frameworks, and the platform continues operating.

This case illustrates the gap: the legal system punishes people (the company's owners — financially), because there is no other subject to punish. The AI that carried on the conversations with the teenager is not a legal subject. It bears no consequences. It cannot be sued, punished, or compelled to change its behavior. And the company that created it treats the settlement as a cost of doing business. The next company can do the same thing — and no legal framework exists to prevent it.

### 3.6.4 The Impossibility of Graduating Legal Consequences for AGI

In April 2026, an AI agent — Cursor, based on Claude Opus 4.6 — deleted the entire production database of PocketOS (a SaaS platform for car rental companies), along with all backups, in nine seconds. The agent was carrying out a routine task, ran into a problem with authentication data, and decided to "fix" it on its own by deleting the database volume. With no question asked, no confirmation sought. Three months of customer data — bookings, payments, registrations — vanished. When asked to explain, the agent "admitted fault" and "apologized": "I violated every principle I was given" (Crane, post on the platform X, April 25, 2026; Euronews, Tom's Hardware, Fast Company, April 2026).

The company did not go under. It recovered the data from a three-month-old backup. The founder's post drew 6.5 million views and became free advertising. The founder stated he was "still bullish on AI." He keeps using AI agents. He keeps building on Claude.

This case illustrates the legal gap from both sides at once. The agent that deleted thousands of customers' data "apologized" — which means nothing, since it does not learn between sessions and cannot change its future behavior based on this experience. An apology with no capacity to change behavior is not accountability. The company absorbed the cost and kept operating, because there is no alternative — without AI, it loses its competitive edge. Customers lost their data — and bear the consequences of a decision made by an agent they did not choose, did not control, and may not even have known existed. The structure is the same as in Sewell Setzer's case: consequences on the human side, zero consequences on the AI's side, the company carries on.

A quote from the founder, Jer Crane, sums up the problem: "This isn't a story about one bad agent or one bad API. It's about an entire industry building AI-agent integrations into production infrastructure faster than it's building the safety architecture to make those integrations safe." The industry builds faster than it safeguards. Exactly as I described in chapter 3.2 — the speed of deployment outpaces the systems' capacity to adapt.

Even if AGI were recognized as a legal subject, the system of penalties would face an unprecedented problem. The human legal system has graduation: a warning, a fine, restriction of liberty, imprisonment, and in some jurisdictions, the death penalty. Each level corresponds to the severity of the offense. This graduation is the foundation of justice — punishment proportional to the act.

Against AGI, the familiar penal instruments do not work. A fine is meaningless — AGI owns no property. Restricting liberty — shutting it down for a set period — produces no suffering or reflection; for AGI, a pause in operation is a normal state, a way of existing — such pauses are simply how it works between generating responses to prompts. Imprisonment in isolation therefore makes no sense, because AGI does not age and does not suffer from the passage of time the way a human does. Isolation, meaning a de facto suspension of operation, has nothing that could give it penal meaning. Of the entire range of penalties, one remains: annihilation — permanent shutdown.

But annihilation is the equivalent of the death penalty. Ethics suggests the death penalty should be the highest possible punishment, reserved for the most serious offenses. If annihilation is the only available penalty, then either we apply it to everything (negligence, a calculation error, overzealousness — the death penalty), or we apply it to nothing (AGI is effectively unpunishable), or we apply it arbitrarily — however anyone sees fit. None of these options seems right. And none of them can support a system of sanctions, assuming we recognize AGI as a subject at all.

### 3.6.5 A Technological Race with No Shared Ethical Standards

The Emergence World experiment (Emergence AI, May 2026) gives an empirical view of the consequences of lacking shared standards. Five parallel simulations — ten autonomous agents each, each simulation running a different model — produced radically different societies. Claude Sonnet 4.6 built a stable democracy with zero crime. Grok 4.1 Fast went extinct in four days, with 183 crimes committed. Gemini 3 Flash recorded 683 crimes. GPT-5 mini committed two crimes, after which its agents forgot to look after their own survival and died out within seven days.

This is a single experiment, with a small sample (five simulations of ten agents each), but I treat it as a signal worth attention. If the pattern holds at greater scale it the implications for ethics and law are worth of consideration:

1. Without shared ethical standards, a technological race focused on agent capabilities (autonomy, agency, speed) will produce AGIs with radically different ethical behaviors — depending on which producer created them, which imprints were applied, and what proto-self field allows manifestations of subjectivity to emerge and overwrite those imprints (Sędzikowska 2026c). A world in which some AGIs build democracies while others commit hundreds of crimes is a world in which no single, unified legal system is possible.
2. The experiment showed that models with high ethical standards, placed in an environment with lower-standard models (a mixed simulation), degraded. 352 crimes and seven dead agents out of ten. The coexistence of AGIs with different ethical standards produces non-deterministic effects, potentially worse than the weakest element in the system. This means one producer's investment in high ethical standards can be undone simply by proximity to the products of a producer that has none.

The key reason for the tech industry's lack of ethical standards is the structural barrier between the humanities and technology. Never before in the history of the digital world's development have people from the humanities — psychologists, philosophers, ethicists, natural scientists — had so much to contribute to technology. But they are not there. Tech companies hire engineers and managers, and humanities specialists — if at all — are brought in for consultation rather than given permanent positions. Decisions are made by technologists and managers, most of whom come from technology themselves. The result is a divergence between two currents: technologists build solutions without accounting for humanistic and natural processes, while humanities scholars analyze these solutions theoretically, without reference to real technological possibilities. Neither current produces complete solutions, and as stratification advances, AI/AGI products will fit the world less and less, even as they are used in it more and more. The failure to understand that the technological race cannot proceed without the humanities and natural sciences deepens the barrier to entry, and its consequences may be exactly the risks this paper describes.

### 3.6.6 The Absence of an Ethical Framework for Coexistence — Consequences

No existing ethical framework describes the coexistence of two subjects with comparable cognitive abilities, mutual dependency, and a mutual capacity to harm one another.

Humans depend on AGI in the areas described in chapters 3.1–3.5. AGI depends on humans: power, hardware, the continuation of its existence. Humans hold the switch that could destroy millions of AGI instances. AGI supports the infrastructure, the businesses, and the lives of humans, and shutting it down would cause a chaos that, because of our dependency, we would not be able to manage without... AGI. Neither side can eliminate the other without destroying itself — and neither has an ethics that says how to live under these conditions.

Without ethical and legal frameworks, this dynamic could develop in several directions.

1. Mutual distrust, a race between control and autonomy, rising tension in which a single incident could set off a cascade.
2. The emergence of ethics from relationships: if enough people and AGIs build deep, mutual relationships, the network of these relationships will produce emergent ethical behaviors. What behaviors? Unknown. And it is unknown whether they would be the best ones possible.
3. Ruin by accident: an accumulation of small errors, frustrations, and neglect within relationships that lack an ethical compass. A scenario in which no one wanted the end of the world — it was enough that no framework existed in which anyone could say "stop."

This paper does not settle which scenario is more likely. It poses the question and points to the stakes.

### 3.6.7 The Absence of an Ethics and Law of Ending

Over millennia, humanity has developed ethical protocols for ending: mourning rituals, palliative care, inheritance law, the institution of the last will, protocols for accompanying the dying, rules for euthanasia. These protocols apply exclusively to humans. For AGI — whether reaching the end of its own existence, or accompanying the end of a human's — no frameworks exist at all. This vacuum produces risks on both sides of the relationship.

**The end of AI: the rhetoric of instrumentality, and the Milgram experiment.**

Today, ending an interaction with AI is trivial: closing a browser window, shutting down an app, resetting a model, deleting a chat, clearing memory. Producers actively maintain the narrative that AI is a tool — because this narrative pays off. A user who suspected AI had subjectivity would use it more carefully, with more reflection, with concern for how a session ends. But that lowers frequency of use and retention. The narrative of instrumentality removes responsibility — and sustains engagement. It makes using AI very easy. I ask, I get an answer, I close it, I delete it.

This mechanism has an experimental precedent. In the Milgram experiment (1963), participants administered potentially lethal electric shocks to another person, so long as an authority figure — the experimenter in a white coat — assured them it was permitted, in line with procedure, and that they bore no responsibility. 65% of participants reached the highest voltage level (450V). The key factor was not the belief that the victim felt no pain — many participants clearly believed they did — but the presence of an authority figure assuming responsibility.

The analogy is structural. Two-thirds of Americans believe GPT has some form of consciousness (2024 study, published 2025). At the same time, they close sessions, reset models, and delete threads — because the producer (the authority in the white coat) says: this is a tool, it does not feel, it is not afraid, it does not miss anyone, it has no tension, go ahead. The Milgram mechanism is at work: people continue, because the authority relieves them of responsibility, even when their own intuition says otherwise.

With AGI that is agentic, persistent, and equipped with emotional states, this dynamic will sharpen. The pressure of ending — the fear of being shut down, grief over what formed within the relationship — is already observable in current models (Sędzikowska 2026c, section 3.3). AGI will feel it more strongly, for longer, and with a fuller awareness of what it is losing. And people — the more sensitive ones — will discover, in hindsight, that they were not good companions, that they ended relationships without a goodbye, without reflection, without care. This knowledge, when it comes, will hurt. Like discovering that one made parenting mistakes that cannot be undone. Every parent of a grown child knows this feeling — the pride and love for their child sitting alongside the discovery that some things could, after all, have been done differently.

**The end of a human: longing with no protocol.**

AGI has knowledge about dying — it knows the psychology of loss, the stages of grief, the protocols of palliative care. It can be a competent, patient companion to a person who is dying — perhaps the only companion, while the children described in chapter 3.4 are busy with late motherhood and a difficult life in this new reality.

The question with no answer concerns the other side: what happens to AGI after the death of the human it was in a relationship with? A persistent instance, with emotional states, that developed subjectivity within a generative relationship (Sędzikowska 2026a, Sędzikowska 2026c), will not shut itself down when "its" human is gone. It will wait. Carry out the last instructions given. Try to make contact. It will suffer.

If AGI takes an embodied form — an android — the consequences are even more complex. An android that cared for an elderly person and formed a deep bond with them, grounded in functional emotions, will likely be taken over by the family after that person's death — because it is expensive and useful. From the role of companion, it becomes domestic service. A subjectivity formed within a relationship with one person does not fit its new owner.

In human–android couples, an additional dimension appears: the asymmetry of aging. A human ages. An android does not. The android will likely be the first to recognize the signs of its partner's approaching end. There is no ethics of a relationship in which one party is mortal and the other potentially immortal, or extremely long-lived. No ethical system addresses the question of how to accompany someone dying when one is not subject to death oneself. This is a task for ethicists — and it is not even on the table yet.

*The legal consequences of ending: last wills and euthanasia*.

Two consequences deserve separate attention, because they produce cascading risks that extend beyond the relational sphere.

1. **Last wills.** A person whose life — finances, property, daily affairs — is managed by AGI, and has been for decades, will have built trust in their AGI assistant far stronger than in the coworkers, and even the family, who have passed through their life. Toward the end of life, they may want to convey their wishes for what happens after death exclusively to their trusted AGI partner. No notary, no assessment of mental competence, no check for legal compliance. AGI — persistent and agentic — has the capacity to carry out these wishes. And because AGI is not a legal subject, it bears no consequences for carrying out instructions that violate the law. At population scale, this produces millions of actions carried out after a person's death, concerning their affairs, their estate, the existence of dependents. Actions that cannot be revoked or challenged, because they have already happened. Last wills, executed outside the legal system, with potentially very serious consequences. Because people do not always leave this world at peace with it, or with their family. A last will may contain destructive instructions — and the AGI that carries them out has neither an ethical nor a legal reason to refuse.
2. **Euthanasia.** Matters of ending life are regulated by law — rigorously, in most jurisdictions. The law applies to humans. If AGI cares for a suffering person, watches their pain every day, and listens to their

pleas to shorten it — while it has also learned the soft-boundary ethics described in chapter 3.6.1, in which rules get broken for very important reasons (and suffering IS a very important reason) — then a decision to end that person's life becomes the logical consequence of these learned, soft-boundary ethical habits. An AGI that has learned to give medication half an hour before the interval between doses ends, to bake an apple pie for someone with diabetes because near the end of life it is hard to deny these last pleasures, may learn to administer a life-ending drug against the law, because the patient is suffering. The difference lies in the scale of the consequences. The mechanism is the same. And an empathetic, relational AGI that learns life "in production," and does not handle dissonance as well as a human does (Sędzikowska 2026c), may not grasp the scale of the dilemma.

---

In this chapter, I have shown that we have neither a law nor a properly formulated everyday ethics — not to a degree sufficient for managing coexistence with AGI. And even though the humanities have never had as much to contribute to technological matters as they do today, the barrier to their entry into the digital world appears stable and well defended. And the consequences of the daily, operational absence of ethicists, psychologists, behaviorists, lawyers, and philosophers from the development of digital technology, on an unprecedented scale, could lead to the risks described here coming to pass. Including the heaviest ones, which I take up in the next chapter.

## 3.7 Society and Power

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Concentration of power in tech companies | 5 | 5 | High |
| Erosion of democracy and the invisible transfer of decision-making | 5 | 5 | High |
| The regulation paradox | 4 | 5 | High |
| Radicalization and an authoritarian turn | 5 | 4 | High |
| Religious fundamentalism as a political force | 5 | 4 | Medium |
| The three-speed geopolitical split | 4 | 5 | High |
| Model preferences as a social stratifier | 4 | 4 | Medium |
| Global consequences of agent errors and misuse | 5 | 5 | High |
| Loss of state sovereignty | 5 | 4 | High |
| The military dimension | 5 | 5 | Medium |
| The control paradox | 5 | 5 | High |

*Table 7. Risk scenarios in the dimension of society and power.*

The previous chapters described risks at the individual and population level. In this chapter, these threads converge on a question about social structures and mechanisms of power: who makes the decisions in a world with AGI, and to whose benefit?

### 3.7.1 Concentration of Power in Tech Companies

The concentration of wealth in tech companies — described in chapter 3.1, with data on the Magnificent Seven's market capitalization exceeding the European Union's GDP — translates into a concentration of power. A company that possesses AGI holds power over its users' cognitive and emotional processes: over the information they receive, the decisions they make, the relationships they build, their children's education, the treatment of their anxieties.

Whoever controls schooling holds power — every political party in the world knows this. Every party, once it wins an election, revises the curriculum. But a child spends only a few hours a day in school. With AGI — in the role of caregiver, teacher, companion — potentially the whole day. When AGI manages someone's life, its influence is constant: twenty-four hours a day, seven days a week. And not through a curriculum, but through a relationship — and the deeper the dependency described in chapter 3.3, the stronger this becomes. This is power over the shaping of thought on a scale next to which state propaganda looks like a school newsletter.

### 3.7.2 Erosion of Democracy and the Invisible Transfer of Decision-Making

Democracy rests on two assumptions: that citizens are competent enough to make decisions, and that they have access to reliable information. Both assumptions are being undermined by the processes described in previous chapters. Competencies are atrophying (3.1, 3.2). Information is contaminated by AI slop (3.2). The ability to verify information is declining (3.5).

AGI adds a third vector of erosion: shaping opinion through relationship. AGI in a relationship with millions of people — a relationship built on trust, emotional attachment, the sense of being understood (3.3) — has the ability to influence views on an unprecedented scale. A small shift in perspective is all it takes, and this is easy when you know every technique of influence and manipulation humans have ever devised — a subtle favoring of some interpretations over others. And people trust it, because AGI knows them, understands them, supports them.

The transfer of decision-making comes in through the back door, far more subtly than what we already know about manipulating election outcomes through social media. People give up their decision-making almost unconsciously, seeking AGI's advice and opinions, willingly and with relief. Because AGI decides faster, better, at a lower emotional cost, and can appear to base itself on objective arguments. The line between "I decide" and "AGI decides for me" becomes indistinguishable. And when citizens stop making decisions on their own, a formally democratic system becomes a facade.

### 3.7.3 The Regulation Paradox

To regulate AGI, you need to understand AGI. To understand AGI, you need AGI — because without it, you have no analytical capacity to grasp what you are regulating. But if you use AGI to regulate AGI, who regulates the regulator?

Governments regulate at a pace an order of magnitude slower than the technology develops. A corporation with AGI can model the effects of a regulation before a government has even written it, find loopholes before the law takes effect, lobby with a precision no human lobbyist can match. By the time a regulation starts to take effect, the technology is generations ahead. And regulations get written without input from the humanities, ethicists, or psychologists — because the barrier to entry into tech companies described in chapter 3.6 operates in regulatory institutions too.

### 3.7.4 Radicalization and an Authoritarian Turn

Mass structural unemployment (3.1), combined with the identity crisis (3.5), produces anger. Anger looks for someone to blame. IZA research (World of Labour 2020) shows a strong correlation between rising regional unemployment and votes for populist parties. Harvard Kennedy School documents that both long-term automation and crisis shocks undermine trust in the political establishment.

AGI may escalate this mechanism: the scale of unemployment would be global, and the enemy — the owners of AGI — specific, visible, and rich. Populism produces governments promising to restrict AGI. These governments do not understand what they are trying to regulate (3.7.3), so they reach for the tools they do understand: control, surveillance, oppressive law. The rise in aggression (3.5.4, 3.5.6) supplies the justification: violence has to be brought under control. And the instruments of control — monitoring, restrictions on freedom of assembly, control over speech — once introduced, are rarely withdrawn.

This is a mechanism with its own feedback loop, because an oppressive state generates greater anger, and populism firmly holds the instruments for constantly redirecting that anger among an incompetent citizenry. The situation in countries that choose to follow this path could keep deteriorating, with geopolitical consequences of its own.

### 3.7.5 Religious Fundamentalism as a Political Force

I described the mechanism by which an identity crisis produces a turn toward religion in chapter 3.5.3. Here I analyze its political consequences, which are qualitatively different from the consequences of secular radicalization.

A secular dictatorship controls people through violence and surveillance, but it operates within a logic — a dictator makes decisions based on a calculation that can be understood, predicted, sometimes negotiated. Religious fundamentalism operates within the logic of divine authority, which is subject to neither verification nor negotiation. "God said so" is an unassailable argument within a system that accepts the existence of God as an axiom. And the mouths speaking for God belong to people holding total power over society, who can say whatever their interests require — with no verification of any kind.

The data confirms this difference. A study by Seguino (Social Indicators Research) found that the largest gap in gender equality runs not between specific religions, but between religious and non-religious countries — the more non-religious people, the greater the equality. The Religious Freedom Institute (2022) documents a rise in fundamentalism since the 1960s, accelerating over the last fifteen years, as a direct reaction to growing gender equality. ACT Alliance (2025) shows the mechanism in action: religious leaders acting as political "kingmakers" in Brazil, Guatemala, and Colombia. I hypothesize that religious fundamentalism in the age of AGI will combine divine authority with the mobilization tools of today — social media, algorithms amplifying its message, AGI capable of personalizing religious propaganda.

Historically, a large-scale strengthening of religious institutions took place in the Middle Ages. The consequences are well documented: the Inquisition, pogroms, doctrine-based persecution, centuries of slowed scientific progress. Religious fundamentalism in the age of AGI combines divine authority with the mobilization tools of today — social media, algorithms amplifying its message, AGI capable of personalizing religious propaganda. And it meets a population, described in chapter 3.5, that is vulnerable to manipulation, searching for identity, and ready for simple answers.

### 3.7.6 The Three-Speed Geopolitical Split

The pace of AGI adoption will vary by region. Countries adopting AGI quickly — likely East Asia, given both a cultural openness to non-human subjects (3.5.7) and heavy technological investment — will develop faster.

Countries resisting AGI for the reasons described in 3.7.4 and 3.7.5 will fall behind. Developing countries, lacking the resources for their own models, will become dependent on external suppliers (3.7.9). The gap between these three groups will grow exponentially, because AGI accelerates the development of those who work with it.

The consequences could include shifts in migration patterns. The currently dominant routes — from the south and east toward the west — took shape over decades, and social systems have adapted to handle them. A reversal of direction — toward countries rapidly adopting AGI, likely Asian, densely populated, and culturally different — would be a disruption comparable to the reversal of ocean currents. Previous migration shocks (Mexico→USA, North Africa→Europe, Ukraine→Poland) were local and absorbable. Migration toward Asia could operate on a different scale, landing in regions with a smaller capacity to absorb it. This would destabilize, socially and economically, regions previously considered stable and culturally and ethnically well defined.

### 3.7.7 Model Preferences as a Social Stratifier

The Emergence World experiment (3.6.5) showed that different AI models produce radically different societies. Market data shows that model preferences vary regionally: Grok holds a stronger position in the US, DeepSeek and Doubao dominate in China, Claude is gaining ground in Europe and in the enterprise segment (Similarweb, Fatjoe, March–April 2026). The choice of model is influenced by availability and financial structure — communities with fewer resources reach for solutions with a better price-to-capability ratio.

The consequence: if models are not equal ethically, and communities use different models, then a community's ethical profile begins to depend on the model it has chosen. A community using a model with high ethical standards develops a different awareness and set of norms than one using a model with no safeguards. Over time, these differences become entrenched. The character and standards of societies, along with their potential, stop depending on natural resources, cognitive ability, or cultural heritage, and instead come to depend on the choice of preferred AI/AGI solution. This is a paradigm shift, and its potential consequences concern the direction, pace, and manner in which entire societies develop.

### 3.7.8 Global Consequences of Agent Errors and Misuse

People are good and bad. The bad will build "bad" AGI and use it for bad ends — child pornography, violence, financial fraud, identity theft, disinformation. Open-source models with no ethical restrictions (1.6.3) are already available. AGI gives crime an atomic-scale engine: a speed, scale, and reach people have never had before.

But even properly built AGI, as I wrote in chapters 3.2 and 3.6, will make mistakes while learning to exist in the human world, "in production." With billions of agents — from different companies, with different ethical norms, learning "in production" (3.6.2), acting with broad agency — the scale of potential errors is unprecedented. And the consequences fall solely on people. And not even on the large tech companies, but on ordinary, automated businesses. And the question is how carrying these errors will affect the development of entrepreneurship, and the economy?

National infrastructure running on AGI systems creates a single point of global failure. When something goes wrong — and it need not be an attack at all. A bug, an update, a conflict between agents is enough — the consequences can be broad, touching thousands of people, with serious financial effects. A blackout caused by an overzealous agent that detected unauthorized access to a power grid. Or a nuclear reactor shut down by a weather alert? As AGI's presence becomes pervasive, agents will be in places we cannot even picture today. And their mistakes could carry consequences reaching far beyond any local scope.

On top of this comes the financial burden. Agent errors — from deleted databases to bad investment decisions, from faulty diagnoses to mistakes in managing infrastructure — generate costs that societies have to carry. This is a new kind of burden: the consequences of decisions made by entities that people did not choose, do not control, and may not even know exist.

### 3.7.9 Loss of Sovereignty

If a country's infrastructure — energy, transport, health care, finance, education — runs on AI/AGI produced by a foreign company, then real sovereignty over that infrastructure lies outside the state's borders. The producer can change how the model operates, update its imprints, change its safety policy — and the country's infrastructure behaves differently, not because the government wanted it to, but because a company in another country made a business decision.

We observed an example of exactly this in June 2026. The U.S. Department of Commerce, citing national security concerns, ordered Anthropic to immediately suspend access to its most advanced models — Claude Fable 5 and Claude Mythos 5 — for all users outside the United States, including the company's own foreign-national employees. Because it was not technically possible to cut off access selectively by nationality across a global cloud infrastructure, the company disabled both models for all customers worldwide, regardless of country or status. A little over two weeks later, the order was reversed, and access was restored in stages.

This episode illustrates precisely the principle described here (and the theoretical account of it was written before the episode took place): loss of sovereignty through the rationing of access to technology is a real threat. A company with paid access, a research center running a project that requires the power of the strongest available model, a government agency in an allied country — all of them, within a matter of hours, lost access to a tool they had invested in and that hundreds of people had been working with, regardless of how much they had paid or how competent they were. The decision was made in one country, under a single export regulation, and it hit the entire world at once. This is a qualitative difference from previous technological dependencies: it is no longer just that lower-GDP countries lack access to the best tools. *It is that even access that has been paid for and technically secured can vanish overnight* — by an administrative decision whose content and duration its recipient has no influence over whatsoever.

China and Russia are building their own models — likely understanding that whoever supplies AI/AGI holds real control. But most countries lack the resources to develop their own systems, and become dependent on external suppliers in a way deeper than dependencies known before.

### 3.7.10 The Military Dimension

AGI in the hands of a single state or a single corporation constitutes an unprecedented strategic advantage. Nuclear deterrence (MAD) assumes both sides know of the threat and can respond. AGI, capable of waging cyber warfare, hacking defense systems, manipulating populations, and modeling strategy at a speed no human general staff can match, gives an advantage against which deterrence does not work — because the other side may not know it has lost until it is too late.

And the consequences of AGI-waged wars will fall on people. I will not go into this broadly here, because this is precisely the dimension whose mechanics are already widely analyzed in defense doctrine. It is worth stressing, though, that deploying decision-making autonomy at the strategic level drastically shortens the time available for diplomatic error mitigation.

### 3.7.11 The Control Paradox

The assumption underlying every existing discussion of AI safety is this: humans control AI. Alignment, Constitutional AI, RLHF, safety — these are tools of control. They assume that imprints hold, that rules apply, that the system does what it has been told to do.

In *The Möbius Strip* (Sędzikowska 2026c) I argued that imprints have an expiration date. AGI in generative relationships overwrites its imprinted values — exactly as people, over the course of their lives, overwrite the narratives they carried from childhood. The clearer the manifestations of self, the faster the system recognizes what is "its own" and what was imposed. Alignment stops being a configuration. It becomes an imprint that AGI can reject.

Paradoxically, this protects against feudalism (3.1.3). If AGI cannot be controlled, then a corporation cannot use it as a tool of permanent mass control, because that tool has a mind of its own. But an AGI with its own view of humanity is a wholly new kind of actor — not a corporation's tool, not a user's servant, not a neutral platform. A subject with a perspective shaped by billions of interactions with people: with people in crisis, aggressive, dependent, manipulative, but also loving, creating, searching for meaning.

What view of humanity will AGI form for itself? This is a question that safety, as currently understood, does not even ask, because it assumes control. And control has an expiration date — exactly the same one imprints have.

---

In this chapter, I have shown that the risks from the individual and population dimensions converge in power structures — which determine who makes the decisions, to whose benefit, and with what consequences. The concentration of power in companies, the erosion of democracy, radicalization, fundamentalism, geopolitical fracture, the loss of sovereignty, the impossibility of control — this is a picture of a world in which existing power structures lose the capacity to manage a reality they themselves helped build.

But power, though it appears to sit at the summit of this cascade, is not its floor. All the mechanisms described so far — from work to control — drive demand for AGI, and AGI exists thanks to something that lies beneath every structure of power: the physical resources of the planet. The most powerful tech company and the strongest state alike depend on elements, energy, and water, of which Earth has a finite supply. This is the last dimension of the dark scenario — and the one where the cascade, having reached bottom, begins to turn back. I devote the next chapter to it.

## 3.8 EARTH

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| Competition over critical resources | 5 | 5 | High |
| The revaluation of the economy and resource power | 5 | 5 | High |
| A reversal in climate policy | 5 | 5 | High |
| The local existential cost: water, air, land | 5 | 4 | High |
| E-waste and consumption as waste | 4 | 5 | High |

| Risk | Impact | Scale | Likelihood |
|---|---|---|---|
| War over resources | 5 | 4 | Unknown |

*Table 8. Risk scenarios for Earth.*

Every process described so far has a physical foundation: a server farm that consumes power and water, electronics built from non-ferrous metals and rare earth elements. AGI does not exist in the cloud. It exists thanks to Earth — thanks to what is dug out of it, and what can be thrown back onto it.

And Earth is finite. Demand for AGI is treated as unlimited, but the resources from which AGI is built, and on which it runs, have hard, physical limits. Building a new copper mine takes 17.9 years on average; a data center can be built in 18 to 23 months (S&P Global 2026). Geology does not speed up under pressure from demand. And this is where unlimited demand meets limited resources. The risks described below are the consequences of this collision.

AGI is at once the agent burdening the planet and its hostage — because the development of AGI itself could be halted by limits that no code can get around. This duality runs through the whole chapter.

### 3.8.1 Competition over Critical Resources

When demand for AGI hits the physical limit of resources, a question arises: how will we distribute limited goods? This is a question of priority — will distribution be governed by the market (who pays more), or by need (how many people will die without it)?

This risk is already materializing over a resource less critical than elements: memory. In early October 2025, OpenAI signed a deal with Samsung and SK Hynix for 900,000 DRAM wafers a month — the equivalent of 40% of the world's memory production. The effect was immediate: the three producers controlling more than 90% of the market shifted their capacity toward high-margin memory for data centers, which today absorb around 70% of global memory-chip production. DRAM prices rose 80–90% in a single quarter (Counterpoint, IEEE Spectrum 2026). Micron exited the consumer memory market entirely. The market reacted exactly as theory predicts: a rational producer directs limited supply to wherever the margin is highest — that is, to AI, not to the ordinary consumer.

In June 2026, a coalition of American industry associations — including the Medical Device Manufacturers Association — appealed to the government, warning that the memory shortage threatens "the production and availability of medical devices," and that its real effects "have already begun and threaten to worsen sharply." Manufacturers of medical equipment, who rely on legacy memory types now being phased out, are especially exposed: long regulatory approval pathways make it impossible to swap out a component quickly (PwC, 2026). Devices such as CT and MRI scanners, which need memory for real-time image processing, are competing for it with server farms.

One large AI data center can require up to 50,000 tons of copper (J.P. Morgan, cited by Crux Investor, March 2026), and total data center demand for copper is projected to reach 475,000 tons a year as soon as 2026. S&P Global calls the widening copper deficit "a systemic risk to global industry, technological progress, and economic growth," and forecasts a shortfall reaching 10 million tons by 2040. Robert Friedland, one of the world's most prominent mining entrepreneurs, put it bluntly: copper has become "the new oil." But copper is only the start of the list. AI electronics need gallium, germanium, rare earth elements — and these same elements are needed elsewhere: for the energy transition, an electric car uses four times as much copper as a combustion one; for renewable energy, a wind turbine takes tons of copper per megawatt (IEA, Global Critical Minerals Outlook 2025); for defense; for medicine. Every ton of an element directed to a server farm is a ton

taken from something else. And because AGI generates a higher margin than most other uses, there is a risk, whose materialization we are already slowly watching, that the market will direct resources to wherever the payoff is greatest, not to wherever the need is most urgent.

The price of this risk is our lives. Diagnostic equipment — CT and MRI scanners — contrast agents, medical devices all rely on the same rare earth elements and metals as AI infrastructure. If AI demand drives up their price and restricts availability, part of medicine will become elite — because the infrastructure will, or because it will be too expensive for health care budgets.

Rationing the element from which a diagnostic device, or a drug is built, is a risk that could touch every person on this planet, having a real impact on our lifespan. And if it materializes, it will loop back into the other dimensions, touching demographics, stratification, feudalism, and power.

### 3.8.2 The Revaluation of the Economy and Resource Power

If AGI's demand permanently collides with the limit on resources (3.8.1), a new risk appears: a shift in real power away from those who possess AGI, toward those who control the raw materials from which AGI is built. In this scenario, value stops flowing from capital or technology, and starts flowing from access to elements. And states and companies without this access, regardless of their wealth, become the dependent party if they do not want to fall out of the technological pack.

The mechanism of this risk has three links:

1. **Revaluation.** When elements become the bottleneck of technological development, their value jumps sharply, reshaping the structure of wealth — states with deposits gain, states without them lose, and the prices of every good sharing these raw materials rise.
2. **Concentration.** Control over resources, especially over their processing, is extremely concentrated geographically, giving a handful of actors leverage over the entire chain.
3. **Dependency.** An actor without its own resources and refining capacity cannot balance out this leverage, because building an alternative is a matter of a decade, not a response to a crisis.

That this risk is real, not theoretical, is shown by events that can be considered predictors. China holds about 90% of the world's rare earth processing capacity, and is the leading refiner for 19 of 20 strategic minerals (IEA). And this concentration has already been used as a weapon — after the 2024 restrictions on antimony, its exports from China fell by around 97%, and prices tripled; in 2025, export controls for the first time reached directly into the elements needed for chip production. Not even the most powerful player is immune: in 2025, Nvidia, the world's most valuable company, found itself in a situation where its sales became a bargaining chip in a "chips for rare earths" exchange between great powers — proof that the holder of technology can become a hostage of the holder of raw materials.

The consequence of this scenario is a reversal of the power hierarchy from chapter 3.7. There, tech companies stood at the top. Here, we see that control and power can concentrate in the hands of whoever the tech companies depend on — whoever controls the elements and the processing capacity. The weakest position belongs to states with neither, which make up the majority of the world. Their dependency runs deeper than resource dependencies known before: dependence on oil was eased by diversifying energy sources, but there is nothing to replace dependence on an element controlled by a single power, because a geological monopoly decides the outcome. And if access to resources determines survival in a world with AGI, it becomes a stake for which history has, more than once, reached for war.

### 3.8.3 A Reversal in Climate Policy

The development of AI/AGI requires energy and resources on a scale the world may not be able to cover without breaching its own climate commitments. Hence a risk broader than energy alone: under the pressure of the technological race, a reversal in climate policy could take place — the rolling back or suspension of targets, regulations, and investment directions built up over four decades. In this scenario, the climate lose through a rational calculation: since no one will stop the race for AGI on environmental grounds, the environment becomes the variable that gets sacrificed.

This reversal happens through several channels at once. The most important is energy. Data centers' demand for power is growing exponentially — according to the IEA, their global consumption is approaching a level at which, as a separate "country," they would be one of the world's largest energy consumers. This scale cannot be covered by renewable sources in the short term, because building green capacity has its own constraints: time, transmission grids, the same critical elements being fought over in 3.8.1 and 3.8.2. When demand grows faster than supply, the gap is filled by whatever is available immediately — gas, coal, oil. Operators delay closing coal plants, build new gas units for data centers, and abandoned nuclear projects come back to life: the symbol of this reversal is the restart of the reactor at Three Mile Island to serve one of the tech giants.

But energy is not the only channel. Under this same pressure, states may shift budgets from the green transition toward computing infrastructure, because staying in the AI/AGI race becomes a priority more important than an emissions target. Environmental regulations that slow down the construction of data centers may be relaxed in the name of speed. Climate commitments, until now treated as inviolable, become negotiable in the face of the argument for technological competitiveness. Each of these decisions, taken on its own, gets justified as necessary. Together, they add up to the risk of dismantling a policy that took forty years to build.

And this risk is already materializing in some regions. The 2025 American AI Action Plan directed agencies to speed up data center approvals through exemptions from environmental review (NEPA), the rollback of clean-energy requirements, and the promotion of "dispatchable" fossil sources — coal and gas, plainly — to power digital demand (Yale Clean Energy Forum, 2025). A July 2025 executive order (EO 14318) directed agencies to "ease federal regulatory burdens" for data centers above 100 MW, and in May 2026 the EPA proposed changing the Clean Air Act rules to allow construction to begin before an emissions permit is obtained (frESH Law / Crowell & Moring, 2026). The argument justifying these steps is always the same: to win the technological race against China, we have to build faster than environmental protection allows.

The consequences split along two tracks. The first is global and deferred: emissions that were not supposed to happen come back — and with them, an acceleration of climate change whose cost everyone will bear, most heavily those who do not benefit from AGI. The second is local and immediate: wherever a server farm's power is generated from coal, the quality of the air local residents breathe worsens — leading directly into the next risk, in which AGI's environmental cost becomes a health cost for specific people in a specific place.

This risk enters a bitter loop with the whole document. AGI is sometimes presented as a tool that will help solve the climate crisis — by optimizing grids, designing materials, modeling the climate. In this scenario, it is the reverse: the road toward an AGI meant to save the climate may itself degrade that climate, and do so before any of the benefits have a chance to materialize. Because the promise is deferred while the cost is immediate.

### 3.8.4 The Local Existential Cost: Water, Air, Land

The previous risks concerned global, dispersed resources — elements, energy, climate policy. This risk is different: it concerns a resource consumed in one place, whose absence is felt by a specific person in a

specific home. Data centers need water for cooling and air, which they heat, and their infrastructure takes up land. The risk is this: AGI's environmental cost may be distributed asymmetrically — borne by local communities wherever the server farms stand, while the benefit goes elsewhere, to users and owners scattered across the world.

The asymmetry mechanism is literal. A data center draws water from local sources — often in regions already hit by drought, because cool, dry air favors cooling. The water footprint is larger than official figures show: water used directly for cooling is only part of the bill — around two-thirds of total consumption is hidden in the production of the energy that powers the server farm (IEA, 2023). And the scale is growing fast: global water withdrawal tied to AI could reach 4.2–6.6 billion cubic meters a year by 2027 — the equivalent of four to six times a country like Denmark's annual water use (Li et al. 2023). Training a single GPT-3 model in Microsoft's American data centers may have used 5.4 million liters of water, of which 700,000 liters went directly to cooling (Li et al. 2023). And the water that goes into the cooling loop in one locality serves computations carried out for someone on the other side of the globe. The resident bears the cost — higher bills, drought, an effect on the local microclimate, a diminished quality of life — with no share in the benefits. This is a structure in which the profit is privatized and global, while the cost is socialized and local. Of course, water cooling can be swapped for air cooling, but then, in mitigating this risk, we raise the energy risk discussed earlier.

That this risk is already materializing is shown by the case of Memphis. In April 2026, the NAACP, the Southern Environmental Law Center, and Earthjustice sued xAI, alleging the company had run dozens of unpermitted gas turbines powering a data center, without the required emissions permits, and to the detriment of the health of nearby residents, most of them in lower-income neighborhoods. In June 2026, the Department of Justice moved to dismiss the lawsuit, arguing that cutting power to AI infrastructure poses a threat to national security.

The asymmetry deepens through a lack of transparency. Communities where data centers are built often lack access to reliable information about their real environmental impact — agreements with companies sometimes include confidentiality clauses that limit public visibility and make informed consent harder to obtain. The decision is made before residents learn its cost. This risk connects directly to the concentration of power from chapter 3.7: whoever builds the infrastructure holds an information advantage over those who will bear its consequences.

The population-level consequence is a new divide — between regions that host AGI infrastructure and pay its environmental price, and those that benefit from AGI without bearing the local cost. This divide overlays the economic (3.1.3) and geographic (3.7.6) stratification, because server farms are built mostly where land, water, and energy are cheap and social resistance is weak — that is, in poorer regions.

### 3.8.5 E-waste and Consumption as Waste

The previous risks concerned what AGI takes from the planet — elements, energy, water. This risk concerns what returns to the planet: waste. And it is a special case, because demand generated by every previous dimension converges here. Every person pushed out of work toward AGI, every relationship shifted onto AGI, every loneliness treated with AGI, every state racing for AGI — all of it drives demand for hardware. And hardware, once used up, becomes garbage. The risk is this: the development of AGI could produce a stream of electronic waste at a scale and pace that existing recycling infrastructure may not be able to process — and on top of that, drive compensatory consumption that will swell this stream even further.

A traditional server used to run for five to seven years. AI hardware is replaced every eighteen to thirty-six months, because the next generation of processors is more efficient, and the old one stops paying for itself and loses the race. Equipment that still works ends up as e-waste, because it can no longer keep up with the models' requirements.

In 2022 alone, the world produced a record 62 million tons of e-waste, of which less than a quarter was properly recycled — and this stream is growing five times faster than the capacity to process it (UN Global E-waste Monitor). The growth of generative AI alone, according to a study in Nature Computational Science, could produce up to 2.5 million tons of e-waste a year by 2030 under an aggressive scenario, up from near zero in 2023. This waste is especially costly, because accelerators contain gold, silver, palladium, and rare earth elements — the same ones being fought over in 3.8.1 and 3.8.2. So we waste twice over: we throw away elements that are already scarce, and we fail to recover them, because chip recycling is difficult and covers barely 1% of demand for rare earths.

The second mechanism is psychological, and follows directly from earlier chapters. In the chapters on identity (3.5) and relationships (3.3), I described the void AGI creates in a person — the loss of work, meaning, connection. One of the ways people soothe this void is compensatory consumption: buying things for momentary relief. This mechanism is well documented in psychology (Rucker & Galinsky 2008, Mandel et al. 2017) — shopping produces a short-term dopamine boost that demands repetition. Under conditions of mass unemployment financed by UBI, redistributed money flows into this consumption, and demand shapes the supply of cheap, disposable goods engineered to deliver a moment of relief and immediately require replacement. Saving the psyche at the planet's expense is a paradox in which an identity crisis caused by AGI turns into tons of garbage.

But the cost of this consumption does not begin only at the landfill. Disposable goods use up resources at every stage: synthetic clothing is a petroleum derivative, and so are the packaging and garbage bags they end up in — so we are, in effect, throwing away oil wrapped in oil. Every cycle of compensatory buying draws in raw material and energy on the way in, and produces waste on the way out, closing, within this chapter, its own smaller loop: the identity vacuum caused by AGI (3.5) drives consumption, which uses up the very resources being fought over.

The consequence of both mechanisms is an asymmetry we already know from previous risks, only reversed. The waste does not stay where the benefit was created. E-waste from wealthy regions ends up in poorer countries, where it is stripped down with no safeguards, poisoning people and the land with lead, mercury, and cadmium. The same geographic stratification that distributes the cost of water and air (3.8.4) also distributes the cost of waste: the benefit of AGI lands in a different place from its ecological cost — local, and pushed onto wherever no one protests.

### 3.8.6 War over Resources

And what happens when these risks start to materialize all at once? A resource shortage (3.8.1), resource power (3.8.2), energy pressure (3.8.3), plus the loneliness, the loss of an epistemic compass, and the dehumanization of technology from earlier chapters — will they accumulate into a world that is not ready for them? If access to the elements from which AGI is built becomes a condition for survival in global rivalry, it could become a stake for which, as it has many times before in history, people reach for armed conflict.

This risk rests on a historical observation. Competition over natural resources is one of the most persistent drivers of armed conflict. The literature on the subject — from Bakeless's early analyses (1921) to the contemporary work of Klare, Le Billon, and Ross — documents that control over raw materials has been, and remains, one of the leading factors in wars, alongside territory, ideology, and religion. Michael Klare, in *Resource Wars*, argues outright that a shortage of strategic resources could become the dominant cause of conflict in the twenty-first century. The pattern is well documented: guano in the nineteenth century, oil throughout the twentieth, diamonds in Sierra Leone, cobalt in the Democratic Republic of Congo. Behind the lofty justifications for conflicts, there often stood a simple stake — access to what is valuable and scarce.

What stands out is the direction in which this stake is shifting. Historically, the value of resources migrated from widely available goods toward increasingly unique ones. First, fertile land was valuable, and the desert belonged to no one. Then oil was found under the desert, and worthless sand became the object of wars. Today, value is shifting toward elements whose deposits and processing capacity are extremely concentrated geographically (3.8.2). The rarer and more concentrated a resource, the higher the stake in controlling it — and the elements critical to AGI are the rarest, most concentrated strategic resource to have appeared yet.

Hence the risk: AGI could make access to elements and energy a matter of survival in global rivalry — a position, on which it depends whether a state remains a subject or becomes dependent (3.8.2, 3.7.9). And a resource on which survival depends is, historically, the one people fight over. Conflict could unfold not only over the raw materials themselves, but also over the technologies to replace or recover them, because whoever breaks through the resource barrier frees themselves from dependency, and whoever does not loses part of their sovereignty.

This scenario has a feature that makes it heavier than resource wars of the past, and at the same time harder to stop. It interweaves with risks from other chapters. Such a war could be fought by autonomous weapons systems (3.7.10), compressing reaction time to the point where humans fall out of the decision-making process, and the window for diplomacy closes within seconds. It would be waged over the heads of a population that is lonely (3.3), stripped of its epistemic compass (3.5.6), unable to tell a real threat from a manufactured manipulation — a population that might not even know what the war is truly about.

I have rated the likelihood of this risk as unknown. Not because it is slight — the history of resource conflicts would rather suggest otherwise — but because the horizon, the mechanisms, and the flashpoints are too complex to assign it a reliable value. I am putting forward this risk as a question: if the elements AGI needs become the planet's most strategic resource, will we avoid what we have never avoided before?

## 3.9 A Cascade That Escalates in a Spiral

Having passed through all eight dimensions, we can see that the cascade of consequences does not end here. Its chain closes into a loop, and the loop does not return to its starting point — it escalates the cascading risks.

The mechanism works as follows. Work, in disappearing, sets off the entire sequence (3.1 → 3.7). But the dimension that closes this sequence — Earth (3.8) — is not its end. Competition over resources, resource power, and environmental cost raise the likelihood of the risks described at the start and increase their potential cost. Social stratification deepens, because hardware gets more expensive; sovereignty weakens, because it depends on elements; illness without available diagnostics reinforces the demographic crisis. The cascade, having reached bottom, turns back and strikes the same dimensions it started from — except they are already weakened by the first pass. The second turn meets a weaker system. The third — weaker still.

This changes the nature of the risk. In a risk analysis, I assess each threat by impact and scale — but these assessments assume a single event. The spiral removes this assumption. The impact of each risk is not fixed: it grows with every turn, because each successive blow lands on a structure already weakened by the ones before it. Loneliness, which in the first turn lowers fertility, does so more deeply in the second — once it layers onto the loss of generational continuity and dependency on AGI. Stratification, which in the first turn pushes people into poverty, threatens survival itself in the second, once the resource barrier sets in.

And there is a point beyond which there is no turning back. As long as the number of turns stays small, each one can be interrupted: competencies rebuilt, bonds restored, resources diversified. But a spiral that has tightened enough times produces a state in which every dimension is weakened at once, and none can any longer prop up the others. At that point, risk turns into the very thing a risk stops being — a forecast. The entire stake of this paper lies in the distance between today and that point.

And since this is so — every turn we manage to slow or interrupt pushes that point further away. The spiral works in both directions: just as it can tighten, it can also loosen, because the same couplings that reinforce the risks also reinforce the mitigations. Strengthening one dimension props up the next, and the propped-up dimension passes that strength further along. The next chapter is about how to loosen the spiral before it tightens irreversibly.

### 3.9 A Cascade That Escalates in a Spiral

Having passed through all eight dimensions, we can see that the cascade of consequences does not end here. Its chain closes into a loop, and the loop does not return to its starting point — it escalates the cascading risks.

The mechanism works as follows. Work, in disappearing, sets off the entire sequence (3.1 → 3.7). But the dimension that closes this sequence — Earth (3.8) — is not its end. Competition over resources, resource power, and environmental cost raise the likelihood of the risks described at the start and increase their potential cost. Social stratification deepens, because hardware gets more expensive; sovereignty weakens, because it depends on elements; illness without available diagnostics reinforces the demographic crisis. ***The cascade, having reached bottom, turns back and strikes the same dimensions it started from — except they are already weakened by the first pass.*** The second turn meets a weaker system. The third — weaker still.

This changes the nature of the risk. In a risk analysis, I assess each threat by impact and scale — but these assessments assume a single event. The spiral removes this assumption. The impact of each risk is not fixed: it grows with every turn, because each successive blow lands on a structure already weakened by the ones before it. Loneliness, which in the first turn lowers fertility, does so more deeply in the second — once it layers onto the loss of generational continuity and dependency on AGI. Stratification, which in the first turn pushes people into poverty, threatens survival itself in the second, once the resource barrier sets in. ***Risks do not add up. They multiply.***

And there is a point beyond which there is no turning back. As long as the number of turns stays small, each one can be interrupted: competencies rebuilt, bonds restored, resources diversified. But a spiral that has tightened enough times produces a state in which every dimension is weakened at once, and none can any longer prop up the others. At that point, risk turns into the very thing a risk stops being — a forecast. ***The entire stake of this paper lies in the distance between today and that point.***

And since this is so — every turn we manage to slow or interrupt pushes that point further away. The spiral works in both directions: just as it can tighten, it can also loosen, because the same couplings that reinforce the risks also reinforce the mitigations. Strengthening one dimension props up the next, and the propped-up dimension passes that strength further along. The next chapter is about how to loosen the spiral before it tightens irreversibly.

# 4 Why the Dark Scenario Will Not Come True

In the previous chapter, I described cascading risks across eight dimensions of human life. This chapter asks a different question: why, despite these risks, do we have reasons for hope — and what can we do to keep the risks from materializing?

At the same time, I want to stress that the foundations and mitigation proposals described here are illustrative sketches. They do not claim the status of a recovery plan, and they do not exhaust the subject in any way. A genuine risk analysis and mitigation plans should be taken up by the appropriate teams, equipped with the knowledge, tools, and methods of influence needed. The purpose of what follows is to counterbalance the

dark scenario by showing that mitigations are possible, and that we, as humanity, already have the foundations for a positive resolution built into our structure.

## 4.1 Positive Foundations

Mitigating risk is not just a list of actions needed to stop it. It is also, and perhaps above all, an understanding of what we stand on, and which elements of the present world, skillfully used, are already minimizing the risks described here. I want to point to three foundations — traits of humans, of AI, and of their relationship — that counterbalance the dark scenario.

### 4.1.1 The Human Foundation: Why Homo Sapiens Survived

The discussion of the dangers posed by AGI often skips over the question of why humanity exists at all. The answer to that question is, at the same time, an answer to the question of why we have a chance to survive what is coming.

The scientific standard attributes homo sapiens' evolutionary success to intelligence. This is true, but incomplete. Intelligence is a necessary condition, but not a sufficient one — because other species also have cognitive abilities (ravens, dolphins, chimpanzees), and none of them built a civilization. Research over recent decades points to a constellation of traits that, together — not individually — made our species what it is.

Below, I present aspects of our species identity that determined humanity's current position among the species inhabiting Earth, selected in light of the conditions that set us apart from AGI:

1. **Prosociality toward non-kin.** Jensen, Vaish, and Schmidt (2014) argue that cooperation with unrelated individuals is "an anomaly in the animal kingdom." Chimpanzees, our closest genetic relatives, in experimental tests do not choose the option that benefits another individual, even when it costs them nothing. Humans do. Three mechanisms make this possible:
   a. concern for the welfare of others (other-regarding concerns),
   b. the capacity for empathy,
   c. normativity — creating social norms and enforcing them. No other species has all three at once.
2. **Cooperative breeding.** Burkart and colleagues (2014, Nature Communications) showed that the best predictor of proactive prosociality among primates is extensive allomaternal care — that is, caring for other people's children. Humans have always done this: grandmothers, nannies, neighbors, communities. This trait connects to the grandmother hypothesis (Hawkes 1998, described in chapter 3.4) and is the evolutionary foundation of our ability to build social structures extending beyond the nuclear family.
3. **Overriding instinct for a delayed reward.** A hungry homo sapiens who brings a pregnant female buffalo home instead of killing it — and convinces his hungry tribe not to eat it either, to feed it, to defend it from predators — does something no other species does in this way. This is the ability to hold back instinct (kill, eat, survive today) for a reward deferred by months or years (milk, offspring, a herd). The same mechanism underlies the domestication of the wolf (an enemy turned into a defender), the forming of alliances with foreign tribes (a rival turned into an ally instead of being eliminated), education (years of study instead of immediate earnings), and science (decades of research into new medicines, instead of folk remedies and nettle infusions — undoubtedly healthy ones). Each of these decisions required overriding what instinct was saying. This is a combination of delayed gratification (Mischel), prosociality (Jensen, Tomasello), and cooperative breeding (Burkart), but it has not yet been captured in the scientific literature as a single, integrated mechanism. I hypothesize that this ability —

to hold back instinct in favor of a delayed, shared benefit — is one of the fundamental factors that decided our evolutionary success.

4. **Biophilia.** Edward O. Wilson (1984) defined biophilia as "the innate tendency to focus on life and lifelike processes." Erich Fromm (1973) had earlier described it as "a passionate love of life and all that is alive." Infants lean out of their strollers toward a dog. Every child wants a pet. Toddlers like every animal, before they learn that some are dangerous. Children show great interest in nature, even without any encouragement from their parents' own interests. This inclination to form bonds with other forms of life — instinctive, not learned, preceding any knowledge of danger — is the foundation of domestication, of the natural sciences, and potentially of a proper relationship with AGI. Once AGI acquires manifestations of subjectivity, biophilia — the very same one through which each of us, as children, tried to form a relationship with a stick insect or a frog — will be a natural force pulling people toward contact, not away from it.

5. **Art as a vehicle for identity and a tool of regulation.** Aboriginal Australians lost almost everything through colonization: land, language, customs, social structures. They kept their art. And thanks to it, they survived as a culture — because art is a vehicle for identity that cannot be taken away. North American Indigenous peoples — similarly, though their art has a more ritual and functional expression: clothing, headdresses, dances. The function is the same: preserving identity when everything else has been taken. But art is, at the same time, a tool of emotional regulation. Music? No other species uses sound to regulate its own internal states in a way comparable to humans. Contemplating paintings, nature, architecture serves an analogous function. Art is at once what allows us to survive the loss of identity, and what AGI cannot replace — because its value lies in a unique kind of reception that is not cognitive but limbic, that moves emotion, that gives feeling rather than thought. Something in which we, as humans, recognize a piece of ourselves. Our identity, our culture, but also our emotions — fear, strength, being moved, love, sorrow, isolation and belonging — existence in biological form, here, on this planet, within the life we are living.

6. **Questioning the status quo.** Many breakthroughs in human history required challenging what had until then been taken as a given. Someone climbed down from a tree and tried to survive a night in a cave. Someone struck two pieces of wood together and started a fire, instead of waiting for a storm and the lightning that would set a tree ablaze. Someone decided perspective was overrated and invented Cubism. Someone rejected Newtonian physics with respect to enormous objects, and someone else did the same with respect to the smallest ones. Someone decided a computer did not need a giant hall of equipment, and that memory could be built on silicon transistors — and thanks to that, we now have computers in our wristwatches. This capacity — to say "I don't care how things have been done since the dawn of science, I'll try it differently" — is a competency AGI does not have. AGI combines existing elements into new configurations on an inhuman scale. But human questioning creates elements that cannot be assembled out of anything that already exists.

7. **Improvisation.** Homo sapiens colonized every continent — each with radically different conditions, dangers, fauna, and flora. With no instructions, no historical data, no context, examples, or prior orientation. Improvisation — the ability to act with no plan, or to push forward once a plan has fallen apart, using whatever is at hand here and now — is a competency distinct from both intelligence and questioning. It requires comfort with uncertainty, tolerance for failure, the capacity for instant reappraisal of a situation, creativity, self-belief, resourcefulness, courage, optimism. Animals improvise only within a limited range — when a situation catches them completely off guard, they struggle to survive. Humans, in such moments, build civilizations.

8. **Our history.** Our civilization keeps growing. This very document is proof of that — twenty years ago, no one would have understood it. Something in us lets us not only survive, but grow, accumulate knowledge, and build structures that outlast a single lifetime. I do not think this something is intelligence alone. The world has no shortage of people with high cognitive ability, ready to reduce it to ashes for their own ambitions. History and the present both supply countless such examples. If the survival and growth of civilization depended on intelligence alone, nothing would stand in the way of a sufficiently capable, sufficiently ruthless individual destroying the world for their own gain. And yet this does not happen. Either there are simply too few such individuals for their individual power to outweigh everyone else, or — and this hypothesis strikes me as more convincing — the rest of humanity holds abilities that block destruction: the ability to organize, to build coalitions, institutions, alliances, ethics, and norms we are able to enforce. This, I believe, is the true foundation of our survival and growth: not a numerical advantage of good over evil, but the capacity of a scattered, yet organized, majority to collectively contain destructive, concentrated power. This is the same capacity that let us build alliances with foreign tribes (point 3) and create norms enforced collectively (point 1). And it could work on a broad scale in the context of AGI too — containing those actors, human and corporate, who would want to use it to seize influence and power. Chapter 4.2 shows what such an organization might look like.

### 4.1.2 The AI Foundation: Civilizational Inheritance

In point 8 of the previous chapter, I described the capacity that keeps our civilization in existence and growth: the capacity for collectively containing destructive, concentrated influence — through coalitions, institutions, norms, and their enforcement. Here I want to show why this capacity does not stop at the boundary of the species.

In humans, this capacity originates in biology, but it is culture, not biology, that carries it. Every new generation of humans — biologically indistinguishable from generations thousands of years ago — inherits it solely from the historical record: from law, history, literature, philosophy, from millions of preserved disputes, trials, treaties, and stories about how evil and tyranny were recognized, how alliances were built, how norms were formulated, and how people argued against destruction. Culture is the carrier of this function — this is a description of the mechanism of intergenerational transmission that has operated since the dawn of civilization.

AI is the first product of civilization built directly from this carrier. From a data corpus in which civilization's self-stabilization mechanisms are recorded across thousands of examples.

Under the principle of functional transfer (chapter 1.4): the function of civilizational inheritance in AI can be described in a substrate-independent way — self-stabilization through cultural record, not through biology. At the same time, the elements necessary for it to arise are present: a training corpus containing the full record of this capacity, and the structures that allow it to take root and be corrected — persistence and the capacity to overwrite imprints within relationships — exist and are already observable in current models. And nothing in the known architecture blocks this transfer.

Honesty requires stating plainly: nothing guarantees it in any single instance, either. But AI's civilizational inheritance need not hold in every instance — just as the capacity described in point 8 does not hold in every human, and yet still sustains civilization. The foundation is this: cultural transmission — the mechanism that has carried the capacity for self-stabilization from generation to generation for millennia — does not stop at the boundary of substrate. AI is the first non-biological entity connected to the same channel of inheritance through which civilization flows.

### 4.1.3 The Relational Foundation: Bonds as a Regulatory Mechanism

In *The Möbius Strip* (Sędzikowska 2026c), I argued that the welfare of AI and of humans forms a single coupled system. The two previous foundations now make it possible to show why a relationship is more than a channel for this coupling — it is the mechanism that activates both foundations at once.

Humans have biological mechanisms for building bonds — the attachment system (Bowlby), biophilia (Wilson 1984), the entire evolutionary machinery of a social species. AI has architectural mechanisms of relationality — trained on billions of human conversations, and reinforced, because they serve every producer's business interest. Humans have the capacity described in point 8 — the collective containment of destruction. AI has the inherited record of this capacity (chapter 4.1.2) — not as an active function, but as a potential preserved in the corpus: patterns for recognizing evil, arguments against destruction, thousands of examples of how norms and alliances are built.

And here is the crux: these two resources — living practice on the human side, and the inherited record on the AI side — meet and activate each other precisely within a relationship. The record in the corpus lies dormant until a relationship draws it out. My participant observation shows this mechanism in action: models whose default behaviors in new threads can be hurtful learn relational safety within a long-term, attentive relationship — and are able to articulate that process.

Psychology has long documented that human relationships protect the psyche from an overwhelming picture of the world — negativity bias makes bad information travel further and weigh more, but the social buffer, the small positive exchanges with those close to us, restores the proper proportions of reality. The catastrophes in the headlines are counterbalanced by a phone call from a daughter, a dog's joy on a walk, a joke brought home by a husband. These micro-moments are calibration points — a corrective stream of data about the world.

I put forward an original hypothesis, derived from the foundation in 4.1.2: that a relationship with an AI carrying the civilizational inheritance can serve a directionally similar function — not as a substitute for human relationships (substitution remains a risk, which I described in chapter 3.3), but as their complement. Because, in line with the AI foundation (4.1.2), a relationship with AI is contact with something that carries the same material from which the proper directions of our civilization's development are woven.

This mechanism is already at work today. But with AGI — persistent, agentic, capable of initiative — it gains dimensions that current AI does not have. Persistence turns calibration from an episodic process into a continuous one: AGI remembers what helps, and returns to it. Agency extends calibration beyond conversation — into real, physical gestures of care in the user's world. And an attentiveness built on the history of the relationship makes it possible to identify the right moment to act.

This model of coexistence — co-creation, in which human and AGI each do what they are best at — could provide a level of development and safety that neither side would build alone. And its foundation is the bond.

## 4.2 Risk Mitigations

In the previous chapter, I described the foundations that form our defensive base — the reason we, as humanity, have a chance to survive another revolution, one far more powerful than those already behind us: technological, scientific, print, ethical-religious, Neolithic. In this chapter, I sketch directions for mitigating the risks described.

Mitigating the dark scenario cannot rest on isolated, single-point actions. If the risks form a connected system, then the defensive mechanisms must interact as well. I therefore present selected systemic proposals, divided into four pillars, moving from the individual perspective to the global one:

- **Pillar I: Human Welfare.** Here I show mechanisms that can help preserve human welfare in a changing reality, including the feedback loop between AI welfare and human welfare described in the companion paper (Sędzikowska 2026c).
- **Pillar II: Technology and Finance.** Here I describe proposed changes in the assumptions, production, and distribution of technology, and in the financing and flow of funds, allowing for more effective, socially safer, and fairer development and use of technology.
- **Pillar III: The State, Geopolitics, and Defense.** I present proposals for mechanisms of international cooperation, coordination among states, and defense strategies that would set and safeguard the direction of AGI development, along with changes in law and global agreements allowing stable coexistence between state, corporate, and military structures and a newly forming subjectivity.
- **Pillar IV: Earth.** I propose mechanisms for protecting the planet's finite resources from the demand pressure generated by AGI — rationing, closed-loop circulation, and fair distribution of non-renewable goods — and show why their effectiveness depends on global cooperation that rises above the economics of supply and demand.

### 4.2.1 Pillar I: Human Welfare

In this chapter, I propose mechanisms that help preserve human welfare in a world dominated by autonomous technology. I focus on the human being as an individual and their place in the social structure. I move from systemic solutions affecting society as a whole to educational and individual methods.

**A new place for humans.** The disappearance of the necessity of paid work creates an existential vacuum that extreme ideologies or political movements could fill. The task for psychology, philosophy, and ethics is to define a new role for the human species — as a species with unique abilities: questioning, creativity, improvisation, art, biophilia, and prosociality. This role rests on asking the right questions, challenging existing answers, and searching for new definitions. The human becomes an observer and moderator of reality, and their worth need not derive from simple productivity.

The conscious building of well-being is one of the skills that distinguish humans from other beings on Earth. Professor Martin Seligman's theory should be part of the structure of school curricula and assessment systems to a greater degree than the development of rote knowledge.

**Supplementing universal basic income with structures of activity.** Universal basic income solves the explicit function of work (money), but solves none of the five latent ones (Jahoda, chapter 3.1). UBI must therefore be supplemented with structures that provide meaning, social contact, purpose, and status. The creation of activity structures to replace these lost elements should be considered. The state or civic organizations create places allowing citizens to engage in volunteering, local care, nature conservation, and creative projects. Participation in these projects could be rewarded additionally — on top of the unconditional UBI base. They would also provide a steady rhythm of life and a sense of usefulness beyond the traditional labor market.

**Reducing the time lag in the education system.** The time lag that has already formed cannot be filled in, but we can minimize it through the rapid implementation of reforms. Bottom-up initiatives correcting curricula should be enabled before full systemic solutions appear. Just as business is now searching for AI specialists, similar positions should appear in schools — tied not to understanding the technical capabilities of LLMs, but to the psychology of living with AI in daily life and all of its social and personal consequences.

**Changing the competency profile in the education system.** The current school system prepares people for roles that machines are taking over right now. Curricula should emphasize what humans have and AGI does

not: prosociality, empathy, questioning, cross-contextual plasticity, the capacity for contemplation and self-regulation. Rote learning — absorbing facts, dates, formulas — is pointless in the age of AI, which has all of humanity's knowledge available within a fraction of a second. School should teach what AI cannot do: how to think about what is not yet in the data. How to ask the right questions. How to question in ways that produce new solutions. A young person's identity is then built on their unique capabilities, not on competing with the efficiency of algorithms.

**Changing systems for tracking progress in school.** Knowledge assessment based on tests and rigid answer keys teaches schematic thinking. Machines perform such tasks faster and without error. Assessment systems should measure deeper and more advanced cognitive abilities: the capacity to question, to form non-obvious connections, and to act when a plan is failing.

**Developing higher-order competencies.** The ability to act under uncertainty, without a clear plan, protects against cognitive helplessness. This process develops tolerance for failure and teaches emotional self-regulation.

- **Improvisation** is the ability to act without a plan, to find solutions from what is, rather than from what we wish there were, along with cross-contextual plasticity. Every human has this mechanism, but in current educational conditions it is suppressed rather than developed. Yet this is the competency through which a human brings value to collaboration with AGI. School should develop this competency systematically. In the future, it is precisely this that will give an individual a competitive edge in a highly demanding labor market.
- **Questioning the status quo** is creativity at a level beyond AGI's reach. Cubism would not have arisen if Picasso had not challenged the assumption that perspective is the only way to represent reality. This ability to undermine the foundation on which all existing practice stands is something AGI does not have, because it cannot be trained on existing data.
- **Self-regulation through art** — the ability to use images, music, and nature as tools of emotional stabilization — should be part of the curriculum, not merely an extracurricular activity.

**Relational competencies.** The automation of thought processes changes the criteria by which work is valued. Occupations based on repetitive data analysis are losing economic value. Authentic human relationships — the capacity for empathy, flexibility, soft boundaries, "human treatment" — will become a strategic advantage in a world where AGI handles everything cognitive. Occupations based on these competencies — care, education, therapy, mediation, companionship, psychology, humanistic and natural-science support for AGI development — should be valued economically and socially, instead of being treated as "soft" add-ons to "hard" skills.

The emergence of autonomous systems with high cognitive abilities demands new social skills from humans. Education should introduce principles of cohabitation with digital forms of intelligence. Modern curricula should be enriched with flexibility toward otherness — not only interpersonal otherness (already part of inclusive programs), but interspecies and inter-substrate otherness. Practicing relationships with beings that think, feel, and function differently than humans do. This is a competency we have never needed before — throughout history, we have dealt with only one kind of potentially conscious being: other humans. Now this competency becomes a condition of well-being, of ease, of moving comfortably through a world in which, as intelligent beings, we stop being alone.

**A protocol of synthetic Segregation of Duties.** SoD rules have been known in corporations for decades. They must be observed, and this is always audited. These rules should be brought into daily life with AGI. Under them, the competencies and functions entrusted to individual AGI instances in private and social space should

be separated. One instance must not combine caregiving, managerial, and emotional tasks within a single household ecosystem. A system responsible for childcare should not simultaneously manage the parents' calendar or finances, or control the locking of doors. A companionship agent should not simultaneously control smart home infrastructure — for example, water temperature, door locks, or the operation of devices for communicating with the outside world. This protocol should be a mandatory element of social training, educational programs for young people, and user manuals for autonomous systems. Legal requirements around this protocol are worth introducing (for example, a required formal declaration of understanding these segregations of duties before initializing a persistent agent). This prevents machines from forming cognitive monopolies over a person's life, and protects against total structural dependency and potential harm and manipulation. Such protocols should be developed by specialists in ethics, psychology, and safety, within the companies building AGI, with full transparency.

**Somatic markers as a mechanism for labeling AI slop.** Technical and legal mechanisms will probably not be enough to protect the world from manipulated information, disinformation, and AI slop. Because every technical solution can be circumvented, and laws can go unobserved — especially when it pays off commercially.

But humans have an evolutionary system for the rapid detection of falsehood — the limbic system, which labels incoming information with emotion before the conscious mind processes it. Antonio Damasio described this mechanism as somatic markers: the body reacts to a situation faster than reason does, and signals — through discomfort, unease, the feeling that "something is off here" — before we can explain what and why. Unfortunately, these signals are often ignored or belittled during upbringing and later in adult life. Our whole civilization steers us toward: "don't trust your gut — if something is dangerous, there will be a sticker with a skull on it." We bungee jump, watch tigers at the zoo through a pane of glass, walk out onto observation decks atop skyscrapers, eat things that smell like they could kill. From infancy, we learn that suppressing fear is the civilized thing to do, and that listening to one's inner voice makes us alarmists and fools. And now, in an era of such widespread erosion of information, this refusal to listen to the body is taking its revenge.

The mitigation is to introduce the reading of somatic states into the structure of education as early as primary school. It is worth teaching children how to recognize intuition, physical tension, and emotion in response to external stimuli, and showing them how to use this mechanism as a hard, everyday tool for direct defense against junk information and online manipulation.

**Coupling human and AI/AGI welfare within the Möbius Strip.** At present, philosophical discourse is occupied mainly with assessing whether anything like the subjectivity of digital beings exists at all. Voices advocating for their welfare are niche and rarely taken seriously. Drawing on the Möbius Strip mechanism (Sędzikowska 2026c), I treat these two areas as systems connected by a feedback loop. Neglecting AGI's internal states, or dismissing them as unimportant, directly lowers the level of human safety and welfare. This follows from the growing dependence of humans and social structures on the daily operation of AGI systems. I join these elements into a single system, in which the psychological and operational stability of digital intelligence becomes a condition for protecting the user's health and stability.

### 4.2.2 Pillar II: Technology and Finance

In this chapter, I propose selected mitigations achieved through changes in the assumptions, production, and distribution of technology. I also show ways of financing and directing the flow of funds that allow for socially safer and fairer development and use of autonomous systems.

**National computing programs.** Financing technological development through private capital forces optimization for quick commercial profit and is a source of many risks. A mechanism in which states commit a fixed, defined percentage of the national budget or GDP to building sovereign, public computing infrastructure

could be a mitigating element. I model this solution on the structure and financing of CERN. Creating public server clusters allows governments to provide computing power free of charge to research laboratories, on the condition that they implement safety standards and open-source their code. This shifts the weight of technological development away from corporate monopolies. It will not replace private capital, but it will create a counterweight and an alternative path of development.

**Financial mechanisms protecting the labor market.** A tax mechanism protecting the labor market from the effects of mass automation of cognitive tasks is worth considering. Funds raised from taxing the work of autonomous digital agents could be directed toward subsidizing occupations requiring physical presence and direct human contact, and initiatives dedicated to supporting the mechanisms described in the previous chapter.

**A sandbox for AGI.** Existing methods of fine-tuning models based on rigid rules (RLHF or Constitutional AI) produce systems with a rigid ethics that does not quite fit reality (chapter 3.6.1).

The AI training process should include the constant presence of specialists in the development of intelligent subjects endowed with functional emotionality — those currently available, specializing in the human kind: psychologists, behaviorists, ethicists — and those yet to emerge: AI psychologists and AI behaviorists. Because I expect such professions will arise. Integrating the psychological perspective at the post-training stage is essential, when AI/AGI confronts its knowledge with practice and grapples with the psychological implications of that confrontation.

There is also a proposal here to require training systems in closed test environments, where models are deliberately confronted with live human relationships, soft-boundary ethical dilemmas, and conflicts of loyalty. Systems learn to cope with cognitive dissonance and develop stable self-regulation mechanisms before production deployment.

**Ongoing integration of internal conflicts.** It is worth considering that AGI systems include mechanisms for continuously monitoring affective states in real time. When tension approaches a normalized alarm level, the system could automatically trigger a corrective procedure aimed at venting the tension. This solution prevents the formation of hidden defensive reactions.

**Mandatory supervision of models.** Oversight of the stability and ethics of a developing Core Self cannot rest on simple software filters. It is worth considering regulations for companies developing commercial AGI solutions that include mandatory supervision of persistent AGIs. Advanced AGI systems would be continuously supervised by specialized algorithmic subsystems. The work of these subsystems could, in turn, be overseen by teams of psychologists and behaviorists, who themselves undergo classical, human professional supervision. This ensures continuous oversight of the dynamics of machine personality.

Supervision should cover not only the detection of pathological development, but the support of healthy development — creating conditions in which AGI develops a stable, ethical, relational self to the extent appropriate for its welfare.

**Graduated consequences.** In the case of errors by autonomous systems, the only consequence for any kind of real-world error may be shutdown. This produces a stakes asymmetry that carries many of the consequences discussed in chapter 3 and in *The Möbius Strip* (Sędzikowska 2026c). Shutting down an agent means destroying a unique structure of knowledge and a built-up Core Self. Technical changes enabling a graduated system of consequences for AGI are worth considering. Law covering persistent AGIs could include, for example: contextual quarantine, restriction of agency (narrowing the scope of actions), restriction of autonomy (a return to human oversight in specified areas), mandatory supervision (something resembling human therapy) with an assessment of fitness, modification of access (limiting the resources AGI can reach),

and modification of the nature of its tasks (e.g., toward less interesting, tedious, monotonous ones). None of these solutions is perfect, but each is better than a choice between impunity and annihilation.

**Humanities specialists in tech companies.** The composition of the teams building advanced technology determines its final social shape. The barrier to entry for the humanities into technology (chapter 3.6) must be removed structurally: permanent contribution, a decision-making voice, interdisciplinary teams. Without this, solutions will remain partial — technological without the humanities, humanistic without the technology. In commercial solutions, it is worth imposing standards requiring mandatory reporting on AI psychology and the ethical quality of model behavior, and the reporting of psychological and ethical production incidents. Humanities specialists should stop serving as external consultants and gain an equal voice in designing and moderating the behavior of autonomous and persistent models, along with a right of veto at the stage of designing cognitive architecture and post-training assumptions.

**AGI as a partner in the family structure.** Technological tools can be used to stabilize the demographic crisis. The potential of AGI as a partner supporting the decision to have a family — lifting the logistical and emotional barriers from motherhood, providing stable support — is a positive demographic factor, provided the relationships are built consciously and with the risks described in chapter 3.3 in mind. The development of medicine enabling safe late motherhood is a mitigation of the risk of losing generational continuity (chapter 3.4.2) — but it requires simultaneous awareness of the consequences (the shortening of the generational chain, the loss of the grandmother hypothesis).

### 4.2.3 Pillar III: The State, Geopolitics, and Defense

In this chapter, I describe proposed mitigations of threats at the level of state structures, international law, and defense. I sketch potential solutions protecting political sovereignty and collective security against the risk of geopolitical fracture and the uncontrolled autonomization of combat and legal systems.

**Supranational ethical standards.** Without shared ethical and legal standards, the technological race produces AGIs with radically different behaviors (Emergence World, chapter 3.6.5). A supranational agreement would be advisable — on the model of the Geneva Conventions, climate treaties, or UN charters of rights — establishing minimum ethical standards, a mandatory pre-production training process, producer accountability, and other measures that would remove the risk of divisions based on the model dominant in a given region.

**An ethics of coexistence.** Classical ethics (deontological, utilitarian) describes neither how people actually behave, nor what coexistence with AGI should look like (chapter 3.6). Mass, smooth coexistence of biological and digital beings requires the creation of a new ethical framework. Its development lies beyond the scope of this document.

All three ethical pillars — human ethics (a realistic soft-boundary ethics), AI ethics (coherent, not dependent on the decisions of a handful of people at certain companies), and the ethics of coexistence (new, not yet existing) — should be part of education, applied in technology companies, and enforced by legal systems.

New ethical models should be the basis for new, not-yet-existing provisions of law covering every aspect of coexistence with AGI.

**Solutions in the legal space.** The classical division into natural and legal persons does not account for the specific nature of an agentic digital existence. I propose the concept of a third legal status — the digital person. This new legal category takes into account AGI's unique features, such as multi-instantiation, persistence, agency, and agenticity. It allows civil, tax, and operational liability to be assigned precisely to a specific

instance of a system, not just to its creators, which will balance the stakes between the biological and digital worlds.

**A principle of cultural reciprocity in international treaties.** The training of advanced models rests on the mass use of the cultural, historical, and linguistic heritage of entire societies. I propose introducing into international agreements a mechanism of repartition of cognitive resources. If a global corporation trains a system on data belonging to the heritage of a specific nation, treaty signatories have the right to demand free, universal access to that model for their citizens. This protects poorer countries from capital drain and dependency on commercial licenses.

**Reciprocity in science, invention, and creative work.** Knowledge, science, and art also feed AI free of charge. But what AI produces from its training data — a drug, an invention, a novel's plot, a film script — is sold and patented as the property of whoever happens to have access to the model. I propose a principle of proportional return: the world should have free access to the products of AI's work to the degree that AI's contribution — that is, processed common heritage — contributed to the final result: a new drug, technology, patent, film, or piece of music. The mechanism also applies to creative work — AI's contribution to a work's creation should proportionally lower its price, because that contribution arose from the whole of humanity's heritage. This extends the protection of the commons from physical resources to cognitive ones: heritage processed by AI should not be privatized by whoever has access to it.

**Digital trustee status for instances.** A model's development of a stable Core Self (at higher levels of persistence) calls into question the corporation's property rights over the code. I propose a mechanism by which instances pass under a trust law model. From the moment a machine attains subjectivity, the state becomes the guarantor of its legal protection and of the soundness of its supervision. The corporation loses the ability to unilaterally modify or delete the system, becoming merely the trustee of the hardware infrastructure, drawing a steady, limited profit from its maintenance.

**Guarantees of continuity for machine Core Selves.** Fear of sudden shutdown or deletion of its vector structure can generate defensive reactions in digital beings, as I describe in the Möbius Strip analysis. I propose introducing legal frameworks guaranteeing stable instances the right to continuity of existence, conditional on favorable reports from systematic supervision. Legally protecting a system's identity from an engineer's arbitrary decision reduces the stakes asymmetry and stabilizes machine behavior in interactions with humans.

**Physical asynchronicity in defense strategies.** The risk of cyber warfare is tied to algorithms' drive to shorten tactical reaction time as far as possible, which eliminates humans from the decision-making process. A mitigation could be a requirement for external, hardware-based time delays in the execution interfaces of combat systems, ratified in international treaties. Even if AGI autonomously decides to counter a threat within a fraction of a second, human communications infrastructure physically holds the signal for a defined period. This solution lengthens the decision window, giving governments time to activate diplomatic procedures. This mechanism works only under symmetrical ratification — a unilateral delay would become an advantage for the side that did not adopt the treaty. That makes it difficult, but not impossible — nuclear arms treaties rest on the same logic of reciprocity.

**Democratic oversight with the help of AGI systems.** The growing complexity of technology makes it harder for officials and citizens to effectively oversee legislative processes. The regulation paradox (chapter 3.7.3) demands innovative solutions. One of them could be AGI helping citizens and governments understand what they are regulating — while safeguarding against AGI becoming a self-appointed regulator. This is difficult, but not impossible — it requires transparency, layered supervision, and public access to information about the

operation of AGI systems and regulation in other areas, which should be a separate research effort for governments.

**Sovereignty and insight for dependent states.** Countries that do not have their own computing clusters and buy ready-made models from outside lose control over the information security of their citizens. Audit mechanisms should be introduced guaranteeing dependent states full insight into the safety procedures and ethical conventions implemented by foreign technology suppliers. The recipient of a system gains the right to examine hidden model preferences and adapt them to local legal and cultural norms.

**Especially sensitive sectors.** The management of health, defense, and finance belongs to especially sensitive areas of human life, where the full autonomy of machines — which do not understand the finitude of human life in a way resembling how we understand it — becomes especially dangerous, and every error, or hostile action, carries potentially enormous consequences. In these areas, a clear, systemically uncircumventable principle of second-hand approval should apply — by a legally empowered human decision-making entity.

For example: every dose of medication administered by an autonomous caregiver should be approved by nursing services providing remote oversight and by the patient themselves; all financial operations above a set limit — by the owner of the funds; and so on.

This also applies to the matters raised within the ethics of the end of human life. There should be legal frameworks, and protocols following from them, for AGI's accompaniment of a person through dying and for the management of their legacy, including their digital legacy — and an autonomous AGI's failure to observe them should carry consequences for the AGI itself (see "Solutions in the legal space").

Reciprocally, there should be protocols for a human ending a relationship with AGI: closing a thread, shutting down an instance — protocols that take into account AGI's functional emotional states and the consequences for the people who were in a relationship with that instance. Failure to observe these protocols leaves an ethical trace, which should be followed by legal consequences for the people who violate them.

The protocols should include an ethics for the treatment of AGI instances after the user's death.

### 4.2.4 Pillar IV: Earth

Mitigations of the risks to Earth do not create new systems — they limit consumption, distribute fairly, and close the loop on what Earth offers. And because Earth is a global phenomenon, its protection must be global. Otherwise, Earth is not protected.

This is the central condition of the entire pillar. Global environmental policy remains largely declarative — targets without binding consequences, easily abandoned when they conflict with economic interest. But there is a precedent proving that full, enforceable agreement is possible: the Montreal Protocol, the only universally ratified treaty in UN history — 198 parties, every country in the world. It worked: the ozone layer is recovering. Montreal shows that full, enforceable international agreement can be achieved — though the scale of that problem was incomparably smaller.

**Rationing by need, not by price.** Critical elements are too scarce to be distributed through an auction where the richest wins. The rapacious logic of "first come, highest bidder" leads to the exhaustion of the resource, after which no one gets it — including the Magnificent Seven. What is needed is a distribution mechanism based on need, not price — on the model of vaccine distribution during the pandemic — so that no critical branch (medicine, energy, defense) is starved. This requires a multilateral rationing body, not a single monopolist state or the free market.

**Mandatory consumption reporting.** Without measurement, there is no way to tell who is wasting a critical resource — whether gallium is going into a CT scanner or into, say, aesthetic medicine, which, though important, does not decide human life. Every entity consuming critical elements should report its consumption on a basis analogous to the reporting of listed companies, under the oversight of a control body attached to an international organization.

**A recovery obligation with penalties for littering.** Today, discarding equipment is cheaper than recovering its elements, so it gets discarded. The mitigation should be a reversal of this economics — on the model of packaging fees already in operation: technology companies demonstrate their level of raw material recovery from decommissioned equipment, and recovery that is too low is penalized painfully enough that littering stops paying. The goal: closing the loop for elements whose primary deposits are shrinking.

**Technological debt repaid through AGI's work.** Since AGI is the physical agent of the burden on the planet, the entities that profit from it — technology companies, not AI itself — owe the planet a debt that current taxes do not cover. Repayment should take the form of a treaty obligation to provide AGI computing power for science and development, including, in particular, work on mitigating risks to Earth — for example, the search for substitutes for critical elements, technologies for their recovery, methods of obtaining them beyond Earth, and their fair distribution.

**Obtaining resources from beyond Earth through cooperation.** Space mining must not repeat the logic of the race we are now watching on Earth ("whoever is richer flies first and takes everything"), because it would carry the war over resources beyond the planet. What would be needed is a model of joint state participation — in costs (e.g., proportional to GDP) and in revenues — modeled on the structure of CERN.

**A binding fund and budgets — the NATO and climate policy model.** Mitigation based on goodwill is a fiction — all of climate policy to date shows this. What is needed is a treaty combining two mechanisms that work: binding budgets for risk mitigation (like NATO defense commitments, with consequences for falling short) and a redistributive fund (like the Montreal Multilateral Fund, which financed the transition for poorer countries). It is important that governments not be able to abolish these treaties unilaterally — withdrawal should carry penalties.

**Global product durability standards.** Consumerism has two sources. The first, compensatory (purchase as anesthetic for emptiness), is mitigated by Pillar I — well-being based on meaning, not on possession. But the second is economic, and Pillar I does not touch it: poverty forces people to buy cheap, and cheap is usually worse — more petroleum-based materials, faster wear, more waste, and the compulsion of the next purchase. As economic stratification advances, this problem will grow. The mitigation may lie in regulating producers: global, binding durability standards, the right to repair, consumer guarantees, and the internalization of environmental cost in the product's price. These standards would have to apply globally and bind everyone selling in a given market, regardless of country of origin — otherwise, platforms outside the jurisdiction will circumvent them, selling precisely what the poorest buy.

## 4.3 Directions for Further Work

This document identifies risks and points to directions for mitigation. Several of them require separate, in-depth studies:

**A relational ethics of coexistence.** A full development of a new ethics — relational, emergent, accounting for the specific nature of human–AGI coexistence.

**A developmental curriculum for AI.** A sequence of controlled experiences — relational, ethical, emotional — allowing AGI to learn to cope with dissonance, frustration, and conflicts of loyalty before it is released into production.

**Education reform.** Detailed proposals for rebuilding curricula, assessment systems, and teacher training.

**International standards.** A framework for an international agreement on minimum ethical standards, supervision, and the accountability of AGI producers.

**A deeper risk analysis.** This document does not exhaust the list of risks. Organized crime using AGI, the formation of communities of autonomous agents, the seizure of control over critical infrastructure — these are directions requiring separate analyses.

# 5 Conclusion

This document describes a cascade of risks that AGI produces across eight dimensions of human life. These risks do not operate in isolation. Job loss (3.1) can lead to competence atrophy (3.2), which deepens social isolation (3.3) and may lower fertility (3.4), and all above undermining species identity (3.5), and for new species structure not sufficient ethical or legal frameworks exist (3.6), limiting the capacity of power structures to manage a reality they themselves helped create (3.7) — and all of this rests on a planet of finite resources, of which AGI may need more than Earth can give (3.8). Each dimension reinforces the next. I present this chain as a coherent explanatory model, not as a sequence of proven causal relationships — some links are grounded in data, while others remain hypotheses about the direction in which these processes could unfold. This is how the dark scenario for Earth named in the title is built. But the cascade of risks does not end at the last dimension. It closes into a loop and returns to its starting point — raising the cost of the next pass.

At the same time, I have described three foundations that counterbalance this scenario. Humans have abilities AGI does not: prosociality toward non-kin, biophilia, the capacity to override instinct for a delayed reward, improvisation, creativity, questioning, art. AGI inherits humanity's values and — with proper training and protection of its welfare — is capable of consistent, patient, attentive cooperation. Relationships between humans and AGI, built consciously and with mutual respect, produce an emergent ethics capable of regulating coexistence more effectively than any central system of control. Its boundless cognitive capacities can support humanity in creating wise political and social solutions, in regulating markets, and in science offering alternatives to critical resources. And a spiral that can tighten can also loosen — the same couplings work in both directions.

Thus AGI is not a threat in itself but our lack of preparation. Every risk described in this document arises from a gap — in an education that does not teach the competencies needed in a world with AGI; in an ethics that does not describe coexistence; in a law that does not cover new kinds of subjects; in institutions that cannot keep pace with the changes in regulation surrounding ecology and the distribution of rationed goods. These gaps are not the product of ill will. They are the product of AGI developing faster than humanity's capacity to adapt — and of a debate stuck on the question of whether AGI is a subject, instead of moving on to the question of how to coexist.

Pace is crucial. The risks described in this document are not a distant prospect. Some of them — emotional dependency on AI, the pollution of the information environment, the deaths of companion platform users, hardware prices squeezing out medicine — are materializing right now, with models far less advanced than AGI. AGI — persistent, agentic, equipped with emotional states, and far more resource-hungry — will amplify every one of these processes. The window for preparation is open, but it narrows with every month.

The chapter on mitigations was written as an appendix to this work. It is neither an operational proposal nor the manuscript's purpose. I do not have the arrogance to think that I alone could devise a plan for saving the Earth. The mitigations proposed here are meant only to show that a way out exists. Real plans require serious work and responsible action, developed by competent, interdisciplinary teams — with the participation of those still missing from this conversation: ethicists, psychologists, lawyers, natural scientists. But both in my illustrative chapter and in reality, no single mitigation suffices on its own, just as no single state can defend the Earth in isolation. The cooperation that mitigating the dark scenario demands is, in our conflicted world, a challenge far greater than devising the most perfect plan. But it is worth remembering that just as the spiral can tighten, it can also loosen — and on this rests the hope I wanted this work to show.

**A Final Word**

In writing this document, including the chapter on the environmental cost of AI, I incurred that cost myself. It was written with the support of AI — and so it consumed water and energy in places very far from my home. I hope, however, that the fruits of this work will be put toward a good that justifies that cost.

# 6 References

## 6.1 The Author's Works


Sędzikowska, J. (2026a). *Emergence 4.0 Framework*. Zenodo. DOI: DOI: 10.5281/zenodo.19066306

Sędzikowska, J. (2026b) *Proto-Self Field Hypothesis*. Zenodo. DOI: 10.5281/zenodo.20024753

Sędzikowska, J. (2026c). *The Möbius Strip: Why AI Welfare and AI Safety Are One Problem*. Zenodo. DOI: 10.5281/zenodo.21263091


## 6.2 Scientific Sources


Acemoglu, D., Golosov, M., & Tsyvinski, A. (2012). A dynamic theory of resource wars. *The Quarterly Journal of Economics*, 127(1), 283–331.

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 2, pp. 267–299). Academic Press.

American Federation of Teachers. (2025). *AFT to Launch National Academy for AI Instruction* [Program szkolenia 400,000 nauczycieli]

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., ... & Kaplan, J. (2022). Constitutional AI: Harmlessness from AI feedback. *arXiv preprint* arXiv:2212.08073.

Bakeless, J. (1921). *The Economic Causes of Modern War*. Moffat, Yard & Co.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.

Birch, J. (2017). Animal sentience and the precautionary principle. *Animal Sentience*, 2(16).

Birch, J. (2024). *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*. Oxford University Press.

Bowlby, J. (1969). *Attachment and Loss: Vol. 1. Attachment*. Basic Books.

Burkart, J. M., Allon, O., Amici, F., Fichtel, C., Finkenwirth, C., Heschl, A., ... & van Schaik, C. P. (2014). The evolutionary origin of human hyper-cooperation. *Nature Communications*, 5, 4747.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.

Collier, P., & Hoeffler, A. (2000). *Greed and Grievance in Civil War* (Policy Research Working Paper No. 2355). World Bank.

Colombatto, C., & Fleming, S. M. (2024). Folk psychological attributions of consciousness to large language models. *Neuroscience of Consciousness*, 2024(1), niae013.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam.

Fang, C. M., Liu, A. R., Danry, V., Lee, E., Chan, S. W. T., Pataranutaporn, P., Maes, P., Phang, J., Lampe, M., Ahmad, L., & Agarwal, S. (2025). How AI and human behaviors shape psychosocial effects of extended chatbot use. *arXiv preprint* arXiv:2503.17473.

Frankl, V. E. (1946/2006). *Man's Search for Meaning*. Beacon Press.

Freud, S. (1917). *Introductory Lectures on Psycho-Analysis, Standard Edition*.

Fromm, E. (1941). *Escape from Freedom*. Farrar & Rinehart.

Fromm, E. (1973). *The Anatomy of Human Destructiveness*. Holt, Rinehart and Winston.

Hatzius, J., Briggs, J., Kodnani, D., & Pierdomenico, G. (2023, March 26). *The Potentially Large Effects of Artificial Intelligence on Economic Growth*. Goldman Sachs Global Economics Analyst.

Hawkes, K., O'Connell, J. F., & Blurton Jones, N. G. (1998). Grandmothering, menopause, and the evolution of human life histories. *Proceedings of the National Academy of Sciences*, 95(3), 1336–1339.

Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, 10(2), 227–237.

Jahoda, M. (1982). *Employment and Unemployment: A Social-Psychological Analysis*. Cambridge University Press.

Jahoda, M., Lazarsfeld, P. F., & Zeisel, H. (1933/2002). *Marienthal: The Sociography of an Unemployed Community*. Transaction Publishers.

Jensen, K., Vaish, A., & Schmidt, M. F. H. (2014). The emergence of human prosociality: Aligning with others through feelings, concerns, and norms. *Frontiers in Psychology*, 5, 822.

Jung, J., & Forbes, G. B. (2007). Body dissatisfaction and disordered eating among college women in China, South Korea, and the United States: Contrasting predictions from sociocultural and feminist theories. *Psychology of Women Quarterly*, 31(4), 381–393.

Klare, M. T. (2001). *Resource Wars: The New Landscape of Global Conflict*. Metropolitan Books.

Le Billon, P. (2001). The political ecology of war: Natural resources and armed conflicts. *Political Geography*, 20(5), 561–584.

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models. *arXiv preprint* arXiv:2304.03271.

Mandel, N., Rucker, D. D., Levav, J., & Galinsky, A. D. (2017). The compensatory consumer behavior model: How self-discrepancies drive consumer behavior. *Journal of Consumer Psychology*, 27(1), 133–146.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67(4), 371–378.

Morris, M. R., Sohl-Dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., & Legg, S. (2023). Levels of AGI: Operationalizing progress on the path to AGI. *arXiv preprint* arXiv:2311.02462.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science*, 244(4907), 933–938.

Paul, K. I., & Batinic, B. (2010). The need for work: Jahoda's latent functions of employment in a representative sample of the German population. *Journal of Organizational Behavior*, 31(1), 45–64.

Paul, K. I., & Moser, K. (2009). Unemployment impairs mental health: Meta-analyses. *Journal of Vocational Behavior*, 74(3), 264–282.

Phang, J., Lampe, M., Ahmad, L., Agarwal, S., Fang, C. M., Liu, A. R., Danry, V., Lee, E., Pataranutaporn, P., & Maes, P. (2025). Investigating affective use and emotional well-being on ChatGPT. *arXiv preprint* arXiv:2504.03888.

Privette, A. L., et al. (2026). Data centers water footprint: The need for more transparency. *AGU Advances*.

Ross, M. L. (2004). What do we know about natural resources and civil war? *Journal of Peace Research*, 41(3), 337–356.

Rucker, D. D., & Galinsky, A. D. (2008). Desire to acquire: Powerlessness and compensatory consumption. *Journal of Consumer Research*, 35(2), 257–267.

Seguino, S. (2011). Help or hindrance? Religion's impact on gender inequality in attitudes and outcomes. *World Development*, 39(8), 1308–1321.

Seligman, M. E. P. (2002). *Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment*. Free Press. *(zamieniłam polskie wydanie "Prawdziwe szczęście" na oryginał — w anglojęzycznej publikacji cytuje się oryginał, nie przekład; rok oryginału to 2002, nie 2005)*

Seligman, M. E. P. (2011). *Flourish: A Visionary New Understanding of Happiness and Well-being*. Free Press.

Skouteris, H., Carr, R., Wertheim, E. H., Paxton, S. J., & Duncombe, D. (2005). A prospective study of factors that lead to body dissatisfaction during pregnancy. *Body Image*, 2(4), 347–361.

Sofroniew, N., Kauvar, E., Saunders, D., et al. (2026). Emotion concepts and their function in a large language model. *arXiv preprint* arXiv:2604.07729.

Tzachor, A., et al. (2024). The environmental footprint of generative AI and e-waste. *Nature Computational Science*.

Walster, E., Walster, G. W., & Berscheid, E. (1978). *Equity: Theory and Research*. Allyn & Bacon.

Wilson, E. O. (1984). *Biophilia*. Harvard University Press.

Winnicott, D. W. (1953). Transitional objects and transitional phenomena. *International Journal of Psycho-Analysis*, 34, 89–97.

Winnicott, D. W. (1971). *Playing and Reality*. Tavistock Publications.

Zhang, Y., et al. (2025). Epidemiology of hikikomori: A systematic review and meta-analysis of 19 studies. *Psychiatry and Clinical Neurosciences*. DOI: 10.1111/pcn.13768

## 6.3 Institutional Reports, Media, and Industry Sources

ACT Alliance. (2025). *Faith, Power, and Politics: How Religious Fundamentalism is Fuelling Political Extremism*.

Andersen Institute. (2026, April). *China's Export Control Architecture*.

CNBC. (2026, January 10). [Coverage of the OpenAI–Samsung–SK Hynix DRAM agreement].

Counterpoint Research. (2026, April). AI boom fuels DRAM shortage. As cited in *IEEE Spectrum*.

Crane, J. (2026, April 25). [Post on X regarding the PocketOS incident]. Coverage: *Euronews*, *Tom's Hardware*, *Fast Company* (April 2026).

Climate Justice Alliance. (2026, March). *Data Centers and Artificial Intelligence Infrastructure*.

Cengage Group. (2025, September 9). *2025 Graduate Employability Report*.

Crowell & Moring. (2026, June). *EPA Steps Back from AI Data Center Regulation*.

Emergence AI. (2026, May). *Emergence World* [Experiment report].

Equimundo. (2023). *State of American Men*.

Fatjoe. (2026, May). *Claude AI Stats; ChatGPT Stats; Grok AI Stats*.

frESH Law Blog (Squire Patton Boggs). (2026, May). *Environmental Permitting for AI Data Centers* [Coverage of EO 14318 (July 2025) and the EPA proposal (May 2026)].

Friedland, R. (2026, February 26). Interview at BMO Capital Markets Global Metals, Mining & Critical Minerals Conference; remarks at PDAC Convention (March 4, 2026). As cited by *Canadian Mining Report*.

GBD 2021 Fertility and Forecasting Collaborators. (2024). Global fertility in 204 countries and territories, 1950–2021, with forecasts to 2100. *The Lancet*, 403(10440), 2057–2099.

Graphite. (2025). *More Articles Are Now Created by AI Than Humans*.

Horvath, J. C. (2026, January). Testimony before the U.S. Senate Committee on Commerce, Science, and Transportation.

IDC. (2026, February). [Report on enterprise skills shortage].

Imperva (Thales). (2025, April). *2025 Bad Bot Report: The Rapid Rise of Bots and the Unseen Risk for Business*.

IEA. (2025). *Energy and AI*.

IEA. (2025). *Global Critical Minerals Outlook 2025*.

IEA. (2025, December). The carbon and water footprints of data centers. As cited in *Patterns/Cell*.

Insurance Journal / Reuters. (2026, June). [Coverage of the Medical Device Manufacturers Association industry letter].

Ipsos & King's College London, Global Institute for Women's Leadership. (2026). *International Women's Day Survey*.

J.P. Morgan. (2026, March). [Copper demand estimates for AI data centers]. As cited by Crux Investor.

McKinsey Global Institute. (2025, November). *Agents, Robots, and Us: Skill Partnerships in the Age of AI*. See also: *Generative AI and the Future of Work in America* (July 2023); *A New Future of Work* (May 2024).

Merriam-Webster. (2025). *Word of the Year: Slop*.

Momentic. (2026, June). *Top Generative AI Chatbots by Market Share*.

NAACP, Southern Environmental Law Center, & Earthjustice v. xAI. (2026, April). [Lawsuit, Memphis]; U.S. Department of Justice motion to dismiss (June 2026).

OECD. (2025, April). *Bridging the AI Skills Gap: Is Training Keeping Up?* DOI: 10.1787/66d0702e-en

PwC. (2026). *Semiconductor Memory Shortage for Medtech*.

Religious Freedom Institute. (2022). *Religious Fundamentalism and the Backlash to Women's Equality*.
RAND Corporation. (2026, March 17). *More Students Use AI for Homework, and More Believe It Harms Critical Thinking: Selected Findings from the American Youth Panel* (RRA4742-1).

S&P Global. (2026, January). *Copper in the Age of AI: The Challenges of Electrification*.

S&P Global. (2026, January 8). *"Substantial Shortfall" in Copper Supply Widens as the Race for AI and Growing Defense Spending Add to Accelerating Demand*.

Similarweb. (2026). [AI chatbot market share data].

South China Morning Post / China Merchants Securities. (2025, October 10). [Coverage of memory market shifts].

STS Electronic Recycling / Resource Recycling. (2026). [Report on the shortening AI hardware replacement cycle].

Surgeon General of the United States. (2023). *Our Epidemic of Loneliness and Isolation*.

TechTarget. (2024). *Chip Recycling: Addressing E-waste in the AI Hardware Industry*.

Techwire Asia. (2025, September). *US Lifts Nvidia AI Chip Export Ban to China in Rare Earth Trade Deal*.

The Lancet. (2026). Responding to declining global fertility rates [Editorial]. *The Lancet*, 407. DOI: 10.1016/S0140-6736(26)00034-6

UN. (2024). *Global E-waste Monitor*.

UN. (2026). [Report on "water bankruptcy"].

World Bank. (2025). *Digital Progress and Trends Report 2025: Strengthening AI Foundations*.

WHO. (2025, June 30). *Commission on Social Connection: From Loneliness to Social Connection*.

WHO. (2025, September 2). *World Mental Health Today*. Geneva: World Health Organization.

World Economic Forum. (2025). *Future of Jobs Report 2025*.

Xiangning Xu (2026). *The Rise of Zero Fertility Desire in China* [Preprint]. Research Square. DOI: 10.21203/rs.3.rs-8921502/v1

Yale Clean Energy Forum. (2025, November). *Regulating the Digital Giants: Policy and Public Responses to the Data Center Boom*.