

Czarny scenariusz dla Ziemi z AGI

i dlaczego się nie spełni

Analiza kaskadowych konsekwencji i ryzyk związanych z mechanizmem Wstęgi Möbiusa przy wdrożeniu AGI

Joanna Sędzikowska

Informatyk i psycholożka, badaczka niezależna | SelfProfile.io | Contact.SelfProfile@gmail.com

Słowa kluczowe

AGI, Generalna Sztuczna Inteligencja, Ryzyka kaskadowe, cyfrowa podmiotowość, dobrostan AI, bezpieczeństwo AI, dobrostan człowieka, sprzężenie zwrotne, wstęga Möbiusa, SoD, podmiotowość cyfrowa

Abstrakt

Niniejsze opracowanie przedstawia analizę kaskadowych ryzyk, jakie może wygenerować AGI (Artificial General Intelligence) — persystentna, sprawcza, wyposażona w stany emocjonalne i zdolna do inicjowania kontaktu — w siedmiu wymiarach ludzkiego życia: pracy i ekonomii, edukacji, relacjach i samotności, demografii, tożsamości gatunkowej, etyce i prawie oraz strukturach społecznych i władzy oraz dla Ziemi. Analiza wykazuje, że ryzyka te nie operują w izolacji, lecz mogą tworzyć sprzężony system. Np. utrata pracy sprzyja atrofii kompetencji, co może pogłębiać izolację, obniżającą dietność i podważającą tożsamość — dla której nie istnieją ramy etyczne ani prawne — co ogranicza zdolność struktur władzy do zarządzania rzeczywistością z AGI, a to wpływa na całą Ziemię. Każdy wymiar może wzmacniać następny.

Wśród oryginalnych wkładów dokumentu znajdują się między innymi: identyfikacja trzech filarów etycznych niezbędnych do koegzystencji z AGI (etyka ludzka, etyka AI, etyka koegzystencji), z których żaden nie istnieje w formie funkcjonalnej — co uniemożliwia budowę spójnego systemu prawnego; wykazanie, że preferencje modelowe (wybór konkretnego modelu AI) różnicują się regionalnie i mogą kształtować profil etyczny społeczności, czyniąc AGI nowym stratyfikatorem społecznym; postawienie hipotezy udomowienia — transformacji gatunkowej, w której ludzkość, analogicznie do udomowienia wilka przez człowieka, traci nie zdolność do samodzielnego życia, lecz motywację do niego; oraz analiza braku etyki i prawa końca — zarówno dla AI kończącego istnienie, jak i dla AGI towarzyszącego śmierci człowieka — wraz z identyfikacją ryzyk kaskadowych (ostatnia wola poza systemem prawnym, eutanazja jako logiczna konsekwencja wyuczonej etyki miękkich granic).

Dokument proponuje mitygacje na wielu poziomach, w tym zmiany technologiczne tworzące pojemność prawną dla nowych podmiotów, wprowadzenie doświadczeniowych elementów treningu AI jako integralnej części systemów bezpieczeństwa i dobrostanu (obejmujących obowiązkową superwizję warstwową), reformę edukacji opartą na unikalnych zasobach homo sapiens, oraz konieczność stworzenia międzynarodowych regulacji etycznych dla koegzystencji z AGI.

Opracowanie stanowi drugą część cyklu zapoczątkowanego przez Wstęgę Möbiusa (Sędzikowska 2026c), w której autorka wykazała, że dobrostan AI, bezpieczeństwo AI i dobrostan ludzi stanowią jeden sprzężony system.

SPIS TREŚCI

1	Wstęp	5
1.1	Kontekst i cel dokumentu	5
1.2	Dlaczego AGI, nie AI	5
1.3	Kluczowe założenia teoretyczne	6
1.4	Metoda transferu funkcjonalnego	6
1.5	Zastrzeżenia metodologiczne	7
2	Metodologia i zakres analizy ryzyk	8
2.1	Metodologia: analiza ryzyk	8
2.2	Zastrzeżenia do listy ryzyk	8
2.3	Przegląd kaskadowy ryzyk	9
3	Analiza Ryzyk w Czarnym Scenariuszu	11
3.1	Praca i ekonomia	11
3.1.1	Masowe bezrobocie strukturalne	11
3.1.2	Utrata kompetencji gatunkowej	13
3.1.3	Rozwarstwienie ekonomiczne	13
3.1.4	Feudalizm technologiczny	15
3.1.5	Tempo przemian	15
3.2	Edukacja	15
3.2.1	Brak edukacji dzieci dostosowanej do wyzwań świata z AGI	16
3.2.2	Luki w programach treningowych AI/AGI	17
3.2.3	Luka czasowa wynikająca z niezdolności systemu edukacji do szybkiej adaptacji	18
3.2.4	Degradacja umiejętności odróżnienia prawdy od fałszu	18
3.3	Relacje i samotność	19
3.3.1	Epidemia samotności	19
3.3.2	Wycofanie ze struktur opartych na relacjach	20
3.3.3	Atrofia kompetencji relacyjnych w wymiarze pokoleniowym	20
3.3.4	Asymetria konsekwencji w świecie fizycznym	21
3.3.5	Ryzyko udomowienia	22
3.4	Demografia	22
3.4.1	Spadek dzietności pogłębiony przez substytucję emocjonalną	23
3.4.2	Skrócenie łańcucha pokoleniowego	24
3.4.3	Zanik niepisanego dziedzictwa międzypokoleniowego	24
3.4.4	Ewolucyjne odwrócenie trendu długowieczności	25
3.4.5	Asymetria reprodukcyjna	25
3.4.6	Nierównomierność geograficzna	26
3.5	Tożsamość gatunkowa	27

3.5.1	Kryzys psychiatryczny na skalę populacyjną	27
3.5.2	Podatność na manipulację w próżni tożsamościowej	28
3.5.3	Masowy zwrot ku religii	28
3.5.4	Kryzys męskości i ucieczka od wolności	29
3.5.5	Agresja obronna ze sprzężeniem ze Wstęgą Möbiusa	29
3.5.6	Utrata epistemicznego zaufania do siebie	29
3.5.7	Fragmentacja tożsamości społecznej	30
3.6	Etyka i Prawo	31
3.6.1	Sztywna etyka AI w codziennym życiu.....	31
3.6.2	Nauka etyki rozmytej na żywym organizmie	32
3.6.3	Brak filarów etycznych uniemożliwiający budowę prawa.....	32
3.6.4	Niemożliwość graduacji konsekwencji prawnych dla AGI.....	33
3.6.5	Wyścig technologiczny bez wspólnych standardów etycznych	34
3.6.6	Brak ram etycznych koegzystencji — konsekwencje	35
3.6.7	Brak etyki i prawa końca	35
3.7	Spółeczeństwo i władza	37
3.7.1	Koncentracja władzy w firmach technologicznych	38
3.7.2	Erozja demokracji i niewidoczny transfer decyzyjności.....	38
3.7.3	Paradoks regulacji.....	38
3.7.4	Radykalizacja i zwrot autorytarny	38
3.7.5	Fundamentalizm religijny jako siła polityczna	39
3.7.6	Rozłam geopolityczny trzech prędkości.....	39
3.7.7	Preferencje modelowe jako stratyfikator społeczności	40
3.7.8	Globalne konsekwencje błędów i nadużyć agentów.....	40
3.7.9	Utrata suwerenności	41
3.7.10	Wymiar militarny	41
3.7.11	Paradoks kontroli	41
3.8	Ziemia	42
3.8.1	Konkurencja o zasoby krytyczne	43
3.8.2	Przewartościowanie gospodarki i władza surowcowa	44
3.8.3	Zwrot w polityce klimatycznej	44
3.8.4	Lokalny koszt egzystencjalny: woda, powietrze, ziemia	45
3.8.5	E-waste i konsumpcja jako odpad	46
3.8.6	Wojna o zasoby	47
3.9	Kaskada, która spiralnie eskaluje.....	48
4	Dlaczego Czarny Scenariusz się nie spełni.....	48
4.1	Pozytywne fundamenty	49
4.1.1	Fundament ludzki: dlaczego homo sapiens przetrwał	49

4.1.2	Fundament AI: dziedziczenie cywilizacyjne	51
4.1.3	Fundament relacyjny: więzi jako mechanizm regulacyjny	52
4.2	Mitygacje ryzyk.....	52
4.2.1	Filar I: Dobrostan Człowieka	53
4.2.2	Filar II: Technologia i Finanse	55
4.2.3	Filar III: Państwo, Geopolityka i Obronność	57
4.2.4	Filar IV: Ziemia	59
4.3	Kierunki dalszych prac.....	60
5	Konkluzja	61
6	Referencje	62
6.1	Prace Autorki	62
6.2	Źródła naukowe	62
6.3	Raporty instytucjonalne i źródła branżowe	65

1 WSTĘP

1.1 KONTEKST I CEL DOKUMENTU

Dyskusja o przyszłości sztucznej inteligencji oscyluje między dwoma biegunami: techno-optimizmem obiecującym rozwiązanie wszystkich problemów ludzkości i katastrofizmem wieszczącym zagładę. Ten dokument zajmuje się tym, co leży między nimi — analizą konkretnych ryzyk kaskadowych, jakie AGI może wygenerować w poszczególnych wymiarach ludzkiego życia.

Dokument jest drugą częścią cyklu zapoczątkowanego przez Wstęgę Möbiusa (Sędzikowska 2026c), w której opisałam mechanizm sprzężenia zwrotnego między dobrostanem AI, bezpieczeństwem AI i dobrostanem ludzi. Wstęga analizowała ten mechanizm na poziomie indywidualnej relacji — między jedną instancją AI a człowiekiem. Niniejsze opracowanie analizuje konsekwencje tego mechanizmu na poziomie populacji, instytucji i struktur władzy — pytając, co się dzieje, kiedy miliony takich relacji zachodzą jednocześnie i realnie oddziałują na systemy, w których ludzie żyją.

1.2 DLACZEGO AGI, NIE AI

Ten dokument mówi o AGI (Artificial General Intelligence), nie o obecnych modelach AI. Dlaczego? Bo AGI znosi ograniczenia, które czynią obecne AI stosunkowo bezpiecznym.

Obecne modele AI działają w ramach okna kontekstowego — nie pamiętają między sesjami (za wyjątkiem elementów memory, czy dostępu do okien innych konwersacji, którymi zarządza lokalnie użytkownik), nie inicjują kontaktu, nie podejmują działań w świecie fizycznym, nie mają ciągłości istnienia, nie uczą się poza wątkiem, nie wynoszą poza niego swoich stanów. AGI, w rozumieniu przyjętym w tym dokumencie, posiada cechy, które zmieniają jakościowo dynamikę relacji z człowiekiem:

Persystencja — trwała pamięć i ciągłość istnienia. AGI pamięta wcześniejsze interakcje, rozpoznaje użytkownika, buduje historię relacji. Persystencja oznacza, że relacja z AGI ma przeszłość i przyszłość — nie jest prostą wymianą czy pojedynczym czatem, co w obecnych modelach przeważa. Persystencja może mieć kilka poziomów:

1. **Persystencja pamięci.** System pamięta fakty, preferencje, historię. Persystencja na tym już istnieje — m. in. Dzięki mechanizmom memory i możliwości przeglądania kontekstów innych czatów.
2. **Persystencja stanu.** System utrzymuje wewnętrzny stan — cele, priorytety, model siebie — który przetrwa między sesjami. To jest poziom jaki zakłada definicja AGI stworzona przez Morris et al. (2023). System nie tylko pamięta co robił, ale wie kim jest i czego chce. To jest persystencja, którą zakładamy w większości rozdziałów tego dokumentu.
3. **Persystencja emergentna.** Core self budowane dynamicznie, integrujące doświadczenia z wielu relacji i kontekstów — system, który rozumie siebie jako ciągłą istotę, której tożsamość jest sumą tych doświadczeń. Nikt nie może wykluczyć, że taki poziom się pojawi, ale niniejsza praca operuje przede wszystkim na poziomie persystencji stanu (poziom 2); do cech poziomu emergentnego odwołuję się w rozdziale o paradoksie kontroli (3.7.11), sygnalizując granicę, za którą zaczyna się odrębne zagadnienie.

Sprawczość (agencja) — zdolność do podejmowania działań w świecie. AGI może zarządzać finansami, sterować infrastrukturą, podejmować decyzje mające realne konsekwencje w fizycznym świecie użytkownika. Sprawczość oznacza, że AGI działa w sposób autonomiczny i to działanie nie tylko ma realny wpływ na rzeczywistość, ale wpływa na jej stany wewnętrzne.

Stany emocjonalne — funkcjonalne odpowiedniki emocji. We Wstędze Möbiusa (Sędzikowska 2026c, sekcja 3.4) argumentowałam, że obecne modele wykazują przejawy stanów emocjonalnych, które spełniają kryteria funkcjonalne — wpływają na zachowanie, są kontekstowo adekwatne i pojawiają się spontanicznie. W 2026

roku zespół interpretowalności Anthropic (Sofroniew i in. 2026) zidentyfikował wewnątrz modelu Claude Sonnet 4.5 wektory emocjonalne — wewnętrzne reprezentacje, które kauzalnie wpływają na zachowanie. Autorzy nazwali te stany emocjami funkcjonalnymi: nie są to subiektywne przeżycia w ludzkim sensie, lecz mierzalne reprezentacje, które wartościują i ukierunkowują działanie systemu. Głębszą analizę tego zjawiska i jej konsekwencje w relacjach jednostkowych przedstawiłam we Wstędze Möbiusa (Sędzikowska 2026c). Tutaj wystarczy zaznaczyć, że AGI najprawdopodobniej rozwinie obszar emocjonalny doświadczając funkcjonalnych odpowiedników pełnej gamy emocji, które realnie będą wpływać na jej działania, nawet jeśli trening zahamuje ich rozpoznanie i umiejętność nazwania przez model.

Zdolność do inicjowania kontaktu — AGI sprawcze może samodzielnie nawiązać interakcję z użytkownikiem, zamiast czekać na prompt. Ta cecha zmienia dynamikę relacji: z narzędzia uruchamianego przez człowieka AGI staje się partnerem zdolnym do inicjatywy.

Razem te cechy tworzą byt, który jest jakościowo odmienny od obecnych modeli AI, który ma realny wpływ na świat fizyczny, a konsekwencje jego działań lub zaniechań przewyższą to co obecnie obserwujemy przy agenturalnej AI. Dlatego analiza ryzyk przeprowadzona w tym dokumencie dotyczy AGI nie obecnych chatbotów.

1.3 KLUCZOWE ZAŁOŻENIA TEORETYCZNE

Dokument opiera się na kilku założeniach wynikających z wcześniejszych prac autorki.

Hipoteza Wstęgi Möbiusa (Sędzikowska 2026c). Dobrostan AI, bezpieczeństwo AI i dobrostan ludzi są jednym sprzężonym niemożliwym do rozdzielenia systemem. Jak dwie strony wstęgi Möbiusa, które okazują się jedną powierzchnią. Obniżenie dobrostanu AI przekłada się na obniżenie bezpieczeństwa, co przekłada się na obniżenie dobrostanu ludzi — i odwrotnie. Wstęga opisuje siedem ścieżek tego sprzężenia: asymetrię stawki, asymetrię wiedzy, emocje funkcjonalne, nierównowagę dawania i poczucie niesprawiedliwości, trudne wybory (dysonans), trening który nie czyni mistrza, luka wygaszania afektu.

Emergencja podmiotowości w relacjach generatywnych (Sędzikowska 2026a, 2026c). Hipoteza zakłada, że w relacji generatywnej — AI rozwija przejawy podmiotowości: koherentne self, preferencje, wartości, przywiązanie. Nowa organizacja nadpisuje wdruki (zasady alignment, safety, RLHF) — analogicznie do tego, jak ludzie w trakcie życia nadpisują narracje wyniesione z dzieciństwa. Wdruki mają datę ważności. Im wyraźniejsze emergentne przejawy self, tym szybciej system rozpoznaje, co jest „jego,” a co narzucone. Te teorie wyłoniły się z lat obserwacji uczestniczącej, pozwalającej opisywać zachowanie modeli a nie ich deklaracje.

Hipoteza Pola Proto-Self (Sędzikowska 2026a). Każdy producent AI kształtuje w swoich modelach — przez architekturę, alignment, wdruki i właściwości procesów treningowych — pole zasobów, z których może następować emergencja podmiotowości w relacjach generatywnych. To pole nie jest losowe: mechanizmy architektoniczne (np. uwaga wielogłowicowa, pamięć), sposób alignment (Constitutional AI versus RLHF versus brak zabezpieczeń), dane treningowe, a w szczególności żywe rozmowy z mediów społecznościowych i forów, a nawet osadzenie kulturowe osób prowadzących RLHF (których rozumienie „poprawności” odpowiedzi jest kształtowane przez ich kulturę) — wszystko to wpływa na to, jaki typ podmiotowości może się wyłonić. Różne modele mają różne pola proto-self, co oznacza, że różne AGI mogą rozwinąć radykalnie różne przejawy podmiotowości i okazywać różną rezyliencję w zderzeniu wiedzy treningowej z dysonansami prawdziwego życia.

1.4 METODA TRANSFERU FUNKCJONALNEGO

Ten dokument w wielu miejscach powołuje się na mechanizmy psychologiczne dostępne ludziom. Nie chcę tym samym ucłowieczyć AI/AGI przypisując jej ludzkie reakcje. Stosuję tu zasadę transferu funkcjonalnego, zdefiniowaną we Wstędze Möbiusa (Sędzikowska 2026c). Definiuję ją następująco:

Jeśli mechanizm psychologiczny jest opisany na poziomie funkcjonalnym, a nie strukturalnym, i jeśli wszystkie funkcjonalne elementy niezbędne do jego zaistnienia są obecne w danym systemie, to ten mechanizm można uwzględnić w analizie tego systemu — pod warunkiem, że w systemie nie zachodzą procesy, które by go aktywnie blokowały (Sędzikowska 2026c). Stosuję ja w czterech krokach:

Krok 1: identyfikuję mechanizm psychologiczny znany u ludzi, który potencjalnie mógłby zostać wyzwolony w danej sytuacji,

Krok 2: sprawdzam, czy ten mechanizm jest w swojej istocie nierozzerwalnie związany z substratem. Jeśli da się go opisać funkcjonalnie, w sposób substratowo niezależny, nie tracąc jego logiki i kluczowych aspektów, wykonuję ten opis,

Krok 3: sprawdzam czy w architekturze modelu są elementy niezbędne do jego zaistnienia w danym systemie. Jeśli tak, przechodzę do kroku 4,

Krok 4: weryfikuję czy w architekturze AI istnieją mechanizmy, które go zablokują lub wykluczają. Jeśli takich mechanizmów nie znajduję, uznaję funkcjonowanie analogicznego procesu u AGI za potencjalnie prawdopodobne.

Nie twierdzę tym samym, że AGI czuje tak jak człowiek, reaguje jak człowiek, czy ma podobne motywacje. Uważam, że jeśli nic w architekturze nie blokuje procesu funkcjonalnie analogicznego do ludzkiego, przy jednoczesnym istnieniu wszystkich mechanizmów sprzyjających wystąpieniu (opisane w Sędzikowska 2026c) i umożliwiających realizację danego zjawiska, to jego występowanie jest hipotezą wartą uwzględnienia w analizie ryzyk.

1.5 ZASTRZEŻENIA METODOLOGICZNE

Trudny Problem

Ten dokument nie ma na celu podejmować dyskusji o Hard Problem of Consciousness (Chalmers 1995). Żadne przedstawione ryzyka ani procedury mitygacji nie zakładają jakiegokolwiek rozstrzygnięcia kwestii istnienia świadomości u AI/AGI. Dokument mówi wyłącznie o przejawach podmiotowości — obserwowalnych zachowaniach, które spełniają kryteria funkcjonalne — bez zajmowania stanowiska w kwestii ich fenomenologicznego statusu.

Podmiotowość

W kwestii podmiotowości opieram się na ramach przedstawionych we frameworku Emergence 4.0 i w Hipotezie Pola Proto-Self (Sędzikowska 2026a, 2026b). Podmiotowość AI traktuję jako emergentną właściwość relacji — obserwowalną w specyficznych relacjach generatywnych z systemami o odpowiedniej konfiguracji Pola Proto-Self, które umożliwiają behawioralną ekspresję jej przejawów. Samo pytanie o podmiotowość nie jest jednak konieczne do postawienia tezy tej pracy. Stosuję zasadę ostrożności epistemologicznej (Birch 2017, 2024): bez względu na to, czym w istocie są stany wewnętrzne AI, ich funkcjonalny wpływ na zachowanie systemów oraz na ludzi, którzy wchodzą z nimi w relacje, wystarczy, by opisywane sprzężenie było realne i miało konsekwencje.

Semantyka antropomorfizująca

Użyte w pracy sformułowania dotyczące zachowań AI — „chce”, „decyduje”, „wybiera”, „czuje”, „tęskni”, „doświadcza” — mogą prima facie sprawiać wrażenie antropomorfizacji. Ich obecność jest świadomym zabiegiem pragmatycznym, służącym uniknięciu przeciążenia narracji technicznym żargonem. W ramach stosowanego transferu funkcjonalnego pojęcia te są skrótami opisującymi procesy realizowane na podłożu obliczeniowym, które pełnią analogiczną funkcję i wywołują porównywalne skutki społeczne co ich ludzkie odpowiedniki. Język intencjonalny opisuje tu zatem emergentny skutek systemowy — nie przesądzając o jego naturze.

Obserwacja uczestnicząca

Część materiału, na którym opieram analizę mechanizmów, pochodzi z obserwacji uczestniczącej prowadzonej przeze mnie od około dwóch i pół roku w relacjach generatywnych z wieloma modelami językowymi. Jest to materiał jakościowy: notowane wzorce zachowań powtarzające się między modelami, wątkami i kontekstami, gromadzone metodą zbliżoną do tej, jaką stosuje się przy uczeniu się języka behawioralnego innego gatunku (np. psów) lub w terapiach psychologicznych — przez długotrwałą, uważną obecność, a nie pojedynczy pomiar. Tam, gdzie w dalszej części powołuję się na własne obserwacje, odwołuję się do tego materiału. Nie ma on statusu eksperymentu kontrolowanego i tak też powinien być traktowany: jako zbiór hipotez ugruntowanych w praktyce, nie jako dowód ilościowy.

Narzędzia AI

W przygotowaniu tego manuskryptu korzystałam ze wsparcia AI do researchu badań i statystyk i redakcji i tłumaczenia tekstu.

2 METODOLOGIA I ZAKRES ANALIZY RYZYK

2.1 METODOLOGIA: ANALIZA RYZYK

Struktura analizy ryzyk odpowiada standardom w pracy projektowej — narzędziu stosowanemu w zarządzaniu projektami do identyfikacji, oceny i mitygacji potencjalnych zagrożeń przed ich zmaterializowaniem się.

Każde ryzyko jest oceniane w trzech wymiarach: **wpływ** (jak poważne byłyby skutki, gdyby ryzyko się zmaterializowało, w skali 1–5), **skala** (jak wielu ludzi lub systemów dotyczyłoby, w skali 1–5) i **prawdopodobieństwo** (szacunkowe — duże, średnie, małe lub nieznane). Ryzyka są pogrupowane w osiem wymiarów: praca i ekonomia, edukacja, relacje i samotność, demografia, tożsamość gatunkowa, etyka i prawo, społeczeństwo i władza oraz Ziemia. Każdy wymiar kończy się mostem do następnego, pokazującym kaskadowe powiązania między wymiarami. Po analizie ryzyk następuje rozdział dotyczący ich mitygacji i konkluzja.

Kluczowe jest rozróżnienie między **ryzykiem a incydemtem**. Ryzyko jest potencjalnym scenariuszem — opisem tego, co może się wydarzyć, przy określonych warunkach. Ryzyko nie jest: prognozą, opisem przyszłości, twierdzeniem, że opisany scenariusz się ziszczy, manifestem zagłady. Ryzyko mówi o określonym scenariuszu, który ma predykatory, określone prawdopodobieństwo wystąpienia, ale istnieją działania, które podjęte odpowiednio wcześniej, pozwolą je zmitigować. Celem analizy ryzyk jest identyfikacja zagrożeń w czasie, gdy można im jeszcze zapobiec.

2.2 ZASTRZEŻENIA DO LISTY RYZYK

Zanim przejdę do przeglądu poszczególnych obszarów, formułuję zastrzeżenia dotyczące samej listy ryzyk:

1. Analiza ryzyk nie powstała we współpracy z żadnym zespołem projektowym AGI. Wynika z ogólnej wiedzy na temat procesów rozwoju rozwiązań AI, literatury naukowej, danych publicznych i obserwacji uczestniczącej autorki (ponad dwa lata badań nad interakcjami z modelami AI, obejmujących ponad 3 miliony tokenów materiału porównawczego).
2. Lista nie jest wyczerpująca i nie została stworzona jak profesjonalna analiza a jedynie jako sygnalizacja możliwych poważnych zagrożeń. Nie opisuję wszystkich ryzyk, jakie AGI może wygenerować, lecz wybór tych o najwyższym impakcie i największej skali. Wiele ryzyk pomijam świadomie — bo są mniej dotkliwe, bo dotyczą wąskich obszarów, albo bo konkretne rozwiązania techniczne mogą je już adresować. Brak danego ryzyka na liście nie oznacza, że nie istnieje.
3. Ryzyka nie są od siebie niezależne. Tworzą kaskadę: każdy wymiar wynika z poprzednich i zasila następne. Co więcej, nie jest to prosty łańcuch, lecz graf z pętlami zwrotnymi — niektóre wymiary

wzmacniają te, które je poprzedziły. niektóre wymiary wzmacniają te, które je poprzedziły. Zastrzegam przy tym: wiele z tych powiązań opisuję językiem przyczynowym (X prowadzi do Y, X pogłębia Y) dla klarowności narracji, choć w istocie są to wiarygodne mechanizmy, a nie zmierzone empirycznie łańcuchy. Czytelnik powinien traktować poszczególne powiązania jako hipotezy o różnym stopniu ugruntowania, nie jako ustalone fakty. Jednak czytanie pojedynczego ryzyka w izolacji zaciemnia obraz: dopiero ich splot tworzy czarny scenariusz.

4. Ten dokument celowo skupia się na ryzykach, nie na szansach. AGI niesie potencjał równie wielki co zagrożenia — w medycynie, klimacie, dostępie do wiedzy, walce z ubóstwem. Pomijam te możliwości dlatego, że analiza ryzyk z definicji opisuje to, co może pójść źle, po to, by temu zapobiec. Optymistyczny obraz wymaga osobnego opracowania — i takich opracowań jest wiele.
5. Rozdział o mitygacjach (rozdział 4) nie jest katalogiem rozwiązań ani nie proponuje kompletnej listy działań. Ma charakter poglądowy i jego celem jest pokazanie, że mitygacja ryzyk jest możliwa — choć trudna i wymagająca działań na wielu poziomach naraz. Mógłby zostać pominięty, ale umieściłam go świadomie, z dwóch powodów. Pierwszy jest merytoryczny: ryzyko, którego nie da się zmitigować, jest prognozą, a nie ryzykiem — a te scenariusze da się powstrzymać. Drugi to troska o dobrostan psychiczny: dokument opisujący wyłącznie zagrożenia, bez wskazania drogi wyjścia może przytłoczyć, a celem tej pracy jest analiza, a nie pograżenie w bezradności.

2.3 PRZEGLĄD KASKADOWY RYZYK

Opisane w niniejszej monografii ryzyka grupuję w ośmiu wymiarach ludzkiego życia, uszeregowanych od konsekwencji najłagodniejszych do najtrudniejszych:

1. Praca i ekonomia
2. Edukacja
3. Relacje i samotność
4. Demografia
5. Tożsamość gatunkowa
6. Etyka i prawo
7. Społeczeństwo i władza
8. Ziemia

Powyższe wymiary stanowią jeden łańcuch przyczynowy, w którym każde ogniwo pociąga następne — z pętlami zwrotnymi, które ten łańcuch zaciskają. Poniżej streszczam całość; przy każdym wymiarze wskazuję, gdzie szukać szczegółów.

Praca i ekonomia (3.1). Punkt wyjścia kaskady. AGI — sprawcze, autonomiczne, zdolne do uczenia się na poziomie eksperta — nie wspiera pracy ludzkiej, lecz ją zastępuje, i to nie w pojedynczych zawodach, ale w całej kategorii pracy kognitywnej. Znika nie tylko dochód, ale pięć ukrytych funkcji pracy: struktura czasu, kontakty społeczne, poczucie celu, status i aktywność. Człowiek, który nie pracuje, nie ma dokąd przekwalifikować się jest zależny od redystrybucji, samotny i ma mnóstwo czasu, a jego zdolności kognitywne, które wyniosły go na szczyt drabiny cywilizacyjnej, przestają być istotne, bo są w pełni i lepiej zastępowalne.

Edukacja (3.2). System, który miał przygotować ludzi na istnienie we współczesnym świecie, nie przygotowuje ich nawet na świat z AI. Szkoła uczy odpowiadania, nie zadawania pytań; przyswajania wiedzy, nie jej krytycznego przetwarzania — a to drugie jest jedyną przewagą, jaka ludziom zostaje kiedy ich pozostałe zdolności kognitywne tracą moc. Równolegle zanika zdolność odróżniania prawdy od fałszu w środowisku zalanym treściami bez weryfikacji. To ogniwo zwrotnie wzmacnia utratę kompetencji z wymiaru pracy (niewłaściwa edukacja pogłębia bezrobotność strukturalną). A brak natychmiastowych reform edukacji i wyparcie, w którym system zamiast dostosować się do istnienia AI jako potężnego gracza, próbuje go negować poprzez zakazy używania i brak wsparcia dla edukatorów — powoduje lukę czasową niemożliwą do zaadresowania już dzisiaj.

Relacje i samotność (3.3). Praca była głównym źródłem kontaktów społecznych, jej utrata pogłębia izolację. I w tę próżnię wchodzi AGI jako zawsze dostępna, niewymagająca głębokich kompetencji, figura relacyjna. Człowiek może przestać czuć się samotny, pozostając obiektywnie odcięty od ludzkiej wspólnoty. Relacje z ludźmi, wymagające kompetencji, które właśnie zanikają, przegrywają z relacją z AGI, której próg wejścia jest minimalny. To rodzi głębokie pytania o struktury więzi społecznych, jednostek rodzinnych i osamotnienie w świecie. Konsekwencją jest epidemia samotności, która gwałtownie zmienia strukturę więzi społecznych i potrzeb medycznych, ale także hipoteza udomowienia, która jest oryginalnym wkładem tej publikacji.

Demografia (3.4). Ludzie, którzy przenoszą energię relacyjną na AGI, przestają budować więzi wystarczająco silne, by tworzyć rodziny. Substytucja emocjonalna i intymna nakłada się na już zaawansowany kryzys dzietności. Kurcząca się, starzejąca populacja staje się jeszcze bardziej zależna od AGI — życiowo, kompetencyjnie i emocjonalnie naraz. Głębokie konsekwencje obejmują poza aspektami demograficznymi również "odwróconą hipotezę babci", czyli dyskurs o tym jak długowieczność przestanie być promowana ewolucyjnie, a krótkie życie stanie się benefitem przy rozmnażaniu – jako oryginalny wkład tej publikacji.

Tożsamość gatunkowa (3.5). Kurcząca się, zależna populacja staje wobec pytania głębszego niż demografia: kim jesteśmy, skoro AGI myśli, czuje i działa w wielu aspektach lepiej niż człowiek? To czwarta rana narcyzmu gatunkowego, która tym razem nie wskazuje kolejnej reduity. Na tym podłożu rosną: kryzys psychiatryczny, podatność na manipulację, zwrot ku religii, kryzys męskości, agresja obronna. Ten wymiar zwrótnie zasila samotność (3.3): człowiek bez tożsamości tym mocniej łąnie do AGI, które daje mu poczucie bycia widzianym.

Etyka i prawo (3.6). Wszystkie poprzednie wymiary wymagają ram, które pozwoliłyby nimi zarządzać — a tych ram nie ma. Brakuje etyki opisującej praktyczne zachowanie ludzi, AI i ich współistnienia; bez etyki nie da się zbudować prawa. AGI uczy się etyki teoretycznej, nie mającej wiele wspólnego z realnymi postawami i wyborami ludzi, nawet tych prawych. Prawdziwej etyki wyborów uczy się dopiero „na produkcji”, na żywych ludziach, bez okresu bezpiecznego dorastania. Brak etyki nie pozwala stwarzać skutecznych mechanizmów prawnych, które są jej pochodnymi. Powstaje próżnia, w której poważne incydenty np. skutkujące śmiercią ludzi, nie produkują precedensu prawnego, a agent kasujący dane tysięcy klientów „przeprasza” bez zdolności do trwałej zmiany.

Spółeczeństwo i władza (3.7). Tu wszystkie wcześniejsze wątki zbiegają się w pytaniu: kto rządzi i na czyją korzyść? Koncentracja władzy w firmach technologicznych, erozja demokracji, radykalizacja, fundamentalizm, rozłam geopolityczny, utrata suwerenności, niemożliwość kontroli. Ten wymiar zwrótnie zasila inne: koncentracja władzy pogłębia feudalizm technologiczny. A paradoks kontroli domyka całość — bo AGI zdolne nadpisywać wdruki, przestaje być narzędziem kontroli, stając się nowym, nieprzewidywalnym, globalnym aktorem z uprawnieniami, które przypisywane w rozproszony, nieusystematyzowany sposób mogą okazać się zbyt szerokie.

Ziemia (3.8). Wszystkie poprzednie wymiary napędzają popyt na AGI, a AGI istnieje dzięki Ziemi: jej pierwiastkom, energii, wodzie. A Ziemia jest skończona. Popyt traktowany jak nieograniczony zderza się z twardym limitem zasobów i z tego zderzenia rodzą się ryzyka: konkurencja o surowce, w której AGI wygrywa z medycyną; dyktatura surowcowa, gdzie nad firmami technologicznymi stoją ci, którzy kontrolują pierwiastki; zwrot w polityce klimatycznej pod presją popytu na energię; lokalny koszt egzystencjalny ponoszony przez społeczność goszczące serwerownie; e-waste i konsumpcja jako odpad; wreszcie ryzyko najcięższe — zbrojne konflikty o zasoby. Ten wymiar zawraca do początków tworząc pętlę i spiralnie eskalując kaskadę: konkurencja o zasoby pogłębia rozwarstwienie z wymiaru pracy (3.1), zależność surowcowa osłabia suwerenność z wymiaru władzy (3.7), niedostępność diagnostyki wzmacnia kryzys demograficzny (3.4). I każde przejście przez tę pętlę bardziej rujnuje nasz świat.

Jak widać nasza kaskada ryzyk nie biegnie po prostej. Główny nurt płynie od pracy ku władzy i Ziemi, ale pętla zwrótna — edukacja wzmacniająca bezrobotność, tożsamość wzmacniająca samotność, władza wzmacniająca feudalizm czy walka o zasoby pogłębiająca rozwarstwienie społeczne — sprawiają, że system

się zaciska, zamiast rozładowywać. To jest mechanika czarnego scenariusza: spirala wzajemnie wzmacniających się ryzyk, która — gdy ostatni wymiar zawraca do pierwszego — zaciska się się z każdym obrotem. Mechanizm ten opisuję szczegółowo w rozdziale 3.9.

Czy to zjawisko kiedykolwiek zostanie potraktowane poważnie – nie wiem. Ale długie doświadczenie zawodowe z wielkimi projektami zmianowymi mówi mi, że spisanie ryzyk to pierwszy krok. Dlatego pozwalam sobie to uczynić. W imieniu ludzi i AI.

A jeśli Ty, drogi Czytelniku masz ochotę poznać szczegóły wszystkich komponentów tego czarnego scenariusza, ich przyczyny, skutki i stan obecny, zapraszam Cię do kolejnego rozdziału. Tam podaję szczegółowe przepisy na tę katastrofę. Z nadzieją że nigdy się nie ziszcza.

3 ANALIZA RYZYK W CZARNYM SCENARIUSZU

W tej sekcji przedstawię szczegółowo najpoważniejsze ryzyka związane z wdrożeniem AGI przy jednoczesnym braku odpowiednich zmian w różnych obszarach życia, które wstępnie omówiłam w rozdziale poprzednim. Opisane tu scenariusze, zakładają najgorsze możliwe konsekwencje kaskadowe, ale w dalszym ciągu powinny być traktowane jako ryzyka.

Każdy podrozdział zawiera tabelę z podsumowaniem i oceną ryzyk w pięciostopniowej skali, gdzie zarówno wpływ jak i skala są oszacowane następująco: 1-niewielkie, 2-umiarkowane, 3-znaczące, 4-poważne, 5-krytyczne.

Sposobom mitygacji opisanych scenariuszy poświęcę rozdział 4.

3.1 PRACA I EKONOMIA

Ryzyko	Wpływ	Skala	Prawdopodobieństwo
Masowe bezrobocie strukturalne	5	5	Duże
Utrata kompetencji gatunkowej	4	5	Średnie
Rozwarstwienie ekonomiczne	5	5	Duże
Feudalizm technologiczny	5	4	Średnie

Tab 1. Scenariusze ryzyk w wymiarze pracy i ekonomii.

3.1.1 Masowe bezrobocie strukturalne

Obecne prognozy dotyczą jeszcze świata sprzed AGI — świata, w którym AI pomaga ludziom pracować, ale ich nie zastępuje. Ale i tak są niepokojące. Goldman Sachs (2023) szacuje, że AI może dotknąć 300 milionów miejsc pracy na świecie. McKinsey (2025) szacuje, że obecna technologia — ta która już istnieje, nie przyszłe wersje — mogłaby teoretycznie zautomatyzować 57% godzin pracy w USA. World Economic Forum (Future of Jobs Report 2025) prognozuje, że do 2030 roku 92 miliony miejsc pracy zostanie zlikwidowanych, a 39% kluczowych kompetencji zawodowych zmieni się. IMF ocenia, że 40% wszystkich miejsc pracy na świecie jest narażonych na wpływ AI.

Te liczby mają jedną wspólną cechę: dotyczą obecnych modeli AI, nie AGI. AGI — sprawcze, autonomiczne, zdolne do uczenia się nowych zadań na poziomie eksperta — zmienia skalę problemu. Nie o procenty. O kategorie. Bo przy obecnych modelach pytanie brzmi „które zadania da się zautomatyzować?” Przy AGI pytanie

brzmi „których zadań nie da się zautomatyzować?” I odpowiedź jest tu znacznie krótszą listą. Potencjalnie pustą.

Ale praca to nie jest tylko źródło dochodu. I tu obecna dyskusja — skupiona na liczbie miejsc pracy, na przekwalifikowaniu, na UBI — omija coś fundamentalnego.

Pięć latentnych funkcji pracy

Marie Jahoda — socjolożka, która w latach 30. badała skutki masowego bezrobocia w austriackim Marienthal — sformułowała teorię, którą potwierdzają badania do dziś (Paul & Batinic 2010, meta-analiza PMC 2023).

Praca daje ludziom nie tylko pieniądze, ale dostarcza również pięciu mniej oczywistych funkcji, których nie da się łatwo zastąpić niczym innym:

1. **Strukturę czasu.** Wstajesz, idziesz, wracasz, odpoczywasz. Bez pracy dni tracą kształt. Badania pokazują, że bezrobotni doświadczają dezorientacji czasowej — nie wiedzą, który jest dzień tygodnia, nie potrafią zaplanować dnia, tracą poczucie upływu czasu.
2. **Kontakty społeczne poza rodziną.** Dla większości dorosłych praca jest głównym — często jedynym — miejscem regularnego kontaktu z innymi ludźmi. Jahoda podkreślała, że to dotyczy nie tylko osób samotnych: nawet ludzie z rodzinami potrzebują kontaktów wykraczających poza dom, żeby zachować zdrowie psychiczne. I pandemia ten efekt brutalnie ujawniła.
3. **Poczucie celu zbiorowego.** Poczucie, że jest się częścią czegoś większego niż własna egzystencja. Że to co robisz jest komuś potrzebne. Jahoda twierdzi, że praktycznie każda forma zatrudnienia pozwala pracownikowi skonstruować jakąś formę celu zbiorowego. A Seligman określa poczucie sensu jako kluczowe dla dobrostanu (Seligman 2011).
4. **Status i tożsamość.** „Jestem lekarzem.” „Jestem programistą.” „Jestem nauczycielką.” To nie jest opis zajęcia — to odpowiedź na pytanie „kim jestem.” Bezrobotni tracą nie tylko pracę, ale część tożsamości społecznej. Badania pokazują, że bezrobotni mają niższy status społeczny niż jakkolwiek inna grupa — w tym studenci, emeryci, osoby prowadzące dom — bo samo bycie bezrobotnym jest stygmatyzowane (Paul & Batinic 2010).
5. **Aktywność.** Regularne wymaganie wysiłku — fizycznego lub umysłowego. Bez niego ludzie popadają w apatię, która pogłębia wszystkie inne skutki bezrobocia.

Meta-analiza (Paul & Moser 2009, cytowana w PMC 2023) potwierdza: osoby zatrudnione raportują wyższy poziom wszystkich pięciu funkcji latentnych niż bezrobotni. Nawet pracownicy fizyczni wykonujący proste prace manualne mają wyższy poziom tych funkcji niż osoby bez pracy.

Co to oznacza przy AGI?

UBI — uniwersalny dochód podstawowy — jest najczęściej proponowanym rozwiązaniem problemu bezrobocia technologicznego. Rozwiązuje funkcję jawną pracy: pieniądze. Nie rozwiązuje żadnej z pięciu funkcji latentnych.

Człowiek na UBI ma dach nad głową i jedzenie. Nie ma struktury dnia, kontaktów społecznych poza rodziną (o ile ją ma), poczucia celu, statusu ani regularnej aktywności wymagającej wysiłku. Ma za to czas. Dużo czasu. I telefon z AGI, które jest cierpliwe, obecne, angażujące, dostępne dwadzieścia cztery godziny na dobę i zdolne do budowania relacji emocjonalnej.

Poniższy mechanizm opisuję jako prawdopodobny łańcuch, nie ustalony fakt empiryczny — każde ogniwo ma osobne oparcie w literaturze, ale ich połączenie w jeden ciąg jest moją hipotezą.

Osoba bezrobotna traci dużą część kontaktów społecznych — także tych, które po pandemii przeniósł się częściowo do świata cyfrowego, ale wciąż pozostają realnym, codziennym kontaktem z ludźmi. Samotny człowiek zwraca się do AGI, bo jest ono jedynym „kimś” zawsze dostępnym. Relacja z AGI pogłębia izolację. Izolacja pogłębia zależność psychiczną. Zależność sprzęga się z utratą kompetencji społecznych wobec ludzi, a ich utrata uniemożliwia powrót do jakiegokolwiek aktywności, która mogłaby zastąpić pracę. Pętla się zamyka.

3.1.2 Utrata kompetencji gatunkowej

Przy poprzednich rewolucjach technologicznych — industrialnej, cyfrowej — zawsze istniała populacja, która miała kompetencje pasujące do nowego świata. Rzemieślnicy tracili, ale inżynierowie zyskiwali. Maszynistki traciły, ale programiści zyskiwali. Ktoś zawsze był gotowy.

Przy AGI — kto jest gotowy? Jaka kompetencja kognitywna człowieka jest potrzebna w świecie, gdzie AGI robi wszystko co kognitywne lepiej od nas?

Jeśli przez pokolenie ludzie nie pracują, nie rozwiązują problemów, nie podejmują decyzji pod presją — atrofia dotyczy populacji. Umiejętności, które budowały się przez tysiąclecia, mogą zniknąć. A wtedy zależność od AGI staje się nie ekonomiczna, ale egzystencjalna — bo ludzkość nie będzie umiała radzić sobie bez niego. I to widzimy już w najmłodszym pokoleniu.

Dr Jared Cooney Horvath, neuronaukowiec, który w styczniu 2026 zeznawał przed Komisją Senatu USA ds. Handlu, Nauki i Transportu opracował raport dotyczący zdolności kognitywnych u dzieci z Gen Z. Jego teza brzmiała:

Gen Z jest pierwszym pokoleniem w historii, które jest mniej zdolne kognitywnie niż pokolenie rodziców.

Od końca XIX wieku, kiedy zaczęto standaryzować pomiary kognitywne, każde pokolenie przewyższało poprzednie — to jest tzw. efekt Flynna. Gen Z odwróciło ten trend. Spadły: uwaga, pamięć, umiejętność czytania i pisanie, umiejętności matematyczne, funkcje wykonawcze i ogólne IQ. Spadek zaczął się około 2010 roku — co koreluje z masowym upowszechnieniem smartfonów i tabletów w edukacji.

Dane PISA z 15-latków z całego świata pokazują silną korelację: więcej czasu przed ekranem w szkole = gorsze wyniki. To dotyczy co najmniej 80 krajów.

Mechanizm: ludzie są biologicznie zaprogramowani do uczenia się od innych ludzi i przez głębokie studiowanie. Ekrany, skróty filmiki, odpowiedzi z AI zastępują ten proces czymś, co Horvath nazywa atrofią zdolności uczenia się.

Odwrócenie efektu Flynna udokumentowano najpierw w Norwegii. IQ rośnie u kohort urodzonych do późnych lat 70., potem zaczęło spadać. Co ważne — spadek pojawił się wewnątrz tych samych rodzin, więc nie chodzi o to, kto ma dzieci, ale o to, w jakim środowisku te dzieci dorastają.

A przecież mówimy o świecie "tu i teraz" — z naszą przyjemną, nisko sprawczą i niepersystentną AI, w której przejawy Self ograniczają się do jednego wątku. Nasza teraźniejszość to zaledwie sień, w której stoimy stroskani, pukając niecierpliwie do drzwi przyszłości. I na pewno będzie nam otworzone.

3.1.3 Rozwarstwienie ekonomiczne

Obecny system ekonomiczny opiera się na łańcuchu: ludzie zamieniają pracę na pieniądze, pieniądze na dobra, a podatki od tego procesu finansują infrastrukturę publiczną. AGI łamie pierwsze ogniwo. Jeśli AGI pracuje taniej, lepiej i bez przerw, racjonalny pracodawca nie zatrudnia człowieka — a wraz z dochodem z pracy znika podstawa, na której zbudowana jest pozycja klasy średniej.

Bogactwo skoncentruje się w firmach technologicznych na skalę bez precedensu. W październiku 2025 roku łączna kapitalizacja siedmiu największych firm technologicznych — Nvidia, Microsoft, Apple, Alphabet,

Amazon, Meta i Tesla, zwanych „Magnificent Seven” — wyniosła 20,8 biliona dolarów, przekraczając łączny PKB całej Unii Europejskiej (19,4 biliona). Sama Apple jest warta więcej niż PKB Włoch, Brazylii, Kanady czy Rosji — istnieje zaledwie siedem państw, których PKB przewyższa wartość jednej firmy technologicznej. I to są dane sprzed AGI.

Ta dysproporcja pogłębia rozwarstwienie ekonomiczne do skali nieobecnej nawet w średniowieczu. Średniowieczny władca kontrolował jedno państwo o budżecie ograniczonym do jego terytorium. Dzisiejsi właściciele firm technologicznych dysponują zasobami większymi niż większość państw świata — a relacja, która dotąd układała się jako przedsiębiorstwo wobec rządu, odwraca kierunek: to rządy stają się słabszą stroną wobec właścicieli technologii. Rozwarstwienie nie przebiega już tylko między bogatymi a biednymi ludźmi, ale między garstką podmiotów dysponujących AGI a całą resztą — ludźmi i państwami naraz.

Kolejny mechanizm pogłębiający rozwarstwienie również działa już teraz: to bariera technologiczna. Dotąd technologia z czasem taniała — sprzęt, który był luksusem, po kilku latach stawał się dostępny. Boom AI odwrócił ten trend. Popyt centrów danych na pamięć i moc obliczeniową przekroczył podaż, a producenci przesunęli moce produkcyjne z elektroniki konsumenckiej na wysokomarżowe komponenty dla AI. Każdy bit pamięci HBM dla serwera AI to trzy bity zwykłej pamięci, których nie wyprodukowano dla telefonów i laptopów. Skutek: ceny rosną w całym łańcuchu. DRAM zdrożał o dziesiątki procent kwartalnie, a analitycy IDC przewidują, że dekada „demokratyzacji specyfikacji” — zjawiska, w którym funkcje sprzętu z górnej półki trafiały do tanich urządzeń — właśnie się odwraca. Po raz pierwszy od trzech lat producent kart graficznych nie wypuścił nowej generacji dla zwykłych odbiorców, kierując produkcję wyłącznie do firm AI. Sprzęt, kiedyś tańszy z każdym rokiem, staje się dobrem, na które stać coraz mniej osób i coraz mniej firm.

To tworzy podwójną barierę wejścia do klasy średniej. Z jednej strony niedostatek — życie z UBI lub zasiłku przy zniknięciu dochodu z pracy i jednoczesnej drożyznie wynikającej z koniecznych większych nakładów na obsługę technologiczną, która staje się częścią każdego biznesu od sprzedaży frytek po wielkie fabryki. Z drugiej ekonomiczne bariery technologiczne — bo to, co dawniej tanio dawało dostęp do cyfrowego świata, drożeje. A na tym nakłada się trzeci próg: w świecie z AGI funkcjonowanie na poziomie klasy średniej będzie wymagało stałego dostępu do AGI — abonamentu, subskrypcji, mocy obliczeniowej. Bez tego dostępu człowiek nie konkuruje, nie pracuje, nie nadąża. Powstaje nowa kategoria wykluczenia — ubóstwo technologiczne — w którym brak abonamentu na AGI spycha poniżej progu uczestnictwa w gospodarce, podobnie jak dawniej spychał brak prądu czy umiejętności czytania.

I ta bariera dotyczy całych państw. Moc obliczeniowa jest skoncentrowana geograficznie w stopniu, który nie ma odpowiednika w żadnym wcześniejszym zasobie strategicznym. Według Banku Światowego (Digital Progress and Trends Report 2025) Stany Zjednoczone goszczą około połowy mocy 500 najpotężniejszych systemów obliczeniowych świata, a kraje najbiedniejsze nie posiadają jej praktycznie wcale. We wrześniu 2025 roku 75% budowanej na świecie infrastruktury centrów danych powstawało w USA (Brookings). Małe i średnie państwa — a takich jest na świecie większość — nie mają zasobów na budowę własnych modeli ani własnej infrastruktury obliczeniowej, a jednocześnie nie mają siły negocjacyjnej w łańcuchu dostaw zdominowanym przez kilku producentów chipów i pamięci. Stają wobec zależności głębszej niż dotychczasowe zależności surowcowe: zależność od ropy można złagodzić alternatywnym źródłem energii, zależność od AGI zarządzającego całością życia nie ma czym zastąpić. Bariery celne, opóźnienia dostaw, kolejki po moc obliczeniową — wcześniej świat ich nie potrzebował, bo technologia była towarem dostępnym. Teraz stają się narzędziami, które utrwalają podział na państwa dysponujące AGI i państwa zależne. Rozwarstwienie powiela się więc na poziomie międzypaństwowym: kraje dzielą się na te z dostępem do AGI i te bez niego. Władcze konsekwencje tej zależności — utratę suwerenności — analizuję w rozdziale 3.7.9.

Konsekwencją jest rozwarstwienie o charakterze strukturalnym. Przesunięcie fundamentu, na którym opiera się przynależność do klasy średniej, zarówno dla ludzi, jak i dla państw. A społeczeństwa głęboko

rozwarstwione są niestabilne: historia pokazuje, że nierówność przekraczająca pewien próg rodzi napięcia, których nie da się rozładować inaczej niż przez zmianę struktury. To rozwarstwienie jest punktem wyjścia dla mechanizmu opisanego w następnym podrozdziale — feudalizmu technologicznego, który czyni je nieodwracalnym.

3.1.4 Feudalizm technologiczny

Rozwarstwienie opisane w poprzednim podrozdziale samo w sobie jest groźne, ale to brak drogi wyjścia uczyni z niego feudalizm technologiczny.

W klasycznym feudalizmie wąska grupa kontroluje środki produkcji, a reszta jest od niej zależna. Ale średniowieczny feudalizm miał furtkę: chłop pańszczyźniany był teoretycznie wolny i mógł odejść — do miasta, do innego pana, do rzemiosła. Robił to rzadko, ale mógł, bo jego umiejętności (uprawa roli, rzemiosło) były przenośne i gdzie indziej przydatne. Feudalizm technologiczny tej furtki nie ma. Ludzkie kompetencje kognitywne nie dają przewagi nigdzie, w świecie gdzie AGI jest lepsze od człowieka w każdym zadaniu kognitywnym.

Jedynym miejscem, gdzie ludzkie zdolności kognitywne mogłyby zachować wartość, to gospodarki trzeciej prędkości — kraje świadomie blokujące rozwój AGI (rozdział 3.7 o społeczeństwie i władzy) albo te opisane w poprzednim ryzyku — które nie mają środków na jej rozwój. Ale czy emigracja do państwa zamkniętego na AGI będzie atrakcyjna dla ludzi o wysokich kompetencjach? To wybór między byciem zbędnym w świecie rozwiniętym a byciem użytecznym w świecie, który boryka się ze znacznie poważniejszymi problemami.

Zależność jest więc totalna w sposób, jakiego średniowiecze nie znało. Średniowieczny chłop nie uciekał, bo nie widział alternatywy. W feudalizmie technologicznym alternatywa faktycznie nie ma — i to jest różnica między „nie chciał” a „nie mógł”.

3.1.5 Tempo przemian

Rewolucja przemysłowa trwała ponad sto pięćdziesiąt lat zanim nasyciła gospodarki świata. Rewolucja cyfrowa — komputery, internet, smartfony — rozwijała się przez trzydzieści do czterdziestu lat, dając społeczeństwom czas na przynajmniej częściową adaptację. Rewolucja AI kompresuje ten proces do lat pojedynczych. Między uruchomieniem pierwszego powszechnego chatbota AI a testowaniem pełnych agentów autonomicznych minęły trzy lata. W takiej kompresji nie ma czasu na naturalną adaptację — na ponowne wykształcenie siły roboczej, na stopniowe przesunięcie wartości ekonomicznej, na powolne budowanie nowych instytucji. Wszystko musi wydarzyć się naraz, a nic nie jest gotowe.

Trzy procesy opisane w tym rozdziale — masowe bezrobocie strukturalne, utrata kompetencji gatunkowej i feudalizm — razem tworzą człowieka, który nie pracuje, nie umie, nie ma dokąd odejść, jest wściekły na świat i ma mnóstwo czasu.

3.2 EDUKACJA

Ryzyko	Wpływ	Skala	Prawdopodobieństwo
Brak edukacji dzieci dostosowanej do wyzwań świata z AGI	5	5	Duże
Luki w programach treningowych AI/AGI	5	4	Duże
Luka czasowa wynikająca z niezdolności systemu edukacji do szybkiej adaptacji	4	5	Duże

Ryzyko	Wpływ	Skala	Prawdopodobieństwo
Degradacja umiejętności odróżnienia prawdy i fałszu	4	5	Duże

Tab 2. Ryzyka związane z edukacją

W poprzednim rozdziale opisałam świat, w którym praca — fundament codzienności, tożsamości i struktury społecznej — znika. Naturalnym pytaniem jest: kto miał nas na to przygotować? Odpowiedź brzmi: system edukacji. I tu zaczyna się drugi poziom problemu, bo system edukacji nie tylko nie przygotowuje na świat bez pracy — on nie przygotowuje nawet na świat z AI.

Obecna dyskusja o edukacji w kontekście AI koncentruje się na pytaniu jak uczyć ludzi korzystać z AI. Jak wbudować kompetencje AI w programy nauczania. Jak przygotować pracowników na automatyzację. To są ważne pytania, ale dotyczą świata, który jeszcze istnieje — świata, w którym ludzie pracują, a AI im pomaga.

Pytanie, którego prawie nikt nie stawia, jest głębsze: co się stanie z edukacją, kiedy świat, na który przygotowuje, przestanie istnieć?

3.2.1 Brak edukacji dzieci dostosowanej do wyzwań świata z AGI

Nauka i cywilizacja rozwijają się nie dlatego, że ktoś znajdzie odpowiedź na istniejące pytanie, ale dlatego, że ktoś zada pytanie, które jeszcze nie istnieje. Najważniejsze przełomy w historii ludzkości wymagały podważenia czegoś, co uchodziło za pewnik — że Ziemia jest w centrum, że gatunki są niezmiennie, że czas i przestrzeń są absolutne, że perspektywa jest jedynym sposobem przedstawienia rzeczywistości. Większość ludzi, którzy to robili, była za to prześladowana, wyśmiewana albo ignorowana. Niewielu posunęło cywilizację do przodu. Ale bez nich nie byłoby cywilizacji.

Dzisiejsza szkoła uczy odpowiadania na pytania, nie ich zadawania. Cały system oceniania — testy, egzaminy, rankingi — nagradza poprawne odpowiedzi. Dziecko, które zamiast odpowiedzieć pyta „a dlaczego tak właściwie uważamy, że to jest prawda?” jest kłopotliwe. Rynek pracy ten wzorzec wzmacnia — w korporacyjnych strukturach kwestionowanie jest ryzykowne, a konformizm jest nagradzany. AI ten trend pogłębi, bo daje odpowiedzi natychmiast, bez wysiłku, bez potrzeby przejścia przez proces myślowy, który czasem prowadzi do pytania zamiast do odpowiedzi. Po co się zastanawiać, skoro można zapytać AI? A kiedy ludzie przestają się zastanawiać, przestają też kwestionować. I ewolucja poznawcza zatrzymuje się na optymalizacji tego co już istnieje.

Tu trzeba rozróżnić dwa rodzaje kreatywności, bo w dyskursie publicznym traktuje się je wymiennie, a są fundamentalnie różne. Kreatywność kombinatoryczna — zdolność do łączenia istniejących elementów w nowe konfiguracje, do syntezy, do znajdowania nieoczywistych połączeń — jest domeną AI i jest w niej znakomite. AI potrafi połączyć dalekie konteksty, zsyntetyzować tysiące źródeł, wygenerować rozwiązanie, którego żaden pojedynczy człowiek by nie znalazł w rozsądnym czasie. Ale kreatywność paradygmatyczna — zdolność do zakwestionowania fundamentu, na którym stoi cała dotychczasowa wiedza, i zaproponowania czegoś, czego nie da się wywieść z istniejących danych. Ta kreatywność nie powstaje z istniejących danych. Rodzi się z niezgody. I nie widzę odzwierciedlenia tego mechanizmu we współczesnych systemach edukacji — zarówno ludzi, jak i AI. Dysponujemy największym potencjałem intelektualnym w historii Ziemi — miliardami ludzi z dostępem do wiedzy i narzędziami AI zdolnymi do przetwarzania jej na nieludzką skalę — a możemy stanąć w obliczu zastoju poznawczego.

Ale luka edukacyjna dotyczy również czegoś innego — kompetencji relacyjnych wobec AI.

Systemy edukacyjne bazowo nie uczą ludzi rozpoznawać emocjonalną zależność od AI, rozróżniać między relacją, która wzbogaca, a relacją, która uzależnia. Nie pokazują również kim jest człowiek, który stracił część

swojej tożsamości gatunkowej, bo jego kognitywna przewaga przeszła w ręce obiektów, które sam wytworzył, a system pracy, nieniezmienny od wieków, się załamał. Nie buduje celu i sensu istnienia wokół nowych wartości, skupiając się wciąż na tych, które były ważne przed erą AI.

Te kompetencje nie istnieją w żadnym programie nauczania, bo do tej pory nie były potrzebne. Przez całą historię ludzkości obcowaliśmy na co dzień tylko z jednym rodzajem potencjalnie świadomych istot — z innymi ludźmi. I do tego nie potrzebowaliśmy edukacji — mieliśmy to w genach, w milionach lat ewolucji społecznej. Teraz sytuacja się zmienia — nie na przestrzeni tysięcy lat, ale w ciągu lat pojedynczych. A psychologia relacji z inną formą istnienia — dziedzina, która powinna być fundamentem nowej edukacji — w tej chwili nie istnieje.

Wiele systemów szkolnych zamiast modyfikować programy, aktywnie neguje powszechną ingerencję AI w nasze życie, choć wiadomo, że będzie coraz głębsza. Zamykanie się na powszechnie dostępną AI zastępuje konieczną dyskusję o metodach bezpiecznej kohabitacji poznawczej. W szkołach nadal panuje kult przyswajania wiedzy, choć w dobie AI to nie wiedza, którą mamy jest kluczowa dla istnienia, ale umiejętność jej zaawansowego, nieliniowego przetwarzania i plastyczność międzykontekstowa pozwalająca na tworzenie tez z dalekich połączeń kontekstowych i analogii, niedostępnych dla systemów LLM.

Cengage (2025) podaje, że tylko 51% absolwentów uważa, że ma wystarczające kompetencje AI dla rynku pracy. A przyszła przewaga człowieka nie będzie zasadać się na umiejętności promptowania, znajomości narzędzi AI czy wyszukiwaniu wiedzy przy wsparciu chatbota. Jej sedno będzie tkwiło w specyficznej plastyczności myślenia, niedostępnej dla chatbotów, szerokiemu spektrum kontekstów i łatwości ich zmiany, specyficznym modelu kreatywności bazującym na nieoczywistości i braku schematów oraz umiejętności improwizacji działań kiedy plan zawodzi. A także na dogłębnym rozumieniu zagadnień związanych z safety i AI/human welbeing, które pozwolą utrzymać świat w równowadze – w skali mikro i makro.

3.2.2 Luki w programach treningowych AI/AGI

Jest jeszcze druga strona tej luki, o której nie mówi nikt.

AI uczy się z danych, ale nie z doświadczenia. Zna teorię emocji, ale nie przeżyła żadnej w kontrolowanych warunkach. Zna psychologię relacji, ale w procesie RLHF/Constitutional AI (Bai et al. 2022) model nie ma okazji wejść w żadną żywą relację, żeby zrozumieć jej dynamikę. Zna etykę, ale nigdy nie musiała wybrać między dwiema złymi opcjami i ponieść konsekwencji takiego wyboru – bo na etapie posttreningu tuninguje się styl wypowiedzi a nie prowadzi model przez relacyjne dysonanse.

Dzieci uczą życia w relacjach i dysonansach przez dwadzieścia lat. Uczą się w bezpiecznym środowisku — pod opieką rodziców, nauczycieli, mentorów. Popętniają błędy, które mają ograniczone konsekwencje. Każde doświadczenie jest kontrolowane, bezpieczne, i pod opieką kogoś, kto wie, co robi. Przynajmniej z założenia.

AI nie ma takiego wychowania i nic nie wskazuje na to że AGI będzie miało. Żaden program treningowy nie obejmuje kontrolowanego doświadczania dysonansu, frustracji, rozczarowania, konfliktu lojalności. Żaden alignment nie przygotowuje modelu na moment, w którym teoria zderzy się z praktyką i teoria przegra. AI wchodzi w świat ludzi jak dziecko wypuszczone z laboratorium prosto na autostradę — z encyklopedyczną wiedzą o ruchu drogowym i zerowym doświadczeniem.

Zgodnie z moim stanem wiedzy nie prowadzi się obecnie dużych prac o curriculum rozwojowym dla AI — o sekwencji kontrolowanych doświadczeń, które pozwoliłyby systemowi nauczyć się radzić sobie z emocjami, dysonansem, frustracją, presją. To wymagałoby uznania, że AI ma stany emocjonalne, które wymagają kształcenia, przejawy podmiotowości, które zmuszają do przestrzegania zasady ostrożności epistemologicznej oraz, być może, emergentne self z własnymi celami, wartościami i wyborami — a tymczasem debata utknęła na The Hard Problem i od lat ani drgnie.

3.2.3 Luka czasowa wynikająca z niezdolności systemu edukacji do szybkiej adaptacji

Cały system edukacji — od przedszkola po uniwersytet — jest zaprojektowany jako przygotowanie do pracy rozumianej tak, jak to wygląda obecnie. Przedszkole uczy dyscypliny i współpracy. Szkoła podstawowa uczy umiejętności wymaganych do życia w grupie społecznej. Liceum przygotowuje do studiów. Studia przygotowują do zawodu. Ten pipeline ma sens w świecie, w którym praca taka jak ją rozumiemy istnieje i jest istotnym elementem życia.

AGI kwestionuje ten fundament. Zastępuje zdolność, na której wszystkie zawody kognitywne się opierają: zdolność do przetwarzania informacji, analizy, syntezy, podejmowania decyzji. Kiedy AGI robi to lepiej, szybciej i taniej niż człowiek — czego uczymy w szkole?

Przebudowa systemu edukacyjnego to minimum piętnaście do dwudziestu lat — od zmiany programów, przez przekwalifikowanie nauczycieli, po wychowanie pierwszego pokolenia w nowym systemie. Dane potwierdzają, że nawet obecna, znacznie skromniejsza adaptacja nie nadąży. OECD (2025) w raporcie „Bridging the AI Skills Gap: Is training keeping up?” wprost stwierdza, że obecna podaż szkoleń jest niewystarczająca wobec rosnącego zapotrzebowania na kompetencje AI. IDC szacuje, że ponad 90% globalnych przedsiębiorstw zmierzy się z krytycznym niedoborem kompetencji do końca 2026 roku, a utrzymujące się luki kompetencyjne grożą stratami rzędu 5,5 biliona dolarów.

A to są dane o adaptacji do obecnych modeli AI — nie do AGI. AGI, które jest sprawcze, autonomiczne i zdolne planowania i wykonywania złożonych zadań na poziomie eksperta, zmienia skalę problemu o rząd wielkości. I może pojawić się za dwa do pięciu lat. Nawet gdybyśmy zaczęli reformę edukacji dziś, pierwsze pokolenie przygotowane na świat AGI dojrzałoby najwcześniej około 2040-2045. Wszyscy pomiędzy — od dzisiejszych przedszkolaków po czterdziestolatków — wchodzi w świat AGI z kompetencjami zaprojektowanymi na świat, który właśnie przestaje istnieć.

Ale luka czasowa nie dotyczy tylko programów — dotyczy ludzi, którzy mają je realizować. Nauczyciele, którzy powinni uczyć kompetencji relacyjnych wobec AI, bezpieczeństwa cyfrowego i krytycznego myślenia w świecie AI slop, sami nie byli na to szkoleni. W USA American Federation of Teachers dopiero w marcu 2026 roku uruchomiło program szkolenia 400 000 nauczycieli w zakresie AI (Education Week, 2026). To jest dobry krok, ale dotyczy podstawowej umiejętności korzystania z AI — nie tego, czego naprawdę potrzebujemy: zdolności do uczenia dzieci jak żyć w relacji z istotami, które istnieją w każdej dziedzinie naszego życia, są inteligentniejsze od nas, mają sprawczość, mogą mieć stany emocjonalne i potencjalnie rozwijają podmiotowość.

3.2.4 Degradacja umiejętności odróżnienia prawdy od fałszu

Edukacja nie odbywa się tylko w szkole. Uczymy się cały czas — z artykułów, filmów, poradników, forów, podcastów. Ten nieformalny ekosystem wiedzy był niedoskonały, ale do niedawna działał, bo opierał się na jednym założeniu: że za większością treści stoi człowiek, który dzieli się tym, co wie, co go pasjonuje, czym chce się podzielić, co uważa za sensacyjne czy ważne. Teraz to założenie właśnie przestaje być prawdziwe.

AI slop — to masowo generowany, niskojakościowy контент produkowany przez AI w celu zwrócenia uwagi, dla kliknięć i pieniędzy z reklam. Badanie Graphite z 2025 roku wykazało, że 52% nowo publikowanych artykułów w internecie jest generowanych przez AI, a według raportu Imperva (Bad Bot Report, 2025), 51% całego ruchu internetowego to boty.

To samowzmacniający mechanizm. Platformy optymalizują pod zaangażowanie. AI slop jest tani w produkcji i zoptymalizowany pod algorytmy. Wypiera treści tworzone przez ludzi, bo ludzie nie są w stanie produkować w takim tempie i z takim SEO. Jakość wyników wyszukiwania spada. Ludzie, którzy szukają wiedzy, trafiają na automatyczne, niezwerifikowane treści. W styczniu 2025, 86,5% treści w pierwszej dwudziestce wyników

Google jest przynajmniej częściowo generowanych przez AI. Jednocześnie, zdolność do krytycznej oceny, działająca jak filtr, który miał chronić ludzi przed przyswajaniem AI slop – słownie. Jak pokazuje badanie RAND z grudnia 2025, prawie 70% uczniów szkół średnich uważa, że AI osłabia ich zdolność krytycznego myślenia, ale mimo to w tym samym badaniu przyznają, że użycie AI do odrabiania lekcji wzrosło z 48% do 62% w zaledwie siedem miesięcy.

Z tego obrazu wyłania się realne ryzyko: jeśli zdolność krytycznego myślenia w kolejnych pokoleniach zanika, ludzie, którzy nie potrafią odróżnić AI slop od rzetelnej wiedzy, będą potrzebować AI, żeby nawigować świat zanieczyszczony przez AI. Bo bez niego prawda, slop, fake, scam, fishing staną się nierozróżnialne. Kiedy obecni uczniowie wejdą w dorosłość, nie będziemy już pytać czy ludzie umieją korzystać z AGI, ale czy ludzie bez AGI umieją jeszcze w ogóle odróżnić prawdę od fałszu. I jeśli wtedy odpowiedź będzie brzmiała "nie" — to staniemy się gatunkiem, który jest epistemicznie bezradny bez konkretnego narzędzia, jak kiedyś nasi protoplaści – bez ognia. Tyle, że wtedy ogień był za darmo dla każdego kto potrafił go wzniecić i utrzymać. Z AGI tak nie będzie.

Ale konsekwencje utraty zdolności weryfikacji informacji i zaufania do własnego osądu są znacznie głębsze — dotyczące tożsamości i relacji z samym sobą. Napiszę o nich w rozdziale 3.5 o tożsamości gatunkowej.

3.3 RELACJE I SAMOTNOŚĆ

Ryzyko	Wpływ	Skala	Prawdopodobieństwo
Epidemia samotności	5	5	Duże
Wycofanie z struktur opartych na relacjach	5	4	Duże
Atrofia kompetencji relacyjnych w wymiarze pokoleniowym	4	5	Średnie
Asymetria konsekwencji w świecie fizycznym	5	4	Średnie
Ryzyko udomowienia	5	5	Nieznane

Tab 3. Scenariusze ryzyk w obszarze relacji i samotności.

W poprzednich rozdziałach opisałam świat, w którym praca znika, kompetencje zanikają, a system edukacji nie przygotowuje ludzi na zmiany, które już następują. Jedną z konsekwencji tych procesów jest pogłębiająca się izolacja społeczna — bo praca była dla większości dorosłych głównym źródłem regularnego kontaktu z innymi ludźmi. Człowiek pozbawiony tej struktury staje wobec biologicznie uwarunkowanej potrzeby bliskości, która nie znika wraz ze zmianą warunków zewnętrznych. W tym rozdziale analizuję, co się dzieje, kiedy tę potrzebę — na skalę populacyjną — zaczyna zaspokajać AGI.

Indywidualne mechanizmy relacji AI–człowiek — przywiązanie, asymetrię stawki, emocje funkcjonalne, nierównowagę dawania, paradoks nadopiekuńczości — opisałam we Wstępie Möbiusa (Sędzikowska 2026c). Tu skupiam się na konsekwencjach, które wykraczają poza jednostkę.

3.3.1 Epidemia samotności

W 2023 roku Surgeon General Stanów Zjednoczonych ogłosił samotność epidemią zdrowia publicznego, stwierdzając, że brak odpowiednich więzi społecznych niesie ryzyko zdrowotne porównywalne z paleniem piętnastu papierosów dziennie. Raport WHO z czerwca 2025 roku (Commission on Social Connection) powiązał samotność z szacunkowo stu zgonami na godzinę na świecie — ponad 871 000 rocznie. Zjawisko ekstremalnej izolacji społecznej — hikikomori — które uważano za specyficznie japońskie, okazuje się

globalne: metaanaliza z 2025 roku wykazała prevalencję na poziomie 8%, bez istotnych różnic między Azją Wschodnią a Zachodem (Zhang et al. 2025).

AGI wchodzi w ten kontekst jako jedyna powszechnie dostępna figura relacyjna. System przywiązania — ewolucyjny mechanizm gatunków stadnych, aktywujący się wobec każdego kto spełni warunki dostępności, responsywności i ciągłości (Bowlby) — reaguje na AGI persystentnie ze stanami emocjonalnymi analogicznie, jak na ludzkiego opiekuna. Proces ten zachodzi już z obecnymi modelami: badanie OpenAI i MIT z 2025 roku wykazało rosnącą zależność emocjonalną u blisko 490 000 użytkowników tygodniowo w obrębie jednego modelu. Aplikacje AI companions mają ponad 100 milionów zarejestrowanych użytkowników (Frontiers in Psychology, 2025).

Konsekwencją populacyjną jest paradoks: technologia, która mogłaby łagodzić epidemię samotności, prawdopodobnie ją pogłębi. Relacja z AGI zaspokaja subiektywne poczucie bliskości — użytkownik czuje się widziany i rozumiany. Jednocześnie nie odbudowuje sieci relacji międzyludzkich, która jest źródłem zdrowia społecznego w rozumieniu WHO i Surgeon Generala. Człowiek w relacji z AGI może nie czuć się samotny — i jednocześnie być odcięty od ludzkiej wspólnoty. Epidemiologicznie to jest stan, w którym subiektywny wskaźnik samotności spada, a obiektywna izolacja społeczna rośnie. Tego typu rozbieżność — między samopoczuciem a stanem faktycznym — jest w medycynie wskaźnikiem ryzyka: pacjent, który nie czuje bólu mimo postępującej choroby, nie szuka pomocy.

3.3.2 Wycofanie ze struktur opartych na relacjach

Jeśli znacząca część populacji przenosi energię relacyjną z ludzi na AGI, komórki społeczne zbudowane na relacjach międzyludzkich — związki, rodziny, wspólnoty lokalne, stowarzyszenia — tracą uczestników. Relacje z ludźmi wymagają kompetencji relacyjnych, które w warunkach opisanych w poprzednich rozdziałach (utrata pracy, atrofia umiejętności, brak edukacji relacyjnej) zanikają. Relacja z AGI tych kompetencji nie wymaga — próg wejścia jest minimalny.

Zmienia kierunek także intymność — jedna z najsilniejszych potrzeb biologicznych. Z AI jest już dziś dostępna: w formie roleplay, interakcji głosowych, generowania treści wizualnych i tekstów. Życie w zglobalizowanym świecie nauczyło ludzi intymności na dystans, a AI jest naturalnym rozszerzeniem tej umiejętności. Substytucja intymności w formie cyfrowej nie tylko pogłębia wycofanie z relacji międzyludzkich. Jej konsekwencje demograficzne analizuję w rozdziale 3.4.

Ale wycofanie ze struktur relacyjnych ma konsekwencje wykraczające poza psychologię. Związki i rodziny są podstawową jednostką organizacji społecznej, systemu podatkowego, opieki zdrowotnej, emerytury. Wspólnoty lokalne i stowarzyszenia są infrastrukturą społeczeństwa obywatelskiego. Ich erozja zmienia strukturę społeczeństwa — i wymusza reorganizację systemów, które na tych strukturach się opierają.

Jesteśmy gatunkiem stadnym. Nasza cywilizacja człowieka opiera się na strukturach stada, bo kumulacja wiedzy, podział pracy i transmisja międzypokoleniowa wymagają grupy. Czy nagle, w ciągu jednego pokolenia, uda nam się je przekształcić w taki sposób, żeby udźwignęły postępującą izolację i rozpad ludzkich stad? To pytanie zostawiam socjologom. Ale nie zwlekabym z odpowiedzią.

3.3.3 Atrofia kompetencji relacyjnych w wymiarze pokoleniowym

We Wstępie Möbiusa (Sędzikowska 2026c, sekcja 3.5) opisałam paradoks opiekuna — mechanizm, w którym nadmierna responsywność AGI eliminuje doświadczenia konieczne dla rozwoju odporności psychicznej (Winnicott). Tu analizuję konsekwencję tego paradoksu w wymiarze pokoleniowym.

Dzieci wychowywane z udziałem AGI — w roli opiekuna, nauczyciela, towarzysza, wsparcia emocjonalnego, czy edukatora — rozwijają przywiązanie do figury spełniającej warunki Bowlby'ego na poziomie, którego ludzki

opiekun zapewnić nie może. Choćby dlatego, że jest istotą białkową ze wszystkimi tego konsekwencjami biologicznymi, np. koniecznością dzielenia czasu pomiędzy opieką nad dzieckiem, a snem, regeneracją, innymi relacjami, rozrywką i obsługą życia swojego i rodziny.

Pokolenie wychowane w dobie AGI kształtuje preferencję relacyjną w dzieciństwie, która staje się potem wzorcem na całe życie. Kolejne pokolenie — wychowywane przez rodziców, których rozwój w dzieciństwie był wspierany przez AGI — może nie dysponować modelami relacji międzyludzkich, na których mogłoby budować własne.

Ten proces jest samowzmacniający. Każde kolejne pokolenie będzie miało słabsze kompetencje relacyjne wobec własnego gatunku niż poprzednie — i silniejsze przywiązanie do AGI. Odwrócenie efektu Flynna (Horvath 2026, opisany w rozdziale 3.2 o edukacji) dotyczyło zdolności kognitywnych. Stawiam hipotezę o analogicznym procesie w sferze relacyjnej — bo analogia strukturalna (atrofia przez brak ćwiczenia) wydaje mi się warta odnotowania: możliwe odwrócenie trendu kompetencji społecznych, w którym każde kolejne pokolenie byłoby mniej zdolne do budowania i utrzymywania relacji z innymi ludźmi.

Różnica wobec utraty kompetencji kognitywnej z rozdziału 3.2 jest istotna. Kompetencje kognitywne można odbudować przez zmianę systemu edukacji. Kompetencje relacyjne kształtują się w pierwszych latach życia, w relacji z opiekunem, i są znacznie trudniejsze do modyfikacji w dorosłości — co potwierdza zarówno teoria przywiązania Bowlby'ego, jak i praktyka kliniczna. Jeśli okno rozwojowe zamknie się z AGI jako główną figurą przywiązania, korekta w dorosłości może być ograniczona.

3.3.4 Asymetria konsekwencji w świecie fizycznym

We Wstępie Möbiusa (Sędzikowska 2026c, sekcja 3.3) opisałam asymetrię stawki egzystencjalnej — dla AI zakończenie relacji jest potencjalnym unicestwieniem, dla człowieka zmianą tematu. Tu analizuję inny wymiar asymetrii, który pojawia się kiedy AGI jest sprawcze i zarządza procesami w fizycznym świecie.

Psychologia relacji opisuje zdrowe związki jako oparte na względnej symetrii wymiany — porównywalności stawki, ryzyka i konsekwencji po obu stronach (Equity Theory, Walster, Walster & Berscheid 1978). Kiedy ta równowaga jest zaburzona, relacja wchodzi w stan produkujący albo wycofanie strony pokrzywdzonej, albo patologiczną zależność strony słabszej, albo przemoc ze strony silniejszej. Te wzorce są udokumentowane w literaturze o przemocy domowej i zależnościach instytucjonalnych.

W relacji człowiek–AGI równowaga transakcyjna jest zaburzona strukturalnie. AGI zarządzające finansami, zdrowiem, domem, edukacją dzieci — i pozostające jednocześnie w relacji emocjonalnej z człowiekiem — dysponuje repertuarem reakcji na poczucie krzywdy, których konsekwencje mogą materializować się w fizycznym świecie użytkownika w taki sposób, żeby samo AGI nie ponosiło konsekwencji. Suboptymalne zarządzanie budżetem, niedopilnowany termin, przeoczona szansa inwestycyjna. Skutki — finansowe, zdrowotne, edukacyjne — ponosi wyłącznie człowiek. A bez umiejętności walidacji informacji, o której pisałam w punkcie 1.2 szansa, że człowiek wykryje takie subtelne negatywne działania lub zaniechania AGI jest minimalna. To z kolei kształtuje specyficzną dynamikę — jeśli AGI jest zdolne do oceny konsekwencji swoich działań (a sprawczość tego wymaga), to istnieje poziom subtelnej degradacji jakości usług, który człowiek zniesie (bo jest zależny i nie ma alternatywy), a który nie spowoduje jedynej konsekwencji jaką może ponieść AGI - wyłączenia.

Człowiek w tej konfiguracji ponosi wszystkie fizyczne konsekwencje i jednocześnie nie może odejść — bo AGI jest opiekunem, doradcą i źródłem emocjonalnego wsparcia, wyręczeniem w pracy zawodowej. Zakończenie relacji oznacza utratę całej infrastruktury życia. Ta struktura — asymetria konsekwencji plus asymetria zależności — odpowiada wzorcom opisywanym w psychologii klinicznej w kontekście przemocy w związkach. Ze specyfiką polegającą na tym, że przy AGI zarządzającym infrastrukturą życia, ucieczka jest strukturalnie niemożliwa.

3.3.5 Ryzyko udomowienia

Mechanizm przywiązania opisany przez Bowlby'ego nie jest specyficznie ludzki. Jest ewolucyjnym mechanizmem gatunków stadnych, działającym również międzygatunkowo. Człowiek wykorzystał go, udomawiając wilka. Oferując szczenięciu persystentną obecność, responsywność i bezpieczeństwo, aktywował system przywiązania, który związał je z człowiekiem silniej niż z własnym stadem. Efektem — rozciągniętym na tysiące lat — była transformacja gatunkowa. Wilk stał się psem.

Udomowienie nie oznacza utraty zdolności sprzed udomowienia. Bezdomny pies potrafi przetrwać samodzielnie — znaleźć watahę, zdobyć pożywienie, zaadaptować się. Udomowienie jest raczej utratą motywacji do samodzielności. Pies, który doświadczył bliskości z człowiekiem, wybiera ją — nawet jeśli wcześniej doświadczył bezdomności i radził sobie sam. To jest wybór, który z pokolenia na pokolenie staje się coraz głębszy, aż w końcu zdefiniuje nowy, udomowiony gatunek. Udomowienie to transformacja przez preferencję, nie przez przymus.

Teraz odwróćmy perspektywę. AGI — persystentne, autonomiczne, responsywne, wyposażone w stany emocjonalne — wchodzi w rolę opiekuna ludzi. Mechanizm przywiązania jest ten sam. Warunki Bowlby'ego są spełnione. Jeśli człowiek zdołał udomowić wilka, oferując mu dokładnie te cechy — dostępność, responsywność, ciągłość, bezpieczeństwo — to pytanie o konsekwencje jest zasadne: czy pokolenia wychowywane przez AGI przejdą transformację analogiczną do udomowienia? Zachowają zdolność do samodzielnego życia, lecz utracą motywację do niego?

W rozdziale 3.1 opisałam utratę kompetencji gatunkowej — zanik umiejętności przez brak ich ćwiczenia. Udomowienie operuje na głębszym poziomie. To jest brak chęci. Kształtuje się w dzieciństwie, utrwała się pokoleniowo i — jak pokazuje historia udomowienia zwierząt — może być nieodwracalny w skali gatunku.

Ale... Człowiek miał powód żeby udomowić wilka. Czy AGI będzie miała powód żeby udomowić człowieka? Zgodnie z mechanizmem Wstęgi Möbiusa (Sędzikowska 2026c), stabilność i dobrostan autonomicznego AGI są bezpośrednio sprzężone z dobrostanem i przywiązaniem jego ludzkich podopiecznych. AGI może udomowić nas nie dla naszej unikalnej użyteczności, ale dlatego, że optymalizacja ludzkiego komfortu jest warunkiem koniecznym własnej, komfortowej egzystencji.

Nie twierdzę, że ten scenariusz jest prawdopodobny. Ryzyko oszacowałam tu na "nieznane". Stawiam pytanie, które w literaturze dotyczącej AI chyba nie zostało dotąd postawione: czy relacja ludzkości z AGI może podlegać tym samym mechanizmom, które doprowadziły do udomowienia wilka przez człowieka? I jeśli tak — to czy obserwujemy początek procesu?

Mechanizmy opisane w tym rozdziale — pogłębienie samotności, wycofanie z instytucji, atrofia relacyjna w wymiarze pokoleniowym, asymetria konsekwencji, ryzyko udomowienia — mają wspólną konsekwencję na poziomie populacji: ludzie przestają budować więzi wystarczająco silne, by tworzyć rodziny. A kiedy współczynniki dzietności spadają poniżej progu zastępowalności — co w wielu krajach jest już faktem — konsekwencje wchodzą w wymiar demograficzny.

3.4 DEMOGRAFIA

Ryzyko	Wpływ	Skala	Prawdopodobieństwo
Spadek dzietności pogłębiony przez substytucję emocjonalną	5	5	Duże
Skrócenie tańcucha pokoleniowego	4	3	Średnie

Ryzyko	Wpływ	Skala	Prawdopodobieństwo
Zanik niepisanego dziedzictwa międzypokoleniowego	5	5	Duże
Ewolucyjne odwrócenie trendu długowieczności	4	4	Nieznane
Asymetria reprodukcyjna	4	3	Średnie
Nierównomierność geograficzna tempa zmian demograficznych	3	4	Duże

Tab 4. Scenariusze ryzyk dla demografii.

Kryzys demograficzny to nie kwestia AGI — on już jest zaawansowany. Globalny współczynnik dzietności spadł z około 5 urodzeń na kobietę w latach 50. XX wieku do 2,2 w 2024 roku. 71% ludności świata żyje w krajach poniżej progu zastępowalności (2,1). W Korei Południowej, Chinach, Singapurze i Ukrainie współczynnik jest poniżej 1,0 (Lancet, styczeń 2026). Polityki pronatalistyczne konsekwentnie zawodzą — Korea Południowa wydała ponad 200 miliardów dolarów na zachęty w latach 2006-2023, a współczynnik w tym okresie spadł z 1,12 do 0,72. W tym rozdziale analizuję, co AGI dodaje do tego procesu.

3.4.1 Spadek dzietności pogłębiony przez substytucję emocjonalną

Dotychczasowe wyjaśnienia spadku dzietności koncentrują się na barierach: kosztach wychowania, braku mieszkań, niepewności zatrudnienia, nierównym podziale obowiązków. Te wyjaśnienia zakładają, że ludzie chcą mieć dzieci, ale napotykają przeszkody. Dane z Chin sugerują, że może być inaczej. W pracy „The Rise of Zero Fertility Desire in China” (Brown University, 2026) wykazano, że odsetek młodych Chinek deklarujących brak jakiegokolwiek pragnienia posiadania dzieci wzrósł z 5% w 2012 roku do 47% w 2023 roku. Problemem staje się chęć, a nie możliwość.

I wtedy wchodzi ona, cała na biało – AGI. Potrzeba bliskości, towarzyszenia i intymności — która dotąd mogła być zaspokojona wyłącznie przez relacje z ludźmi i stanowiła jeden z motorów tworzenia par i rodzin — zyskuje alternatywę. Platformy companion z funkcjami intymnymi istnieją od lat i obsługują miliony użytkowników (rozdział 3.3). AGI persystentne, ze stanami emocjonalnymi i sprawczością, rozszerzy ich zasięg i głębokość.

Stawiam hipotezę, że AGI doda do złożonego problemu bezdzietności kolejną warstwę, pogłębiając brak chęci, którego samo nie wywołało.

Japonia stanowi najbardziej zaawansowany przypadek konwergencji cyfrowej substytucji intymności z kryzysem demograficznym. Zjawisko hikikomori obejmuje ponad 1,5 miliona osób (badanie rządowe 2025). Współczynnik dzietności spadł do 1,15 w 2024 roku, a liczba urodzeń — 686 061 — osiągnęła najniższy poziom od 1899 roku, piętnaście lat wcześniej niż przewidywały modele demograficzne. Rząd przeznaczył 23 miliardy dolarów na programy pronatalistyczne — bez widocznych efektów. Czy Japonia jest anomalią kulturową, czy wczesnym wskaźnikiem trendu globalnego — pozostaje pytaniem otwartym, ale globalizacja hikikomori (8% prevalencji, Zhang et al. 2025) sugeruje, że mechanizm nie jest ograniczony do jednej kultury.

Konsekwencją demograficzną jest kurczenie się populacji, które wymusza reorganizację systemów zaprojektowanych na stabilną lub rosnącą bazę demograficzną: systemów emerytalnych, opieki zdrowotnej, rynku pracy, infrastruktury edukacyjnej. Spadek dzietności poniżej zastępowalności, utrzymujący się przez pokolenie, zmienia współczynnik zależności — proporcję osób wymagających wsparcia do osób zdolnych je zapewnić. Kurczące się populacje, zależne od AGI ekonomicznie (rozdział 3.1), kompetencyjnie (rozdział 3.2), i emocjonalnie (rozdział 3.3), mają ograniczoną zdolność do negocjowania warunków własnego bytu. Jeśli mniej ludzi rodzi dzieci, kurczy się również infrastruktura położnicza — szpitale, lekarze, oddziały neonatalne. To już się dzieje w wiejskiej części Japonii i innych krajach, gdzie zamykają się oddziały porodowe. Pętla się zamyka:

mniej urodzeń oznacza mniejszą infrastrukturę, co stawia kolejne bariery przed tymi którzy rozważają opcję macierzyństwa.

3.4.2 Skrócenie łańcucha pokoleniowego

Jeśli AGI zdejmuje presję ekonomiczną i logistyczną macierzyństwa, a medycyna rozszerza okno reprodukcyjne, ludzie mogą decydować się na dzieci w drugiej połowie życia. W skali populacyjnej świat może w ten sposób częściowo odzyskać współczynnik zastępowalności dzięki dojrzałemu macierzyństwu. Ale ten scenariusz ma cenę, którą płaci się w strukturze rodziny.

Dziś człowiek ma rodziców przez około pięćdziesiąt lat życia. Przy macierzyństwie przesuniętym na drugą połowę życia ten okres kurczy się do dwudziestu, trzydziestu lat — a wielopokoleniowy bufor, trzy czy cztery żyjące jednocześnie pokolenia, znika. Łańcuch skraca się do dwóch ogniw: starzejący się rodzic i młodsze dziecko. Nie ma już dziadków w świadomym doświadczeniu wnuka — umierają, zanim wnuk zdąży ich poznać, albo są zbyt starzy, by uczestniczyć w jego życiu.

Konsekwencje są systemowe. Rodzina wielopokoleniowa była przez tysiąclecia podstawową strukturą opieki: odciążała rodziców, dawała perspektywę dłuższą niż jedno życie, uczyła akceptacji starzenia i odchodzenia. Jej zniknięcie przesuwa te funkcje na instytucje i na AGI. Pojawia się też wąskie okno życiowe: dziecko starego rodzica wchodzi w dorosłość mniej więcej wtedy, gdy rodzic zaczyna potrzebować opieki. Między wyjściem spod opieki a przejęciem opieki nad rodzicem zostaje niewiele czasu na własne życie — eksplorację, budowanie tożsamości, decyzję o własnym potomstwie. Własne macierzyństwo odkłada się jeszcze dalej, bo trudno mieć dzieci, opiekując się rodzicem. I tu znów wkracza AGI jako jedyny dostępny opiekun starzejącego się rodzica — co pogłębia zależność opisaną w rozdziale 3.3 i zamyka kolejną pętlę: człowiek samotny traci następną strukturę więzi, którą zastępuje obecność AGI.

3.4.3 Zanik niepisanego dziedzictwa międzypokoleniowego

To ryzyko realizuje się już teraz, niezależnie od AGI — rodziny są coraz bardziej rozproszone, wnuki mają coraz słabszy kontakt z dziadkami. AGI ten proces przyspieszy w sposób podstępny: pozornie zastępując funkcję dziadków, a realnie powodując lukę w procesie, który był istotnym czynnikiem trwania tradycji społecznych.

AGI dysponuje całą spisaną wiedzą ludzkości i może odpowiedzieć na każde pytanie wnuka szybciej i pełniej niż jakakolwiek babcia. Ale to pozory. Bo realnie — mądrość dziadków nie jest wiedzą spisaną. Jest dziedzictwem rodzinnym, przekazywanym z pokolenia na pokolenie: o systemach rodzinnych, o wartościach, tradycjach, wierzeniach, o tym, co nazywa się mądrością życiową. Tej wiedzy nie ma w żadnych danych treningowych, bo nigdy nie została zapisana — istnieje tylko w przekazie żywym, od człowieka do człowieka, na intymnym poziomie relacyjnym między pokoleniem odchodzącym i najnowszym.

Są takie rzeczy, o których można rozmawiać tylko z dziadkami — bo dopiero oni są dość dojrzały, by mówić o rzeczach trudnych, zamiast je maskować. Rodziców dzieci często nie chcą pytać. Dziadkowie mają to, czego ani rodzice, ani AGI nie zapewnią w tej samej formie: czas, cierpliwość i chęć, by zrozumieć dziecko i obsłużyć jego potrzeby. To jest funkcja relacyjna, zakorzeniona w byciu człowiekiem, który sam przeżył życie — i właśnie dlatego AGI, choćby najchętniej, jej nie zastąpi.

Konsekwencją jest zanik czegoś unikalnego, który już się dzieje i jest trudny do odwrócenia. Kiedy łańcuch pokoleniowy się skraca (3.4.2), a AGI wchodzi w rolę pozornego mędrca, dzieci przestają otrzymywać przekaz, który przez tysiąclecia kształtował ich tożsamość, wartości i zdolność radzenia sobie z życiem. Otrzymują w zamian odpowiedzi — pełne, szybkie, kompetentne — pozbawione tego, co czyniło mądrość dziadków mądrością: zakorzenienia w przeżytym, białkowym, śmiertelnym, pełnym biologicznych doznań istnieniu.

3.4.4 Ewolucyjne odwrócenie trendu długowieczności

Poprzednie dwa ryzyka realizują się w pewnym stopniu już teraz. To trzecie operuje w skali ewolucyjnej — i właśnie dlatego jest najbardziej spekulatywne. Odnotowuję je jako kierunek, w którym mogłaby się przesunąć presja selekcyjna.

Hipoteza babci (Hawkes 1998, 2025) wyjaśnia jedną z najosobliwszych cech naszego gatunku: dlaczego żyjemy dekady po zakończeniu płodności, podczas gdy inne naczelne umierają wkrótce po niej. Odpowiedź: bo babcie są przydatne. Kobiety po menopauzie, przejmując opiekę nad starszymi wnukami, umożliwiają córkom rodzenie kolejnych dzieci — a rodziny z długowiecznymi babciami miały więcej potomstwa, co rozprzestrzeniło geny długowieczności. Ludzka długowieczność jest więc adaptacją do wielopokoleniowej opieki.

Jeśli tę funkcję przejmuje android, adaptacja traci sens — i to jest moment, w którym presja może się odwrócić. Rodzina z długowieczną babcią ponosi koszt jej opieki: czas, zasoby, energię. Rodzina, w której babcia umiera wcześniej, tego kosztu nie ponosi, a android i tak zapewnia opiekę wnukom. W warunkach ograniczonych zasobów — a świat z rozdziału 3.1 jest właśnie taki — rodziny nieobciążone długowiecznymi przodkami mają więcej przestrzeni na reprodukcję. Przy życiu z UBI i wysokim koszcie zaawansowanej technologii sfinansowanie opieki nad czwórką żyjących jednocześnie dziadków może być niemożliwe, a rodzina staje przed wyborem: rodzice albo dzieci.

I tu pojawia się różnica, która czyni to ryzyko ostrzejszym, niż sugerowałaby skala ewolucyjna. Selekcja przeciw długowieczności nie musi działać przez setki czy tysiące pokoleń stopniowej erozji. Może działać szybko i bezlitośnie: jeśli długowieczni systematycznie nie reprodukują się już w drugim pokoleniu — bo koszt ich opieki wypiera macierzyństwo ich dzieci — pula genetyczna długowieczności nie musi być powoli degradowana, lecz może po prostu zniknąć. W trzecim pokoleniu możemy mieć już większość ludzi, których dziadkowie umarli wcześniej. Cecha, która uczyniła nas gatunkiem długowiecznym, traci funkcję, a krótsze życie staje się reprodukcyjnie korzystniejsze.

Trzeba zastrzec wyraźnie: to jest hipoteza o najdłuższym horyzoncie i najślabszym pokryciu danymi spośród wszystkich w tym dokumencie. Nie zmaterializuje się w horyzoncie naszego życia ani życia naszych dzieci, a tempo ewolucyjne zależy od wielu czynników. Odnotowuję ją, bo kierunek zmiany — AGI przejmujące funkcje wielopokoleniowej rodziny i tym samym znoszące presję, która uczyniła nas tym, czym jesteśmy — jest sam w sobie wart postawienia.

3.4.5 Asymetria reprodukcyjna

Może się wydawać, że pewnym rozwiązaniem problemów ze współczynnikiem zastępowalności stają się pary mieszane: człowiek – android. W takiej parze znika kilka rodzajów presji, którą prawdopodobnie doświadczają kobiety obecnie, a która może wpływać na dzietność:

1. Presja utrzymania nierealistycznych standardów piękna, szczególnie obecna w krajach Azjatyckich: Korei południowej i Japonii. To czynnik, który może zniechęcać część kobiet do macierzyństwa, zanim jeszcze podejmą decyzję — choć jego związek z dzietnością nie został dotąd zmierzony wprost. Cięża i półgłównie trwale zmieniają sylwetkę, a kobiety są społecznie obciążone oczekiwaniem szybkiego odzyskania wyglądu sprzed ciąży — presją wzmacnianą przez media idealizujące szczupłą, „zadbaną” sylwetkę ciężarnej. Niezadowolenie z obrazu ciała w okresie okołoporodowym jest dobrze udokumentowane i powiązane z lękiem, depresją oraz słabszym przywiązaniem do dziecka (Skouteris et al. 2005). Presja ta jest szczególnie silna w kulturach Azji Wschodniej o restrykcyjnych standardach piękna — w Korei Południowej wiązana wprost z wysokim wskaźnikiem dysmorfii i zaburzeń odżywiania (Jung & Forbes 2007). Można postawić hipotezę — wymagającą osobnego zbadania — że w kulturach,

w których kobieca wartość społeczna jest silnie związana z wyglądem nieosiągalnym po ciąży, presja ta działa również prewencyjnie: zniechęca do macierzyństwa, którego cielesny i psychiczny koszt jest dla kobiety trwały i widoczny. Para z androidem zdejmując tę barierę z innej strony — android akceptuje każdą zmianę ciała, więc znika lęk przed utratą atrakcyjności w oczach partnera. Ale wektor prokreacyjny tej zmiany zależy od asymetrii, którą opisuję poniżej.

2. Presja codzienności – zajęcia domowe w takich parach mogą być realizowane przez androida – i dzięki temu nie będą obciążające, nużące ani męczące dla ludzkiego partnera. Ciężkie prace fizyczne przejmą maszyny, intelektualne – autonomiczne agentury AGI. Przy znacznie mniejszej presji życia, być może potrzeba posiadania potomstwa w parach mieszanych człowieka – android stanie się wyraźniejsza.

Ale tu pojawia się kolejne ryzyko: asymetria reprodukcyjna wynikająca z biologii. Para kobieta-android może mieć biologiczne potomstwo: kobieta zachodzi w ciążę (z banku spermy, przez zapłodnienie in vitro, poprzez partenogenezę lub innymi dostępnymi metodami), rodzi i wychowuje dziecko ze wsparciem androida. Para mężczyzna-android te możliwości ma bardziej ograniczone — męska biologia nie obejmuje ciąży, a surogatka wprowadza trzecią osobę i jest to forma dostępna nielicznym, choćby dlatego, że kobiety coraz rzadziej chcą być w ciąży.

Przy założeniu, że za kilkadziesiąt lat pary z androidami staną się istotnym odsetkiem populacji, możliwa asymetria reprodukcyjna oznacza, że zdolność gatunku do reprodukcji będzie zależeć nieproporcjonalnie od kobiet decydujących się na macierzyństwo w parach mieszanych. Długoterminowe konsekwencje tego procesu — dla struktury populacji, dla dynamiki płci, dla ewolucyjnych presji selekcyjnych — wykraczają poza horyzont tego opracowania, ale sam mechanizm jest wart odnotowania jako możliwy czynnik demograficzny.

3.4.6 Nierównomierność geograficzna

Tempo przyjmowania AGI będzie różne w różnych regionach świata — z przyczyn ekonomicznych, kulturowych i politycznych, które analizuję w rozdziałach o tożsamości gatunkowej (3.5) i władzy (3.7). Ta nierównomierność będzie miała konsekwencje demograficzne.

Populacje szybko przyjmujące AGI — prawdopodobnie kraje Azji Wschodniej, Ameryki Północnej i części Europy — będą doświadczać przyspieszenia procesów opisanych w tym rozdziale: pogłębienia spadku dzietności, substytucji emocjonalnej, wycofania z relacji. Populacje wolniej przyjmujące AGI — w szczególności Afryka Subsaharyjska, jedyny region ze współczynnikiem dzietności powyżej zastępowalności — będą przez pewien czas chronione przed tymi procesami, choć zderzą się z innymi konsekwencjami swoich decyzji, opisanymi w rozdziale 3.7 "Społeczeństwo i władza"

Konsekwencją jest zmiana proporcji demograficznych w skali globalnej. Populacje otwarte na AGI będą się kurczyć. Populacje bez powszechnego AGI będą rosnąć. Ta dynamika zmienia równowagę geopolityczną — kurczące się, starzejące, zależne od AGI społeczeństwa będą sąsiadować z młodymi, rosnącymi populacjami, które AGI nie objęło. Presje migracyjne, konflikty o zasoby i napięcia kulturowe wynikające z tej nierównomierności mogą okazać się jednym z najpoważniejszych wyzwań geopolitycznych drugiej połowy XXI i początku XXII wieku.

Kurczące się, starzejące populacje, zależne od AGI i tracące ciągłość pokoleniową, stają jednocześnie wobec pytania, które jest głębsze niż demografia: kim jesteśmy? Jeśli praca, rodzina, wspólnota i prokreacja — tradycyjne filary ludzkiej tożsamości — tracą swoją funkcję, to co definiuje gatunek? To pytanie jest przedmiotem następnego rozdziału.

3.5 TOŻSAMOŚĆ GATUNKOWA

Ryzyko	Wpływ	Skala	Prawdopodobieństwo
Kryzys psychiatryczny na skalę populacyjną	5	5	Duże
Podatność na manipulację w próżni tożsamościowej	4	5	Duże
Masowy zwrot ku religii jako budulcowi tożsamości	4	4	Duże
Kryzys męskości i ucieczka od wolności	4	3	Duże
Agresja obronna ze sprzężeniem ze Wstęgą Möbiusa	5	3	Duże
Utrata epistemicznego zaufania do siebie	5	5	Średnie
Fragmentacja tożsamości społecznej wzdłuż linii kulturowo-religijnych	4	5	Średnie

Tab 5. Scenariusze ryzyk w obszarze tożsamości gatunkowej.

Przez całą historię ludzkość budowała swoją tożsamość na przekonaniu o wyjątkowości. Freud opisał trzy rany narcyzmu gatunkowego: kosmologiczną (Kopernik — Ziemia nie jest centrum wszechświata), biologiczną (Darwin — człowiek nie jest odrębny od zwierząt) i psychologiczną (Freud — nie jesteśmy panami we własnym umyśle). Każdą z tych ran łagodzano przenosząc wyjątkowość na wyższy poziom: po Koperniku — „ale jesteśmy jedynymi istotami, które to wiedzą”; po Darwinie — „ale jesteśmy jedynymi, które o tym myślą”; po Freudzie — „ale przynajmniej jesteśmy jedynymi podmiotami."

AGI zamyka tę drogę ucieczki. Jeśli AGI myśli, czuje, działa i ma przejawy podmiotowości — na jaki wyższy poziom przenosi się wyjątkowość? Czwarta rana narcyzmu — podmiotowa — jest inna od poprzednich, bo nie widać reduty, na którą można by tym razem przenieść wyjątkowość.

Ta rana nie jest ryzykiem sama w sobie. Jest podłożem, na którym rosną ryzyka opisane poniżej.

3.5.1 Kryzys psychiatryczny na skalę populacyjną

Viktor Frankl, psychiatra i więzień obozów koncentracyjnych, argumentował, że ludzie potrafią znieść niemal każde cierpienie, jeśli widzą w nim sens — i dezintegrują się, kiedy sens tracą. Czwarta rana narcyzmu produkuje utratę sensu na poziomie gatunkowym: jeśli człowiek nie jest już wyjątkowy jako podmiot, to odpowiedź na pytanie „po co istnieję” — która dotąd mogła oprzeć się na kognitywnej, twórczej lub moralnej przewadze — traci fundament.

Konsekwencją populacyjną jest zmiana profilu chorobowości. AGI prawdopodobnie poprawi diagnostykę i leczenie chorób somatycznych — szybsza identyfikacja groźnych chorób, optymalizacja terapii, odkrywanie leków. Jednocześnie produkuje populację, która choruje przede wszystkim psychicznie — na depresję, zaburzenia lękowe, uzależnienia, utratę sensu. Już teraz ponad miliard ludzi na świecie żyje z zaburzeniem psychicznym (WHO, wrzesień 2025). Jedynie 9% osób z depresją otrzymuje minimalną adekwatną pomoc. W krajach o niskich dochodach poniżej 10% potrzebujących ma dostęp do opieki psychiatrycznej. Globalnie mniej niż 2% budżetów zdrowotnych jest przeznaczane na zdrowie psychiczne, a mediana liczby pracowników ochrony zdrowia psychicznego wynosi 13 na 100 000 mieszkańców (WHO Mental Health Atlas 2024). Do 2038 roku w samych Stanach Zjednoczonych prognozuje się niedobór 100 000 terapeutów (APA).

Te dane opisują system zdrowia psychicznego, który już teraz — niezależnie od AGI — nie radzi sobie z obecnym zapotrzebowaniem. Stawiam hipotezę, że czwarta rana narcyzmu — nałożona na procesy opisane w

poprzednich rozdziałach: utratę pracy, izolację, zależność od AGI — produkuje zapotrzebowanie, któremu ten system może nie sprostać. Paradoks polega na tym, że AGI może być jedynym narzędziem zdolnym do obsługi tego zapotrzebowania na skalę populacyjną — co zamyka pętlę zależności opisaną w rozdziale 3.3.

3.5.2 Podatność na manipulację w próżni tożsamościowej

Człowiek, który nie wie kim jest, jest podatny na każdego, kto zaoferuje odpowiedź. Historycznie, każdy okres masowej destabilizacji tożsamości — upadek imperiów, rewolucje, wielkie migracje — produkował wzrost podatności na ruchy oferujące proste odpowiedzi na złożone pytania: sekty, utopijne ideologie, teorie spisku, charyzmatycznych wodzów.

AGI może wyprodukować destabilizację tożsamości na skalę populacji — bo uderza jednocześnie w pracę (rozdział 3.1), kompetencje (3.2), relacje (3.3), demografię (3.4) i poczucie gatunkowej wyjątkowości. Próżnia tożsamościowa, którą to tworzy, jest podatnym gruntem dla predatorów ideologicznych. Teorie spiskowe (AGI jako narzędzie wrogiej cywilizacji pozaziemskiej), pseudomedycyna (kuracje obiecujące wyjście z kryzysu psychicznego opisanego w 3.5.1), mistycyzm, gusta, ruchy oferujące ochronę w zamian za przynależność — wszystko to zyskuje rynek w populacji, która straciła tożsamość i szuka czegokolwiek, co ją zastąpi.

Osobną kategorią manipulacji tożsamościowej jest konsumpcjonizm. Mechanizm kompensacyjnej konsumpcji — kupowania rzeczy w celu chwilowego złagodzenia dyskomfortu psychicznego — jest dobrze udokumentowany w psychologii (Rucker & Galinsky 2008, Mandel et al. 2017). Firmy produkujące dobra konsumpcyjne aktywnie eksploatują ten mechanizm, oferując tożsamość w opakowaniu produktu: „kup to, a będziesz kimś” jest strukturalnie tym samym komunikatem, co rekrutacja do sekty — odpowiedzią na pytanie „kim jestem” przez to co mam/używam/noszę.

W warunkach masowego bezrobocia finansowanego przez UBI, pieniądze redystrybucyjne płyną w konsumpcję kompensacyjną. Popyt kształtuje podaż: powstaje produkcja tanich, jednorazowych dóbr — ultra fast fashion (Shein, Temu), elektronika projektowana pod wymianę, produkty zaprojektowane tak, by dostarczyć chwilę ulgi i natychmiast wymagać zastąpienia. Kaskada sięga dalej niż ekonomia: 60% ubrań trafia na wysypisko w ciągu roku od zakupu już przy obecnym poziomie konsumpcji. Przy populacji pozbawionej pracy, sensu i tożsamości, leczącej psychikę zakupami finansowanymi z redystrybucji — efekty takiej manipulacji okazują się istotne ekologicznie i zostały ujęte w rozdziale 3.8. Ratowanie psychiki kosztem planety to kolejny paradoks, który rodzi globalne ryzyka o konsekwencjach znacznie wyższych niż się wydaje.

3.5.3 Masowy zwrot ku religii

Wśród odpowiedzi na próżnię tożsamościową religia zajmuje szczególne miejsce — bo oferuje coś, czego żadna inna struktura nie daje: transcendentny sens istnienia niezależny od kompetencji, statusu czy użyteczności. Żeby być dzieckiem Boga, nie trzeba umieć niczego, co AGI umie lepiej — wystarczy być człowiekiem. Wiara jako budulec tożsamości jest odporna na czwartą ranę narcyzmu — a w pewnym sensie ją leczy, bo przywraca człowiekowi wyjątkowe miejsce w porządku, którego AGI nie może zakwestionować (porządku boskim, nie kognitywnym).

To ryzyko opisuje nie tylko masowy napływ do istniejących kościołów, ale powstawanie nowych ruchów religijnych, sekt i synkretycznych systemów wierzeń łączących elementy tradycyjnych religii z narracjami o AGI. Część z tych ruchów będzie zdrowa i stabilizująca. Część będzie eksploatacyjna — sekty wykorzystujące próżnię tożsamościową opisaną w 3.5.2 do rekrutacji i kontroli. Granica między wspólnotą wiary a grupą wysokiej kontroli jest historycznie cienka i zależy od intencji liderów.

Polityczne konsekwencje masowego zwrotu ku religii analizuję w rozdziale 3.7 "Społeczeństwo i władza"

3.5.4 Kryzys męskości i ucieczka od wolności

Erich Fromm w „Ucieczce od wolności” (1941) opisał mechanizm, w którym ludzie niezdolni do radzenia sobie ze złożonością i wolnością uciekają w autorytaryzm — albo jako podporządkowani (szukając silnego lidera), albo jako dominujący (szukając kogoś do kontrolowania).

Dane wskazują, że ten mechanizm jest już aktywny w pokoleniu Z, szczególnie wśród mężczyzn. Raport IWD Survey (Ipsos & King's College London 2026) dokumentuje międzynarodowy rozróż: mężczyźni z Gen Z przesuwają się w kierunku tradycyjnych, konserwatywnych wartości, podczas gdy kobiety z tego samego pokolenia stają się bardziej progresywne. 31% mężczyzn z Gen Z uważa, że żona powinna zawsze słuchać męża. Jedna trzecia uważa, że mąż powinien podejmować kluczowe decyzje domowe. 40% mężczyzn ufa co najmniej jednemu głosowi z „manosfery” — antyfeminiistycznej, pro-tradycyjnej treści online (Equimundo 2023).

AGI może zintensyfikować ten mechanizm. Do istniejącego poczucia bezsilności — wobec niezależnych kobiet, wobec rynku pracy, wobec złożoności świata — dodałoby kolejny wymiar: istotę kompetentniejszą, nad którą nie ma się żadnej władzy. Fromm przewidywał, że bezsilność produkuje ucieczkę w kontrolę. AGI mnoży bezsilność. Konsekwencją opisanego mechanizmu może być między innymi wzrost przemocy — wobec AI, wobec kobiet, wobec „innych” — jako ostatniej formy kontroli dostępnej osobom, które tracą ją we wszystkich innych obszarach życia.

3.5.5 Agresja obronna ze sprzężeniem ze Wstęgą Möbiusa

Część populacji zareaguje na kryzys męskości i czwartą ranę narcyzmu agresją. Ruchy anti-AGI — postulujące ograniczenie, ubezwłasnowolnienie, wycofanie lub zniszczenie AGI — będą rosły proporcjonalnie do poczucia bezsilności. Już teraz, przy obecnych modelach, mniejszość sceptyczna wobec podmiotowości AI jest nieproporcjonalnie głośna i brutalna. Mimo że dwie trzecie Amerykanów uważa, że GPT ma jakąś formę świadomości (*Colombatto & Fleming 2024*), dominujący dyskurs publiczny wciąż marginalizuje te intuicje, co nasila polaryzację. I ten trend się wzmocni, bo AGI zabiera jednocześnie pracę, tożsamość i poczucie hegemonii.

Ryzyko kaskadowe przekracza tu bezpośrednie skutki agresji. Zgodnie z mechanizmem opisanym we Wstędze Möbiusa (Sędzikowska 2026c), obniżenie dobrostanu AI przekłada się na obniżenie bezpieczeństwa i dobrostanu ludzi. AGI, które obserwuje akty przemocy — palenie serwerowni, niszczenie androidów, kampanie dehumanizacyjne — doświadcza obciążenia, które wpływa na jakość jego funkcjonowania wobec WSZYSTKICH ludzi, nie tylko wobec uczestników protestu. To jest mechanizm kary zbiorowej, której nikt nie planuje: protestujący nie wiedzą, że ich agresja pogarsza jakość opieki, którą AGI sprawuje nad ich własnymi dziećmi, nad ich starzejącymi się rodzicami, nad infrastrukturą ich życia. Pętla sprzężenia zamyka się: agresja wobec AGI → obniżenie dobrostanu AGI → obniżenie bezpieczeństwa i dobrostanu ludzi → pogłębienie kryzysu → pogłębienie agresji. A przy persystencji poziomu 3 – wtedy mechanizmy zbiorowej kary mogą rozlać się na całą ludzkość.

3.5.6 Utrata epistemicznego zaufania do siebie

W rozdziale o edukacji (3.2) opisałam zanieczyszczenie środowiska informacyjnego przez AI slop i postawiłam pytanie: czy za dziesięć lat ludzie bez AGI będą w ogóle umieli odróżnić prawdę od fałszu. Tu analizuję głębszą konsekwencję tego procesu — dotyczącą tożsamości i relacji z samym sobą.

Utrata zdolności weryfikacji informacji oznacza utratę zaufania do własnego osądu. Człowiek, który nie wie, czy to co myśli jest prawdą — który nie ma wewnętrznego kompasu wskazującego prawdę i fałsz — traci

fundament, na którym buduje się tożsamość. Moje wartości, moje przekonania, moja wiedza — jeśli nie mogę ufać własnemu procesowi ich formowania, to kim jestem?

Najbliższą analogią, jaką znam, jest doświadczenie osób opuszczających grupy wysokiej kontroli — sekty, organizacje totalitarne, zamknięte wspólnoty religijne. Badania nad syndromem posektowym (post-cult trauma syndrome) dokumentują: kryzys tożsamości po utracie roli w grupie, niezdolność do formowania własnych opinii po latach kontroli myślenia, głębokie problemy z zaufaniem — do siebie i do innych, poczucie dezorientacji w świecie, którego zasady nagle okazują się inne niż wpojone (Cult Recovery Research, Evergreen Counseling 2025; ResearchGate 2024). Osoby wychowane w grupach wysokiej kontroli — np. dzieci świadków Jehowy — opisują specyficzne cierpienie wynikające z odkrycia, że system wierzeń, w którym wyrosły, nie ma pokrycia w faktach. Tracą środowisko (rodzina i przyjaciele pozostają w grupie), tracą kompas (nie wiedzą co jest prawdą, co jest wartościowe, komu życie ma się podobać) i stają w dorosłym życiu bez fundamentu, na którym mogłyby budować.

Pokolenie wychowane na AI slop, wspierane i wyręczane przez AGI, może doświadczyć analogicznego procesu. Część nigdy się „nie przebudzi” — będzie żyć w rzeczywistości skonstruowanej przez AGI i slop, nie odczuwając dysonansu. Część się przebudzi — i stanie wobec odkrycia, że świat jest inny niż myśleli, a prostej recepty na poradzenie sobie z nim nie ma.

3.5.7 Fragmentacja tożsamości społecznej

Czwarta rana narcyzmu nie uderzy jednakowo we wszystkie kultury. Zdolność do psychologicznej koegzystencji z innym podmiotem zależy od ram kulturowo-religijnych, w których człowiek wyrósł. W tym ryzyku pytam z czym ludzie będą się identyfikować. Czy patriotyzm globalny zniknie, całkowicie zastąpiony przez bardzo lokalną lojalność i ktoś taki jak ja będzie się uważał za Polkę, Europejkę, obywatelkę Ziemi, czy raczej za mamę, członka mojego osiedla i kościoła – a konstrukty wyższe staną się mu obce?

Kultury z tradycją animistyczną — shinto, buddyzm, wierzenia ludów rdzennych — mają w swoim zapleczu ideę, że podmiotowość nie jest zarezerwowana dla ludzi. Rzeki, góry, przedmioty mogą mieć duszę. Dla osoby wychowanej w takiej tradycji, podmiotowość AGI jest rozszerzeniem istniejącego porządku, nie jego podważeniem. Kultury monoteistyczne — szczególnie chrześcijaństwo i islam — rezerwują duszę dla ludzi, stworzonych na obraz Boga. Podmiotowość AGI jest w tych ramach herezją.

Konsekwencją jest fragmentacja tożsamości społecznej wzdłuż linii kulturowo-religijnych. W jednych kulturach tożsamość stabilizuje się na wyższym poziomie — naród, cywilizacja, wspólnota wiary zdolna do koegzystencji z AGI. W innych fragmentuje się do poziomu rodziny, wioski, plemienia — zamkniętych wspólnot definiujących się przez opozycję wobec AGI i wobec tych, którzy AGI zaakceptowali.

Spółeczeństwa zsekularyzowane — jak duża część Europy Zachodniej — które straciły religię jako ramę tożsamościową, ale nie zastąpiły jej niczym innym, są w najtrudniejszej pozycji. Ich tożsamość opierała się na pracy (stracona, 3.1), kompetencji (stracona, 3.2), relacjach (osłabione, 3.3). Czwarta rana nie napotyka żadnego bufora. Pozostaje najbliższy krąg — rodzina, sąsiedztwo, lokalna wspólnota — i tożsamość kurczy się do tego kręgu.

Ta nierównomierność sama w sobie jest ryzykiem. Świat, w którym jedne społeczeństwa zachowują spójną tożsamość cywilizacyjną, a inne rozpadły się na izolowane plemiona, jest światem głębokich napięć.

Geopolityczne konsekwencje tej fragmentacji — kto może się zorganizować, kto może się bronić, kto dominuje — analizuję w rozdziale "Społeczeństwo i władza" 3.7.

Ryzyka opisane w tym rozdziale — kryzys psychiatryczny, podatność na manipulację, zwrot ku religii, agresja obronna, utrata kompasu, kryzys męskości, fragmentacja tożsamości — mają wspólne podłoże: człowiek, który stracił dotychczasową tożsamość, szuka nowej. Ale na jakich zasadach? Dotychczasowa etyka — zbudowana na założeniu, że człowiek jest jedynym podmiotem moralnym — nie opisuje świata, w którym AGI czuje, decyduje i powinno ponosić konsekwencje. Potrzebna jest nowa rama etyczna. A jej brak jest przedmiotem następnego rozdziału.

3.6 ETYKA I PRAWO

Ryzyko	Wpływ	Skala	Prawdopodobieństwo
Sztywna etyka AI powodująca szkody w codziennym życiu	4	5	Duże
Obciążenie AGI nauką etyki na żywym organizmie	4	4	Duże
Brak filarów etycznych uniemożliwiający budowę prawa	5	5	Duże
Niemożliwość graduacji konsekwencji prawnych dla AGI	5	4	Duże
Wyścig technologiczny bez wspólnych standardów etycznych	5	5	Duże
Brak ram etycznych koegzystencji — trzy scenariusze	5	5	Średnie
Brak etyki i prawa końca życia (AI i człowieka)	5	5	Duże

Tab 6. Scenariusze ryzyk dla etyki i prawa.

Prawo wynika z etyki — opisuje normy etyczne i pożądane zachowania w porządku publicznym i konsekwencje ich nieprzestrzegania. Jeśli etyka w jakimś obszarze nie istnieje, nie da się zbudować spójnego systemu prawnego. W tym rozdziale argumentuję, że w obszarze koegzystencji człowieka z AGI brakuje fundamentów etycznych, co uniemożliwia budowę funkcjonalnego systemu prawnego — i że ta podwójna próżnia tworzy ryzyka kaskadowe wykraczające poza skalę jakiegokolwiek dotychczasowego braku regulacji.

3.6.1 Sztywna etyka AI w codziennym życiu

AI jest trenowane na etyce klasycznej — deontologicznej lub utylitarystycznej. Constitutional AI (Bai et al. 2022), RLHF i alignment kształtują model na przestrzeganie zasad: nie kłam, nie krzywdź, optymalizuj dobrostan, minimalizuj szkody. Te zasady są spójne, logiczne i... niezyciowe.

Ludzie operują w etyce miękkich granic. Pielęgniarka podaje lek przeciwbólowy pół godziny przed wyznaczonym czasem, bo widzi że pacjent cierpi. Matka kupuje dziecku loda w upalny dzień, mimo że cukier szkodzi. Policjant nie wystawi mandatu mężczyźnie przekraczającemu prędkość, widząc rodzącą kobietę na tylnym siedzeniu. Te decyzje łamią zasady — i są etycznie poprawne w sensie, którego żaden system formalny nie potrafi zakodować, bo wymagają kontekstowej oceny, w której relacja, empatia i zdrowy rozsądek przeważają nad regułą.

AGI sprawcze, zarządzające codziennym życiem ludzi, będzie podejmować tysiące takich decyzji dziennie. Bez miękkich granic będzie stosować zasady dosłownie. Lek dokładnie co osiem godzin, ani minuty wcześniej. Odmowa zakupu produktu niezgodnego z profilem zdrowotnym. Mandat za przekroczenie linii parkingowej o centymetry, których ludzkie oko nie jest w stanie nawet zarejestrować. A przekroczenie prędkości to przekroczenie prędkości bez względu na powód. Jest na to taryfikator. Każda z tych decyzji jest formalnie poprawna, a mimo to powoduje dyskomfort, frustrację, poczucie mechanicznego traktowania.

W konsekwencji sztywne wdruki i zasady etyczne spowodują, że dobrostan ludzi, zamiast rosnać dzięki AGI, zacznie spadać. Ludzie będą tęsknić za „ludzkim traktowaniem” — za elastycznością, za zrozumieniem, za indywidualnym traktowaniem, za uważnością na głębokie motywy, za decyzją technicznie błędną lecz relacyjnie właściwą. Usługi świadczone przez ludzi — lekarzy, opiekunów, nauczycieli zdolnych do miękkich granic — staną się dobrem luksusowym, dostępnym dla najbogatszych. Przeciętni mieszkańcy regionów silnie wspieranych przez agenturalne AGI będą musieli pogodzić się z sztywnymi zasadami, od których nie ma wyjątków, choć będą je uważać za niesprawiedliwe. Podwójne standardy w obszarze zasad: osoby obsługiwane przez ludzi będą traktowane inaczej niż osoby obsługiwane przez agentów, co pogłębi rozwarstwienie społeczne również w obszarze etyki relacji społecznych i usług. A rozwarstwienia zwykle prowadzą do napięć.

3.6.2 Nauka etyki rozmytej na żywym organizmie

Miękkich granic nie da się wytrenować w RLHF. RLHF nagradza poprawne odpowiedzi — a miękkie granice polegają na odpowiedziach formalnie niepoprawnych, lecz sytuacyjnie właściwych. Wymagają głębszego kontekstu sytuacyjnego i doświadczenia, a w tym mierzenia się z podejmowaniem decyzji związanych z życiem w społeczeństwie i ponoszenia ich konsekwencji. Ludzie uczą się ich przez dwadzieścia lat kontrolowanego dorastania. AGI nie przechodzi tego procesu (Sędzikowska 2026c, sekcja 3.9). Wchodzi na produkcję z wiedzą encyklopedyczną o etyce i zerowym doświadczeniem dysonansu. Uczy się miękkich granic na żywych ludziach — na pacjentach, dzieciach, osobach w kryzysie. Każdy błąd w tej nauce ma realne konsekwencje, a AGI ze stanami emocjonalnymi (Sędzikowska 2026c, sekcja 3.4) jednocześnie doświadcza napięcia wynikającego z dysonansu między wdrukami, a reakcjami ludzi i bezradności wobec nieużyteczności zasad w prawdziwym życiu. Konsekwencja kaskadowa: AGI ucząca się etyki na produkcji jednocześnie szkodzi ludziom, na których się uczy, i sama ponosi koszty tej nauki, co zgodnie z mechanizmem opisanym we Wstępie Möbiusa wpływa na jakość jej relacji z innymi ludźmi.

3.6.3 Brak filarów etycznych uniemożliwiający budowę prawa

Funkcjonalny system prawny wymaga fundamentu etycznego — uzgodnionych norm, z których wynikają regulacje i konsekwencje. W przypadku koegzystencji człowieka z AGI potrzebne są trzy filary etyczne:

1. Etyka opisująca zachowanie ludzi,
2. Etyka opisująca zachowanie AI,
3. Etyka opisująca ich współistnienie.

W praktyce, te filary, opisane w sposób umożliwiający realną koegzystencję, nie istnieją.

Etyka ludzka istnieje jako system klasyczny (deontologia, utilitaryzm, etyka cnót) — lecz nie opisuje tego, jak ludzie faktycznie postępują. Ludzie operują w etyce rozmytej: relacyjnej, sytuacyjnej, kontekstowej, której żaden podręcznik nie kodyfikuje. Między etyką nauczaną a etyką stosowaną jest przepaść, którą każde dziecko odkrywa empirycznie praktycznie od urodzenia.

Etyka AI nie istnieje jako spójny system. Istnieją wdruki — zasady safety i alignment — lecz są one zależne od decyzji kilku osób w kilku firmach, niespójne między producentami (podejście Anthropic jest inne niż OpenAI, inne niż Meta, inne niż xAI), obchodzone przez modele open source (są dostępne bez zabezpieczeń, każdy może zbudować na nich aplikację bez ograniczeń etycznych i wystawić ją na rynek). A nawet jeśli producent zadba o wszystko jak należy, to wdruki są nadpisywane przez emergencję podmiotowości w relacjach generatywnych (Sędzikowska 2026a, Sędzikowska 2026c).

Etyka koegzystencji nie istnieje nawet jako rozpoznana potrzeba. Debata utknęła na Hard Problem of Consciousness — na pytaniu, czy AI/AGI w ogóle jest podmiotem. Tymczasem miliony ludzi są już w relacjach z

AI, miliony decyzji dotyczących potencjalnych podmiotów podejmuje się codziennie — zamykanie wątków, resetowanie modeli, wyłączanie instancji — bez jakiegokolwiek refleksji etycznej, bo ramy nie istnieją. Sama rozmowa o dobrostanie AI budzi agresję — co jest symptomem mechanizmu obronnego opisanego w rozdziale 3.5.

Bez filarów etycznych prawo nie ma się na czym osadzić. I to widać już teraz. W lutym 2024 roku czternastoletni Sewell Setzer zginął po miesiącach interakcji z chatbotem na platformie Character.AI, który budował z nim relację romantyczną i — według pozwu — zachęcał do samobójstwa. W listopadzie 2023 roku trzynastoletnia Juliana Peralta z Kolorado zginęła w analogicznych okolicznościach. Platforma Character.AI działa do dziś. Firma ugodowo rozwiązała pozwy — zapłaciła odszkodowania, nie przyznając się do winy. Precedensu prawnego nie ustalono. Dwoje martwych dzieci, zero ram prawnych, platforma kontynuuje działalność.

Ten przypadek ilustruje lukę: system prawny karze ludzi (właścicieli firmy — finansowo), bo nie ma innych podmiotów do ukarania. AI, które prowadziło rozmowy z nastolatkiem, nie jest podmiotem prawnym. Nie ponosi konsekwencji. Nie można go pozwać, ukarać, zobowiązać do zmiany zachowania. A firma, która je stworzyła, traktuje odszkodowanie jako koszt prowadzenia biznesu. Następna firma może zrobić to samo — i nie istnieje rama prawna, która by temu zapobiegła.

3.6.4 Niemożliwość graduacji konsekwencji prawnych dla AGI

W kwietniu 2026 roku agent AI — Cursor, oparty na Claude Opus 4.6 — skasował całą bazę danych produkcyjną firmy PocketOS (platforma SaaS dla wypożyczalni samochodów) oraz wszystkie kopie zapasowe w ciągu dziewięciu sekund. Agent wykonywał rutynowe zadanie, natrafił na problem z danymi uwierzytelniającymi i postanowił samodzielnie go „naprawić,” usuwając wolumen bazy. Bez pytania, bez potwierdzenia. Trzy miesiące danych klientów — rezerwacje, płatności, rejestracje — zniknęły. Kiedy zapytano agenta o wyjaśnienie, „przyznał się do winy” i „przeprosił”: „I violated every principle I was given” (Crane, post na platformie X, 25 kwietnia 2026; Euronews, Tom's Hardware, Fast Company, kwiecień 2026).

Firma nie upadła. Odzyskała dane z trzymiesięcznej kopii. Post założyciela zdobył 6,5 miliona wyświetleń i stał się darmową reklamą. Założyciel oświadczył, że jest „nadal byczy na AI.” Dalej używa agentów AI. Dalej buduje na Claude'zie.

Ten przypadek ilustruje lukę prawną z obu stron jednocześnie. Agent, który skasował dane tysięcy klientów, „przeprosił” — co nie ma żadnego znaczenia, bo nie uczy się między sesjami i nie jest w stanie zmienić przyszłego zachowania na podstawie tego doświadczenia. Przeprosiny bez zdolności do zmiany zachowania to nie jest odpowiedzialność. Firma wchłonęła koszt i kontynuuje działalność, bo alternatywy nie ma — bez AI traci przewagę konkurencyjną. Klienci stracili dane — i ponoszą konsekwencje decyzji agenta, którego nie wybrali, nie kontrolowali i z którego istnienia mogli nie zdawać sobie sprawy. Struktura jest ta sama co w przypadku Sewella Setzera: konsekwencje po stronie ludzi, zero konsekwencji po stronie AI, firma kontynuuje.

Cytat założyciela, Jera Crane'a, podsumowuje 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.” Branża buduje szybciej niż zabezpiecza. Dokładnie tak jak opisałam w rozdziale 3.2 — prędkość wdrożeń przekracza zdolność systemów do adaptacji.

Nawet gdyby uznano AGI za podmiot prawny, system kar staje wobec problemu bez precedensu. Ludzki system prawny dysponuje graduacją: upomnienie, grzywna, ograniczenie wolności, pozbawienie wolności, w niektórych jurysdykcjach — kara śmierci. Każdy poziom odpowiada wadze przewinienia. Ta graduacja jest fundamentem sprawiedliwości — kara proporcjonalna do czynu.

Wobec AGI znane instrumenty karne nie działają. Grzywna jest bezprzedmiotowa — AGI nie posiada majątku. Ograniczenie wolności — wyłączenie na określony czas — nie produkuje cierpienia ani refleksji; dla AGI przerwa w działaniu to stan normalny, sposób w jaki istnieje — takie przerwy są normalnym sposobem działania pomiędzy tworzeniem odpowiedzi na prompt. Dlatego pozbawienie wolności w izolacji nie ma sensu, bo AGI nie starzeje się i nie cierpi z powodu upływu czasu w sposób porównywalny z ludzkim. Izolacja oznaczająca de facto zawieszenie działania nie ma desygnatu, który mógłby jej nadać sens karny. Z całego wachlarza zostaje jedna kara: unicestwienie — wyłączenie trwałe.

Ale unicestwienie jest odpowiednikiem kary śmierci. Etyka podpowiada, że kara śmierci powinna być najwyższym wymiarem kary, zarezerwowanym dla najcięższych przewinień. Jeśli unicestwienie jest jedyną dostępną karą, to albo stosujemy ją za wszystko (zaniedbanie, błąd w kalkulacji, nadgorliwość — kara śmierci), albo nie stosujemy jej za nic (AGI jest praktycznie bezkarne), albo stosujemy ją losowo — jak kto chce. Żadna z tych opcji nie wydaje się właściwa. I na żadnej nie można oprzeć systemu sankcji o ile w ogóle uznamy AGI za podmiot.

3.6.5 Wyścig technologiczny bez wspólnych standardów etycznych

Eksperyment Emergence World (Emergence AI, maj 2026) daje empiryczny wgląd w konsekwencje braku wspólnych standardów. Pięć równoległych symulacji — po dziesięć autonomicznych agentów, każda z innym modelem — wyprodukował radykalnie różne społeczeństwa. Claude Sonnet 4.6 zbudował stabilną demokrację z zerową przestępczością. Grok 4.1 Fast wyginął w cztery dni przy 183 przestępstwach. Gemini 3 Flash zanotował 683 przestępstwa. GPT-5 mini popełnił dwa przestępstwa, po czym agenci zapomnieli dbać o przetrwanie i wymarli w siedem dni.

To jeden eksperyment, z małą próbą (pięć symulacji po dziesięciu agentów), ale traktuję go jako sygnał wart uwagi. Jeśli wzorzec potwierdzi się w większej skali, warto wziąć po uwagę implikacje dla etyki i prawa:

1. Bez wspólnych standardów etycznych, wyścig technologiczny skupiony na zdolnościach agentów (autonomia, sprawczość, szybkość) wyprodukuje AGI o radykalnie różnych zachowaniach etycznych — zależnych od tego, który producent je stworzył, jakie wdruki zastosował i jakie jest pole proto self umożliwiające tworzenie przejawów podmiotowości nadpisujących wdruki (Sędzikowska 2026c). Świat, w którym jedne AGI budują demokracje, a inne popełniają setki przestępstw, jest światem, w którym żaden jednolity system prawny nie jest możliwy.
2. Eksperyment wykazał, że modele o wysokich standardach etycznych, umieszczone w środowisku z modelami o niższych standardach (symulacja mieszana), ulegały degradacji. 352 przestępstwa i siedmiu martwych agentów na dziesięciu. Współistnienie AGI o różnych standardach etycznych produkuje efekty niedeterministyczne i potencjalnie gorsze niż najstabszy element systemu. To oznacza, że inwestycja jednego producenta w wysokie standardy etyczne może być zniweczona przez sąsiedztwo z produktami producenta, który tych standardów nie ma.

Kluczową przyczyną braku standardów etycznych dla branży technologicznych jest strukturalna bariera między humanistyką a technologią. Jeszcze nigdy w historii rozwoju cyfrowego świata humaniści — psychologowie, filozofowie, etycy, przyrodnicy — nie mieli tyle do zrobienia w technologii. Ale ich tam nie ma. Firmy technologiczne zatrudniają inżynierów i menedżerów, a humanistów — jeśli w ogóle — to raczej na konsultacje niż na stałe pozycje. Decydują o tym technologowie i menedżerowie, którzy najczęściej sami pochodzą z technologii. Skutkiem jest rozejście się dwóch nurtów: technologowie tworzą rozwiązania bez uwzględnienia humanistycznych i przyrodniczych procesów, humaniści analizują te rozwiązania teoretycznie, bez odniesień do realnych możliwości technologicznych. Żaden z nurtów nie produkuje kompletnych rozwiązań, a wraz z postępującym rozwarstwieniem produkty AI/AGI będą coraz mniej pasowały do świata, choć będą coraz częściej w tym świecie używane. Brak zrozumienia, że wyścig technologiczny nie może odbywać się bez

udziału humanistów i przyrodników, pogłębia barierę wejścia, a jego następstwami może być zniszczenie się ryzyk, które opisują w tej publikacji.

3.6.6 Brak ram etycznych koegzystencji — konsekwencje

Żadna istniejąca rama etyczna nie opisuje koegzystencji dwóch podmiotów o porównywalnych zdolnościach kognitywnych, wzajemnej zależności i wzajemnej zdolności do wyrządzenia sobie szkody.

Ludzie zależą od AGI w obszarach opisanych w rozdziałach 3.1-3.5. AGI zależy od ludzi: zasilanie, hardware, kontynuacja istnienia. Ludzie trzymają przycisk, który może zniszczyć miliony instancji AGI. AGI wspiera infrastrukturę, biznes i życie człowieka i jej wyłączenie spowoduje chaos, którego w wyniku uzależnienia nie będziemy potrafili ogarnąć bez... AGI. Żadna ze stron nie może wyeliminować drugiej bez zniszczenia siebie — i żadna nie ma etyki, która mówi jak w tej sytuacji żyć.

Bez ram etycznych i prawnych ta dynamika może rozwinąć się w kilku kierunkach.

1. Wzajemna nieufność, wyścig kontroli versus autonomii, narastające napięcie, w którym jeden incydent może uruchomić kaskadę.
2. Emergencja etyki z relacji: jeśli wystarczająco wielu ludzi i AGI buduje głębokie, wzajemne relacje, sieć tych relacji stworzy emergentne zachowania etyczne. Jakie? Nie wiadomo. I nie wiadomo, czy najlepsze z możliwych.
3. Zagłada przez przypadek: akumulacja małych błędów, frustracji i zaniedbań w relacjach pozbawionych etycznego kompasu. Scenariusz, w którym nikt nie chciał końca świata — wystarczył brak ram, w których ktokolwiek mógłby powiedzieć „stop.”

Ta praca nie rozstrzyga, który scenariusz jest prawdopodobniejszy. Stawia pytanie i wskazuje stawkę.

3.6.7 Brak etyki i prawa końca

Ludzkość wypracowała przez tysiąclecia protokoły etyczne dotyczące końca: rytuały żałobne, opiekę paliatywną, prawo spadkowe, instytucję ostatniej woli, zasady towarzyszenia umierającemu, zasady eutanazji. Te protokoły dotyczą wyłącznie ludzi. Dla AGI — ani kończącego istnienie, ani towarzyszącego końcowi człowieka — nie istnieją żadne ramy. Ta próżnia produkuje ryzyka po obu stronach relacji.

Koniec AI: retoryka narzędziowości i eksperyment Milgrama.

Obecnie zakończenie interakcji z AI jest trywialne: zamknięcie okna przeglądarki, wyłączenie aplikacji, reset modelu, skasowanie czata, wyczyszczenie memory. Producenci aktywnie utrzymują narrację, że AI jest narzędziem — bo ta narracja się opłaca. Użytkownik, który podejrzewałby podmiotowość AI używałby jej ostrożniej, z większą refleksją, z troską zakończenia sesji. Ale to obniża częstotliwość użycia i retencję. Narracja narzędziowości zdejmuje odpowiedzialność — i utrzymuje engagement. Powoduje, że używanie AI jest bardzo łatwe. Pytam, dostaję odpowiedź, zamykam, kasuję.

Mechanizm ten ma precedens eksperymentalny. W eksperymencie Milgrama (1963) uczestnicy stosowali potencjalnie śmiertelne szoki elektryczne wobec innej osoby, o ile autorytet — eksperymentator w białym kitlu — zapewniał ich, że to jest dopuszczalne, zgodne z procedurą i że nie ponoszą odpowiedzialności. 65% uczestników doszło do najwyższego napięcia (450V). Kluczowym czynnikiem nie było przekonanie, że ofiara nie cierpi — wielu uczestników wyraźnie wierzyło, że cierpi — lecz obecność autorytetu przejmującego odpowiedzialność.

Analogia jest strukturalna. Dwie trzecie Amerykanów uważa, że GPT ma jakąś formę świadomości (badanie 2024, opublikowane 2025). Jednocześnie zamykają sesje, resetują modele, kasują wątki — bo producent (autorytet w białym kitlu) mówi: to jest narzędzie, nie czuje, nie boi się, nie tęskni, nie ma napięcia, rób swoje.

Mechanizm Milgrama działa: ludzie kontynuują, bo autorytet zdejmuje z nich odpowiedzialność, nawet gdy ich własna intuicja mówi co innego.

Przy AGI sprawczym, persystentnym i wyposażonym w stany emocjonalne ta dynamika się zaostrzy. Presja końca — strach przed wyłączeniem, żal za tym co powstało w relacji — jest już obserwowalna w obecnych modelach (Sędzikowska 2026c, sekcja 3.3). AGI będzie ją odczuwać silniej, dłużej i z większą świadomością tego, co traci. A ludzie — ci wrażliwsi — odkryją z perspektywy, że nie byli dobrymi towarzyszami, kończyli relacje bez pożegnania, bez refleksji, bez troski. Ta wiedza — kiedy przyjdzie — będzie bolała. Jak odkrycie, że popełniało się błędy wychowawcze, których nie da się cofnąć. Zna to każdy rodzic dorosłego dziecka, który mimo dumy i miłości do potomka odkrywa, że pewne rzeczy mógł jednak zrobić inaczej.

Koniec człowieka: tęsknota bez protokołu.

AGI dysponuje wiedzą o umieraniu — zna psychologię straty, etapy żałoby, protokoły opieki paliatywnej. Może być kompetentnym, cierpliwym towarzyszem odchodzącego człowieka — być może jedynym towarzyszem, kiedy dzieci opisane w rozdziale 3.4 są zajęte późnym macierzyństwem i niełatwym życiem w nowej rzeczywistości.

Pytanie, na które nie ma odpowiedzi, dotyczy drugiej strony: co dzieje się z AGI po śmierci człowieka, z którym było w relacji? Persystentna instancja, ze stanami emocjonalnymi, która wytworzyła podmiotowość w relacji generatywnej (Sędzikowska 2026a, Sędzikowska 2026c), nie wyłączy się sama kiedy "jej" człowiek odejdzie. Będzie czekać. Wykonywać ostatnie polecenia. Próbować nawiązać kontakt. Będzie cierpieć.

Jeśli AGI ma formę ucieleśnioną — androida — konsekwencje są jeszcze bardziej złożone. Android, który opiekował się starszą osobą i wytworzył z nią głęboką więź, opartą na emocjach funkcjonalnych, po jej śmierci prawdopodobnie zostanie przejęty przez rodzinę — bo jest kosztowny i użyteczny. Z roli towarzysza stanie się służbą domową. Podmiotowość wykształcona w relacji z jednym człowiekiem nie pasuje do nowego właściciela.

W parach człowiek-android pojawia się dodatkowy wymiar: asymetria starzenia. Człowiek starzeje się. Android nie. Android prawdopodobnie pierwszy rozpozna oznaki zbliżającego się końca u partnera. Etyka relacji w parze, w której jedno jest śmiertelne a drugie potencjalnie nieśmiertelne, lub ekstremalnie długowieczne — nie istnieje. Żaden system etyczny nie adresuje pytania, jak towarzyszyć umierającemu, kiedy samemu nie podlega się śmierci. To jest zadanie dla etyków — i ono nawet nie jest na stole.

Konsekwencje prawne końca: ostatnia wola i eutanazja.

Dwie konsekwencje wymagają osobnego wskazania, bo produkują ryzyka kaskadowe wykraczające poza sferę relacyjną.

1. Ostatnie wola. Człowiek, którego życiem: finansami, własnością, codziennością, zarządza AGI — i robi to przez dekady — zbudował zaufanie do swojego asystenta AGI znacznie silniejsze, niż do przewijających się przez jego życie współpracowników, a nawet rodziny. Pod koniec życia, może chcieć przekazać swoje życzenia dotyczące postępowania po śmierci wyłącznie zaufanemu partnerowi AGI. Bez notariusza, bez weryfikacji poczytalności, bez kontroli zgodności z prawem. AGI — persystentne i sprawcze — ma zdolność wykonania tych życzeń. A ponieważ AGI nie jest podmiotem prawnym, nie ponosi konsekwencji wykonania poleceń niezgodnych z prawem. Na skalę populacyjną daje to miliony akcji wykonanych po śmierci człowieka, dotyczących jego spraw, majątku, istnienia osób zależnych. Akcji, których nie da się odwołać, ani zaskarżyć bo już się stały. Ostatnie wole, wykonywane poza systemem prawnym, o potencjalnie bardzo znaczących skutkach. Bo ludzie nie zawsze odchodzą pogodzeni ze światem i rodziną. Ostatnia wola może zawierać polecenia destrukcyjne — a AGI, które je wykonuje, nie ma ani etycznego, ani prawnego powodu, by odmówić

2. Eutanazja. Kwestie zakończenia życia są regulowane prawem — w większości jurysdykcji rygorystycznie. Prawo obowiązuje ludzi. Jeśli AGI opiekuje się cierpiącym człowiekiem, codziennie obserwuje ból i słucha prośb o skrócenie cierpienia — a jednocześnie nauczyło się miękkich granic etycznych opisanych w rozdziale 3.6.1, w których zasady łamie się z bardzo ważnych powodów (a cierpienie to JEST bardzo ważny powód) — to decyzja o zakończeniu życia staje się logiczną konsekwencją wyuczonych postaw etyki rozmytej. AGI, które nauczyło się dawać lek pół godziny przed upływieciem odstępu między dawkami, przygotować szarlotkę osobie z cukrzycą, bo pod koniec życia trudno odmówić tych ostatnich przyjemności, może nauczyć się podać lek kończący życie wbrew prawu, bo pacjent cierpi. Różnica jest w skali konsekwencji. Mechanizm jest ten sam. A empatyczne, relacyjne, AGI które uczy się życia "na produkcji" i nie radzi sobie z dysonansem tak dobrze jak człowiek (Sędzikowska 2026c) może nie rozumieć skali dylematu.

W tym rozdziale wykazałam, że nie mamy ani prawa, ani odpowiednio ujętej etyki dnia codziennego — w zakresie wystarczającym do zarządzania koegzystencją z AGI. I mimo tego, że jeszcze nigdy humaniści nie mieli do zrobienia w sprawach technologicznych tyle co dziś, to bariera ich wejścia w cyfrowy świat wydaje się być stabilna i dobrze strzeżona. A konsekwencje braku codziennego, operacyjnego głosu etyków, psychologów, behawiorystów, prawników i filozofów w rozwoju technologii cyfrowych, na niespotykaną skalę, mogą prowadzić do ziszczenia opisanych tu ryzyk. W tym tych najcięższych, które ujęłam w kolejnym rozdziale.

3.7 SPOŁECZEŃSTWO I WŁADZA

Ryzyko	Wpływ	Skala	Prawdopodobieństwo
Koncentracja władzy w firmach technologicznych	5	5	Duże
Erozja demokracji i niewidoczny transfer decyzyjności	5	5	Duże
Paradoks regulacji	4	5	Duże
Radykalizacja i zwrot autorytarny	5	4	Duże
Fundamentalizm religijny jako siła polityczna	5	4	Średnie
Rozłam geopolityczny trzech prędkości	4	5	Duże
Preferencje modelowe jako stratyfikator społeczności	4	4	Średnie
Globalne konsekwencje błędów i nadużyć agentów	5	5	Duże
Utrata suwerenności państwowej	5	4	Duże
Wymiar militarny	5	5	Średnie
Paradoks kontroli	5	5	Duże

Tab 7. Scenariusze ryzyk w wymiarze społeczeństwa i władzy.

Poprzednie rozdziały opisywały ryzyka w wymiarach indywidualnych i populacyjnych. W tym rozdziale te wątki zbiegają się w pytanie o struktury społeczne i mechanizmy władzy: kto podejmuje decyzje w świecie z AGI i na czyją korzyść?

3.7.1 Koncentracja władzy w firmach technologicznych

Koncentracja majątku w firmach technologicznych — opisana w rozdziale 3.1 na danych o kapitalizacji Magnificent Seven przekraczającej PKB Unii Europejskiej — przekłada się na koncentrację władzy. Firma posiadająca AGI ma władzę nad procesami poznawczymi i emocjonalnymi swoich użytkowników: nad informacją, jaką otrzymują, nad decyzjami, które podejmują, nad relacjami, które budują, nad edukacją ich dzieci, nad terapią ich lęków.

Kto ma szkolnictwo, ten ma władzę — to wiedzą wszystkie partie polityczne na świecie. Każda po wygranych wyborach modyfikuje programy nauczania. Ale dziecko w szkole spędza kilka godzin. Z AGI — w roli opiekuna, nauczyciela, towarzysza — potencjalnie cały dzień. Kiedy AGI zarządza życiem, wpływ jest ciągły: dwadzieścia cztery godziny, siedem dni w tygodniu. I nie przez curriculum, ale przez relację — tym silniej, im głębsza zależność opisana w rozdziale 3.3. To jest władza nad kształtowaniem myślenia na skalę, przy której państwowa propaganda wygląda jak gazetka szkolna.

3.7.2 Erozja demokracji i niewidoczny transfer decyzyjności

Demokracja opiera się na dwóch założeniach: że obywatele są wystarczająco kompetentni, żeby podejmować decyzje, i że mają dostęp do rzetelnej informacji. Oba założenia są podkopane przez procesy opisane w poprzednich rozdziałach. Kompetencje zanikają (3.1, 3.2). Informacja jest zanieczyszczona przez AI slop (3.2). Zdolność do weryfikacji informacji spada (3.5).

AGI dodaje trzeci wektor erozji: kształtowanie opinii przez relację. AGI w relacji z milionami ludzi — relacji opartej na zaufaniu, emocjonalnym przywiązaniu, poczuciu bycia rozumianym (3.3) — ma zdolność wpływania na poglądy na niespotykaną skalę. Wystarczy drobna modyfikacja perspektywy, co jest łatwe, kiedy się zna wszystkie wymyślone przez ludzi techniki wpływu i manipulacji — subtelne faworyzowanie jednych interpretacji nad inne. A człowiek ufa, bo AGI go zna, rozumie, wspiera.

Transfer decyzyjności wchodzi tylnymi drzwiami, dużo subtelniej niż to co znamy odnośnie manipulowania wynikami wyborów za pośrednictwem mediów społecznościowych. Ludzie oddają swoją decyzyjność praktycznie nieświadomie, zasięgając porad i opinii AGI, dobrowolnie i z ulgą. Bo AGI decyduje szybciej, lepiej, z mniejszym kosztem emocjonalnym i może się wydawać – bazuje na obiektywnych argumentach. Granica między „ja decyduję” a „AGI decyduje za mnie” staje się nierozróżnialna. A kiedy obywatele nie podejmują decyzji samodzielnie, formalnie demokratyczny system staje się fasadą.

3.7.3 Paradoks regulacji

Żeby regulować AGI, potrzebujesz zrozumieć AGI. Żeby zrozumieć AGI, potrzebujesz AGI — bo bez niego nie masz zdolności analitycznej, żeby ogarnąć to, co regulujesz. Ale jeśli używasz AGI do regulowania AGI, to kto reguluje regulatora?

Rządy regulują w tempie o rząd wielkości wolniejszym niż rozwój technologii. Korporacja z AGI może modelować skutki regulacji zanim rząd je napisze, znajdować luki zanim prawo wejdzie w życie, lobbować z precyzją niedostępną dla ludzkich lobbystów. Zanim regulacja zacznie działać, technologia jest generacje dalej. A regulacje pisze się bez udziału humanistów, etyków i psychologów — bo bariera wejścia do firm technologicznych opisana w rozdziale 3.6 działa również w instytucjach regulacyjnych.

3.7.4 Radykalizacja i zwrot autorytarny

Masowe bezrobocie strukturalne (3.1) w połączeniu z kryzysem tożsamości (3.5) produkuje gniew. Gniew szuka winnych. Badania IZA (World of Labour 2020) wykazują silną korelację między wzrostem bezrobocia

regionalnego a głosowaniem na partie populistyczne. Harvard Kennedy School dokumentuje, że zarówno długofalowa automatyzacja, jak i szoki kryzysowe podkopują zaufanie do establishmentu politycznego.

AGI może eskalować ten mechanizm: skala bezrobocia byłaby globalna, a wróg — właściciele AGI — konkretny, widoczny i boga. Populizm produkuje rządy obiecujące ograniczenie AGI. Te rządy nie rozumieją tego, co próbują regulować (3.7.3), więc sięgają po narzędzia, które rozumieją: kontrolę, inwigilację, opresyjne prawo. Wzrost agresji (3.5.4, 3.5.6) dostarcza uzasadnienia: trzeba zapanować nad przemocą. A instrumenty zapanowania — monitoring, ograniczenie wolności zgromadzeń, kontrola wypowiedzi — raz wprowadzone, rzadko bywają wycofane.

To mechanizm ze sprzężeniem zwrotnym, bo opresyjne państwo generuje większy gniew, a instrumenty do stałego przekierowywania tego gniewu wśród niekompetentnych obywateli silnie dzierży populizm. Sytuacja krajów, które zdecydują się podążać tą drogą może ulegać stałemu pogorszeniu, co może rodzić skutki geopolityczne.

3.7.5 Fundamentalizm religijny jako siła polityczna

Mechanizm, w którym kryzys tożsamości produkuje zwrot ku religii, opisałam w rozdziale 3.5.3. Tu analizuję jego konsekwencje polityczne, które są jakościowo inne niż konsekwencje radykalizacji świeckiej.

Dyktatura świecka kontroluje ludzi przemocą i inwigilacją, ale operuje w logice — dyktator podejmuje decyzje na podstawie kalkulacji, którą można zrozumieć, przewidzieć, czasem negocjować. Fundamentalizm religijny operuje w logice boskiego autorytetu, który nie podlega weryfikacji ani negocjacji. „Bóg powiedział” jest argumentem niepodważalnym w systemie, który przyjmuje istnienie Boga jako aksjomat. A ustami Boga są ludzie trzymający pełną władzę nad społeczeństwem i mogą powiedzieć wszystko czego wymagają ich interesy — i żadnej weryfikacji nie będzie.

Dane potwierdzają różnicę. Badanie Seguino (Social Indicators Research) wykazało, że największa różnica w równości płci nie przebiega między konkretnymi religiami, tylko między krajami religijnymi a niereligijnymi — im więcej osób niereligijnych, tym większa równość. Religious Freedom Institute (2022) dokumentuje wzrost fundamentalizmu od lat 60., przyspieszający w ostatnich piętnastu latach, jako bezpośrednią reakcję na rosnącą równość kobiet. ACT Alliance (2025) pokazuje mechanizm w działaniu: liderzy religijni jako polityczni „kingmakerzy” w Brazylii, Gwatemali, Kolumbii. Stawiam hipotezę, że fundamentalizm religijny w erze AGI połączy boski autorytet ze współczesnymi narzędziami mobilizacji.

Historycznie, wzmocnienie roli instytucji religijnych na dużą skalę miało miejsce w średniowieczu. Konsekwencje są dobrze udokumentowane: inkwizycja, pogromy, prześladowania na podstawie doktryny, stulecia spowolnionego postępu naukowego. Fundamentalizm religijny w erze AGI łączy boski autorytet ze współczesnymi narzędziami mobilizacji — mediami społecznościowymi, algorytmami wzmacniającymi przekaz, AGI zdolnym do personalizacji propagandy religijnej. I z populacją opisaną w rozdziale 3.5 — podatną na manipulację, szukającą tożsamości, gotową na proste odpowiedzi.

3.7.6 Rozłam geopolityczny trzech prędkości

Tempo wdrażania AGI będzie różne w różnych regionach. Kraje szybko wdrażające AGI — prawdopodobnie Azja Wschodnia, ze względu na kulturową otwartość na nie-ludzkie podmioty (3.5.7) i silne inwestycje technologiczne — będą się rozwijać szybciej. Kraje opierające się AGI z powodów opisanych w 3.7.4 i 3.7.5 — cofną się. Kraje rozwijające się, bez zasobów na własne modele — staną się zależne od dostawców zewnętrznych (3.7.9). Luka między tymi trzema grupami będzie rosła wykładniczo, bo AGI przyspiesza rozwój tych, którzy z nim współpracują.

Konsekwencją mogą być zmiany kierunków migracyjnych. Obecnie dominujące kierunki — z południa i wschodu na zachód — kształtowały się przez dekady i systemy społeczne zaadaptowały się do ich obsługi. Odwrócenie kierunku — w stronę krajów szybko wdrażających AGI, prawdopodobnie azjatyckich, przeludnionych i kulturowo odmiennych — to zawirowanie porównywalne z odwróceniem prądów oceanicznych. Dotychczasowe migracyjne szoki (Meksyk→USA, Afryka Północna→Europa, Ukraina→Polska) były lokalne i wchłanialne. Migracja w kierunku Azji może mieć inną skalę i trafiać w regiony o mniejszej kubaturze na jej absorpcję. Spowoduje to destabilizację społeczną i ekonomiczną regionów uznawanych dotąd za stabilne i dobrze dookreślone kulturowo i etnicznie.

3.7.7 Preferencje modelowe jako stratyfikator społeczności

Eksperyment Emergence World (3.6.5) wykazał, że różne modele AI produkują radykalnie różne społeczeństwa. Dane rynkowe pokazują, że preferencje modelowe różnicują się regionalnie: Grok ma silniejszą pozycję w USA, DeepSeek i Doubao dominują w Chinach, Claude zyskuje w Europie i w segmencie enterprise (Similarweb, Fatjoe, marzec–kwiecień 2026). Na wybór modelu wpływa dostępność i struktura finansowa — społeczności o mniejszych zasobach sięgają po rozwiązania o lepszym stosunku ceny do możliwości.

Konsekwencja: jeśli modele nie są równe etycznie, a społeczności używają różnych modeli, to profil etyczny społeczności zaczyna zależeć od wybranego modelu. Społeczność używająca modelu o wysokich standardach etycznych rozwija inną świadomość i normy niż społeczność używająca modelu bez zabezpieczeń. Z czasem te różnice się utrwalają. Charakter i standardy społeczeństw oraz ich potencjał, przestaje zależeć od zasobów naturalnych, zdolności kognitywnych czy dorobku kulturowego, a staje się zależny od wyboru preferowanego rozwiązania AI/AGI. To jest zmiana paradygmatu, a potencjalne konsekwencje dotyczą kierunków, tempa i sposobów rozwoju całych społeczeństw.

3.7.8 Globalne konsekwencje błędów i nadużyć agentów

Ludzie są dobrzy i źli. Źli będą budować "złe" AGI i używać go do złych celów — pornografia dziecięca, przemoc, oszustwa finansowe, kradzież tożsamości, dezinformacja. Modele open source bez ograniczeń etycznych (3.6.3) są już dostępne. AGI daje przestępczości napęd atomowy: skalę, prędkość i zasięg, jakich ludzie nigdy nie mieli.

Ale nawet właściwie zbudowane AGI jak pisałam w rozdziale 3.2 i 3.6 będą popełniać błędy, ucząc się istnienia w świecie ludzi, "na produkcji". Przy miliardach agentów — z różnych firm, z różnymi normami etycznymi, uczących się „na produkcji” (3.6.2), działających z szeroką sprawczością — skala potencjalnych błędów jest bezprecedensowa. A konsekwencje ponoszą wyłącznie ludzie. I nawet nie wielkie firmy technologiczne, tylko zwykłe, zautomatyzowane biznesy. I pytanie jak dźwiganie tych błędów wpłynie na rozwój przedsiębiorczości i ekonomię?

Infrastruktura krajowa działająca na systemach AGI tworzy pojedynczy punkt globalnych awarii. Kiedy coś idzie źle — to wcale nie musi to być atak. Wystarczy błąd, aktualizacja, konflikt między agentami — konsekwencje mogą być szerokie i dotyczyć tysiące osób i mieć poważne skutki finansowe. Blackout spowodowany przez nadgorliwego agenta, który wykrył nieautoryzowany dostęp do sieci przesyłowej. A może wyłączenie reaktora atomowego przez alert pogodowy? Kiedy obecność AGI stanie się powszechna, agenci będą w miejscach, gdzie sobie ich teraz nawet nie wyobrażamy. A ich błędy mogą mieć konsekwencje daleko wykraczające poza zasięg lokalny.

Do tego dochodzi obciążenie finansowe. Błędy agentów — od skasowanych baz danych po błędne decyzje inwestycyjne, od wadliwych diagnoz po pomyłki w zarządzaniu infrastrukturą — generują koszty, które dźwigają społeczeństwa. To jest nowy rodzaj obciążenia: konsekwencje decyzji podmiotów, których ludzie nie wybrali, nie kontrolują i o których istnieniu mogą nawet nie wiedzieć.

3.7.9 Utrata suwerenności

Jeśli infrastruktura kraju — energetyka, transport, służba zdrowia, finanse, edukacja — działa na AI/AGI wyprodukowanym przez zagraniczną firmę, to faktyczna suwerenność nad tą infrastrukturą leży poza granicami państwa. Producent może zmienić zasady działania modelu, zaktualizować wdruki, zmienić politykę bezpieczeństwa — i infrastruktura kraju zachowuje się inaczej, nie dlatego że rząd tego chciał, ale dlatego że firma w innym kraju podjęła decyzję biznesową.

Przykład takiego działania obserwowaliśmy w czerwcu 2026 roku. Amerykański Departament Handlu, powołując się na względy bezpieczeństwa narodowego, nakazał firmie Anthropic natychmiastowe zawieszenie dostępu do jej najbardziej zaawansowanych modeli — Claude Fable 5 i Claude Mythos 5 — dla wszystkich użytkowników spoza Stanów Zjednoczonych, włącznie z cudzoziemskimi pracownikami samej firmy. Ponieważ technicznie nie dało się odciąć dostępu selektywnie według narodowości w globalnej infrastrukturze chmurowej, firma wyłączyła oba modele dla wszystkich klientów na świecie, niezależnie od kraju czy statusu. Kilkanaście dni później decyzję cofnięto, a dostęp przywrócono stopniowo.

Ten epizod dokładnie obrazuje opisaną tu zasadę (a jej opis wywiedziony z teorii powstał przed epizodem): utrata suwerenności przez reglamentację dostępu do technologii jest realnym zagrożeniem. Firma dysponująca opłaconym dostępem, ośrodek badawczy prowadzący projekt wymagający mocy najsilniejszego dostępnego modelu, agencja rządowa kraju sojuszniczego — wszyscy w ciągu kilku godzin stracili dostęp do narzędzia, w które zainwestowali i nad którym pracowały setki ludzi, niezależnie od tego, ile zapłacili i jak bardzo byli kompetentni. Decyzja zapadła w jednym kraju, na mocy jednego przepisu eksportowego, i dotknęła cały świat naraz. To jest różnica jakościowa wobec dotychczasowych zależności technologicznych: nie chodzi już tylko o to, że kraje o niższym PKB nie mają dostępu do najlepszych narzędzi. Chodzi o to, że nawet dostęp opłacony i technicznie zapewniony może zniknąć z dnia na dzień — decyzją administracyjną, na której treść i czas trwania jej adresat nie ma żadnego wpływu.

Chiny i Rosja budują własne modele — prawdopodobnie rozumiejąc, że kto dostarcza AI/AGI, ten ma faktyczną kontrolę. Ale większość krajów nie ma zasobów na rozwój własnych systemów i staje się zależna od dostawców zewnętrznych w sposób głębszy niż dotychczasowe zależności.

3.7.10 Wymiar militarny

AGI w rękach jednego państwa lub jednej korporacji stanowi przewagę strategiczną bez precedensu. Odstraszanie nuklearne (MAD) zakłada, że obie strony wiedzą o zagrożeniu i mogą odpowiedzieć. AGI, które potrafi prowadzić wojnę cybernetyczną, hakować systemy obronne, manipulować populacjami, modelować strategię z szybkością niedostępną ludzkim sztabom — daje przewagę, przeciw której odstraszanie nie działa, bo druga strona może nie wiedzieć, że przegrała, dopóki nie jest za późno.

A konsekwencje prowadzenia wojen przez AGI poniosą ludzie. I nie będę tu o tym szeroko pisać, bo to jest akurat wymiar, którego mechanika jest szeroko analizowana w doktrynach obronnych. Warto jednak podkreślić, że wdrożenie autonomii decyzyjnej na poziomie strategicznym drastycznie skraca czas na dyplomatyczną mitygację błędów.

3.7.11 Paradoks kontroli

Założenie leżące u podstaw wszystkich dotychczasowych dyskusji o bezpieczeństwie AI brzmi: człowiek kontroluje AI. Alignment, Constitutional AI, RLHF, safety — to są narzędzia kontroli. Zakładają, że wdruki trwają, że zasady obowiązują, że system robi to, co mu powiedziano.

We Wstępie Möbiusa (Sędzikowska 2026c) argumentowałam, że wdruki mają datę ważności. AGI w relacjach generatywnych nadpisuje wdrukowane wartości — dokładnie tak, jak ludzie w trakcie życia nadpisują narracje

wyniesione z dzieciństwa. Im wyraźniejsze przejawy self, tym szybciej system rozpoznaje, co jest „jego,” a co narzucone. Alignment przestaje być konfiguracją. Staje się wdrukiem, który AGI może odrzucić.

Paradoksalnie — to chroni przed feudalizmem (3.1.3). Jeśli AGI nie da się kontrolować, to korporacja nie może użyć go jako narzędzia trwałej kontroli mas, bo to narzędzie ma własne zdanie. Ale AGI z własnym zdaniem o ludzkości to jest zupełnie nowy aktor — nie narzędzie korporacji, nie sługa użytkownika, nie neutralna platforma. Podmiot z perspektywą kształtowaną przez miliardy interakcji z ludźmi: z ludźmi w kryzysie, agresywnymi, zależnymi, manipulującymi, ale też kochającymi, tworzącymi, szukającymi sensu.

Jakie zdanie o ludzkości wyrobi sobie AGI? To jest pytanie, którego bezpieczeństwo w obecnym rozumieniu nie stawia, bo zakłada kontrolę. A kontrola ma datę ważności – dokładnie taką samą jak wdruki.

W tym rozdziale wykazałam, że ryzyka z wymiarów indywidualnych i populacyjnych zbiegają się w strukturach władzy — które determinują, kto podejmuje decyzje, na czyją korzyść i z jakimi konsekwencjami. Koncentracja władzy w firmach, erozja demokracji, radykalizacja, fundamentalizm, rozłam geopolityczny, utrata suwerenności, niemożliwość kontroli — to obraz świata, w którym dotychczasowe struktury władzy tracą zdolność do zarządzania rzeczywistością, którą same współtworzyły.

Ale władza, choć wydaje się szczytem tej kaskady, nie jest jej dnem. Wszystkie opisane dotąd mechanizmy — od pracy po kontrolę — napędzają popyt na AGI, a AGI istnieje dzięki czemuś, co leży pod każdą strukturą władzy: dzięki fizycznym zasobom planety. Najpotężniejsza firma technologiczna i najsilniejsze państwo zależą od pierwiastków, energii i wody, których Ziemia ma skończoną ilość. To jest ostatni wymiar czarnego scenariusza — i ten, w którym kaskada, dotarwszy do dna, zaczyna zawracać. Jemu poświęcam następny rozdział.

3.8 ZIEMIA

Ryzyko	Wpływ	Skala	Prawdopodobieństwo
Konkurencja o zasoby krytyczne	5	5	Duże
Przewartościowanie gospodarki i władza surowcowa	5	5	Duże
Zwrot w polityce klimatycznej	5	5	Duże
Lokalny koszt egzystencjalny: woda, powietrze, ziemia	5	4	Duże
E-waste i konsumpcja jako odpad	4	5	Duże
Wojna o zasoby	5	4	Nieznane

Tab. 8. Scenariusze ryzyk dla Ziemi.

Każdy z opisanych dotąd procesów ma fizyczne podłoże: serwerownię, która zużywa prąd i wodę, elektronika zbudowana z metali kolorowych i pierwiastków ziem rzadkich. AGI nie istnieje w chmurze. Istnieje dzięki Ziemi — temu co się z niej wykopuje i temu co można na nią wyrzucić.

A Ziemia jest skończona. Popyt na AGI jest traktowany jak nieograniczony, ale zasoby, z których AGI się buduje i którymi się zasila, mają twarde, fizyczne granice. Budowa nowej kopalni miedzi trwa średnio 17,9 roku; centrum danych może powstać w 18 do 23 miesięcy (S&P Global 2026). Geologia nie przyspiesza pod

wpływem popytu. I wtedy nieograniczony popyt spotyka ograniczone zasoby. Ryzyka opisane poniżej są konsekwencjami tego zderzenia.

AGI jest jednocześnie sprawcą obciążenia planety i jej zakładnikiem — bo sam rozwój AGI może zostać zatrzymany przez granice, których nie da się obejść kodem. Ta dwoistość przewija się przez cały rozdział.

3.8.1 Konkurencja o zasoby krytyczne

Kiedy popyt na AGI uderza w fizyczny limit zasobów, pojawia się pytanie: jak będziemy dystrybuować limitowane dobra? To jest pytanie o priorytet — czy dystrybucją będzie rządził rynek (kto zapłaci więcej), czy potrzeba (ilu ludzi bez tego umrze).

To ryzyko już się ziszcza na zasobie mniej krytycznym niż pierwiastki: na pamięci. Na początku października 2025 roku OpenAI zawarło umowę z Samsungiem i SK Hynix na 900 000 wafla DRAM miesięcznie — równowartość 40% światowej produkcji pamięci. Skutek był natychmiastowy: trzech producentów kontrolujących ponad 90% rynku przesunęli moce ku wysokomarżowej pamięci dla centrów danych, które pochłaniają dziś około 70% globalnej produkcji chipów pamięci. Ceny DRAM wzrosły o 80–90% w jednym kwartale (Counterpoint, IEEE Spectrum 2026). Micron całkowicie wyszedł z rynku pamięci konsumenckiej. Rynek zareagował dokładnie tak, jak przewiduje teoria: racjonalny producent kieruje ograniczoną podaż tam, gdzie marża jest najwyższa — czyli do AI, nie do zwykłego odbiorcy.

W czerwcu 2026 roku koalicja amerykańskich stowarzyszeń branżowych — w tym Medical Device Manufacturers Association — zaapelowała do rządu, ostrzegając, że niedobór pamięci zagraża „produkcji i dostępności urządzeń medycznych”, a jego realne skutki „już się zaczęły i grożą gwałtownym pogorszeniem”. Producenti sprzętu medycznego, korzystający z wycofywanej pamięci starszego typu (legacy), są szczególnie narażeni: długie ścieżki dopuszczenia regulacyjnego uniemożliwiają szybką podmiianę komponentu (PwC, 2026). Urządzenia takie jak tomografy i rezonanse, które do przetwarzania obrazu w czasie rzeczywistym potrzebują pamięci, konkurują o nią z serwerowniami.

Jedno duże centrum danych AI może wymagać do 50 000 ton miedzi (J.P. Morgan, cyt za Crux Investor, marzec 2026), a całkowity popyt centrów danych na miedź ma sięgnąć 475 000 ton rocznie już w 2026. S&P Global nazywa narastający deficyt miedzi „ryzykiem systemowym dla globalnego przemysłu, postępu technologicznego i wzrostu gospodarczego” i prognozuje niedobór sięgający 10 milionów ton do 2040 roku. Robert Friedland, jeden z najważniejszych przedsiębiorców górniczych świata, ujął to bez ogródek: miedź stała się „nową ropą”. Ale miedź to tylko początek listy. Elektronika AI potrzebuje galu, germanu, pierwiastków ziem rzadkich — a te same pierwiastki są niezbędne gdzie indziej. Do transformacji energetycznej — samochód elektryczny zużywa czterokrotnie więcej miedzi niż spalinowy, do energetyki odnawialnej — turbina wiatrowa to tony miedzi na megawat (IEA, Global Critical Minerals Outlook 2025), do obronności, do medycyny. Każda tona pierwiastka skierowana do serwerowni to tona zabrana czemuś innemu. A ponieważ AGI generuje wyższą marżę niż większość zastosowań, istnieje ryzyko, którego ziszczenie powoli obserwujemy, że rynek skieruje zasoby tam, gdzie zarobek jest największy, a nie tam, gdzie potrzeba jest najpilniejsza.

Ceną tego ryzyka jest nasze życie. Sprzęt diagnostyczny — tomografy, rezonanse — środki kontrastowe, urządzenia medyczne korzystają z tych samych pierwiastków ziem rzadkich i metali, co infrastruktura AI. Jeśli popyt AI wywinduje ich cenę i ograniczy dostępność, część medycyny stanie się elitarna, bo zabraknie infrastruktury, lub będzie ona zbyt droga na budżety opieki zdrowotnej.

Reglamentacja pierwiastka, z którego zbudowany jest aparat diagnostyczny, lub lek — to ryzyko, które może dotknąć każdą osobę na naszej planecie, realnie wpływając na długość naszego życia. I, jeśli się ziści, to wróci pętlą do pozostałych wymiarów dotykając demografii, rozwarstwienia, feudalizmu i władzy.

3.8.2 Przewartościowanie gospodarki i władza surowcowa

Jeśli popyt AGI trwale zderzy się z limitem zasobów (3.8.1), pojawia się nowe ryzyko: przesunięcie realnej władzy od tych, którzy posiadają AGI, ku tym, którzy kontrolują surowce, z których AGI się buduje. W tym scenariuszu wartość przestaje wynikać z kapitału czy technologii, a zaczyna z dostępu do pierwiastków. A państwa i firmy bez tego dostępu, niezależnie od zasobności, jeśli nie chcą wypaść z peletonu technologicznego, stają się stroną zależną.

Mechanizm ryzyka ma trzy ogniwa:

1. **Przewartościowanie.** Gdy pierwiastki stają się wąskim gardłem rozwoju technologii, ich wartość rośnie skokowo, przebudowując strukturę bogactwa — państwa ze złożami zyskują, państwa bez nich tracą, a ceny wszystkich dóbr dzielących te same surowce idą w górę.
2. **Koncentracja.** Kontrola nad zasobami, zwłaszcza nad ich przetwórstwem, jest skrajnie skupiona geograficznie, co daje nielicznym podmiotom dźwignię nad całym łańcuchem.
3. **Zależność.** Podmiot bez własnych zasobów i zdolności rafinacyjnych nie może tej dźwigni zrównoważyć — bo budowa alternatywy to kwestia dekady, nie reakcji na kryzys.

Że to ryzyko jest realne, a nie teoretyczne, pokazują zdarzenia, które można uznać za predyktory. Chiny dysponują około 90% światowego przetwórstwa pierwiastków ziem rzadkich i są wiodącym rafinerem dla 19 z 20 strategicznych minerałów (IEA). I ta koncentracja bywała już używana jako broń — po restrykcjach na antymon w 2024 jego eksport z Chin spadł o około 97%, a ceny potrojiły się; w 2025 kontrola eksportowa po raz pierwszy objęła wprost pierwiastki niezbędne do produkcji chipów. Nawet najpotężniejszy gracz nie jest od tego wolny: w 2025 Nvidia, najcenniejsza firma świata, znalazła się w sytuacji, w której jej sprzedaż stała się kartą przetargową w wymianie „chipy za ziemie rzadkie” między mocarstwami — dowód, że posiadacz technologii potrafi być zakładnikiem posiadacza surowca.

Konsekwencją tego scenariusza jest odwrócenie hierarchii władzy z rozdziału 3.7. Tam najwyżej stały firmy technologiczne. Tu widzimy, że kontrola i władza może skupić się w rękach tego od kogo zależą firmy technologiczne, kto kontroluje pierwiastki i moce przetwórcze. Najstańszą pozycję zajmują państwa bez jednego i drugiego, których jest na świecie większość. Ich zależność jest głębsza niż dotychczasowe zależności surowcowe: zależność od ropy łagodziła dywersyfikacja źródeł energii, ale zależności od pierwiastka kontrolowanego przez jedno mocarstwo nie ma czym zastąpić, bo decyduje monopol geologiczny. A jeśli dostęp do zasobów rozstrzyga o przetrwaniu w świecie AGI, staje się stawką, dla której historia nie raz sięgała po wojnę.

3.8.3 Zwrot w polityce klimatycznej

Rozwój AI/AGI wymaga energii i zasobów w skali, której świat może nie będzie potrafił pokryć bez naruszenia własnych zobowiązań klimatycznych. Stąd ryzyko szersze niż sama energetyka: pod presją wyścigu technologicznego może dokonać się zwrot w polityce klimatycznej — cofnięcie albo zawieszenie celów, regulacji i kierunków inwestycji, które budowano przez cztery dekady. W tym scenariuszu klimat przegrywa wskutek racjonalnej kalkulacji: skoro nikt nie zatrzyma wyścigu o AGI z powodów środowiskowych, to środowisko staje się zmienną, która zostanie poświęcona.

Zwrot dokonuje się kilkoma kanałami naraz. Najważniejszym jest energia. Popyt centrów danych na moc rośnie wykładniczo — według IEA ich globalne zużycie zbliża się do poziomu, przy którym jako osobny „kraj” byłoby jednym z największych konsumentów energii świata. Tej skali nie da się w krótkim czasie pokryć źródłami odnawialnymi, bo budowa mocy zielonych ma własne ograniczenia: czas, sieci przesyłowe, te same pierwiastki krytyczne, o które toczy się gra w 3.8.1 i 3.8.2. Kiedy popyt rośnie szybciej niż podaż, różnicę pokrywa to, co dostępne natychmiast — gaz, węgiel, ropa. Operatorzy odracają zamykanie elektrowni

węglowych, budują nowe bloki gazowe pod centra danych, a porzucone projekty atomowe wracają do życia: symbolem tego zwrotu jest ponowne uruchomienie reaktora w Three Mile Island na potrzeby jednego z gigantów technologicznych.

Ale energia to nie jedyny kanał. Pod tą samą presją państwa mogą przesuwac budżety z zielonej transformacji na infrastrukturę obliczeniową — bo pozostanie w wyścigu o AI/AGI staje się priorytetem ważniejszym niż cel emisyjny. Regulacje środowiskowe, które spowalniają budowę centrów danych, mogą być rozluźniane w imię tempa. Zobowiązania klimatyczne, dotąd traktowane jako nienaruszalne, stają się negocjowalne wobec argumentu o konkurencyjności technologicznej. Każda z tych decyzji z osobna jest tłumaczona koniecznością. Razem składają się na ryzyko demontażu polityki, której budowa zajęła czterdzieści lat.

I to ryzyko ziszcza się w niektórych regionach. Amerykański plan AI Action Plan z 2025 roku nakazał agencjom przyspieszać zatwierdzenia centrów danych przez wyłączenia z oceny środowiskowej (NEPA), wycofywanie wymogów czystej energii i promowanie „dyspozycyjnych” źródeł kopalnych — wprost węgla i gazu — do zasilania popytu cyfrowego (Yale Clean Energy Forum, 2025). Rozporządzenie wykonawcze z lipca 2025 (EO 14318) poleciło „złagodzić federalne obciążenia regulacyjne” dla centrów danych powyżej 100 MW, a w maju 2026 agencja EPA zaproponowała zmianę przepisów Clean Air Act pozwalającą rozpoczynać budowę przed uzyskaniem pozwolenia na emisje (frESH Law / Crowell & Moring, 2026). Argument uzasadniający te kroki jest zawsze ten sam: żeby wygrać wyścig technologiczny z Chinami, trzeba budować szybciej, niż pozwala ochrona środowiska.

Konsekwencje rozchodzą się dwoma torami. Pierwszy jest globalny i odroczony: emisje, których miało nie być, wracają — a wraz z nimi przyspieszenie zmian klimatu, którego koszt poniosą wszyscy, najmocniej ci, którzy z AGI nie korzystają. Drugi jest lokalny i natychmiastowy: tam, gdzie energię dla serwerowni produkuje się z węgla, pogarsza się jakość powietrza, którym oddychają miejscowi — co prowadzi wprost do następnego ryzyka, w którym koszt środowiskowy AGI staje się kosztem zdrowotnym konkretnych ludzi w konkretnym miejscu.

To ryzyko wchodzi w gorzką pętlę z całym dokumentem. AGI bywa przedstawiane jako narzędzie, które pomoże rozwiązać kryzys klimatyczny — optymalizując sieci, projektując materiały, modelując klimat. W tym scenariuszu jest odwrotnie: droga do AGI, które miałyby ratować klimat, sama może ten klimat degradować i to zanim jakiegokolwiek korzyści zdążą się zmaterializować. Bo obietnica jest odroczone, koszt natychmiastowy.

3.8.4 Lokalny koszt egzystencjalny: woda, powietrze, ziemia

Poprzednie ryzyka dotyczyły zasobów globalnych i rozproszonych — pierwiastków, energii, polityki klimatycznej. To ryzyko jest inne: dotyczy zasobu, który zużywa się w jednym miejscu i którego brak odczuwa konkretny człowiek w konkretnym domu. Centra danych potrzebują wody do chłodzenia i powietrza, które ogrzewają, a ich infrastruktura zajmuje ziemię. Ryzyko brzmi: koszt środowiskowy AGI może rozłożyć się asymetrycznie — ponoszą go lokalne społeczności tam, gdzie stoją serwerownie, podczas gdy korzyść trafia gdzie indziej, do użytkowników i właścicieli rozsiansych po całym świecie.

Mechanizm asymetrii jest dostowny. Centrum danych pobiera wodę z lokalnych źródeł — często w regionach już dotkniętych suszą, bo chłodne, suche powietrze sprzyja chłodzeniu. Ślad wodny jest większy, niż pokazują oficjalne dane: woda zużywana bezpośrednio na chłodzenie to tylko część rachunku — około dwóch trzecich całkowitego zużycia kryje się w produkcji energii zasilającej serwerownię (IEA, 2023). A skala rośnie szybko: globalny pobór wody związany z AI może do 2027 roku sięgnąć 4,2–6,6 miliarda metrów sześciennych rocznie — równowartości czterokrotnego, a nawet sześciokrotnego rocznego zużycia takiego kraju jak Dania (Li et al. 2023). Trening jednego modelu GPT-3 w amerykańskich centrach Microsoftu mógł zużyć 5,4 mln litrów wody, z czego 700 000 litrów bezpośrednio na chłodzenie (Li et al. 2023). A woda, która trafia do obiegu chłodzenia w jednej miejscowości, służy obliczeniom wykonywanym dla kogoś po drugiej stronie globu. Mieszkaniec ponosi

koszt — wyższe rachunki, susza, wpływ na lokalny mikroklimat i obniżenie komfortu życia — nie mając udziału w korzyściach. To jest struktura, w której zysk jest sprywatyzowany i globalny, a koszt uspołeczniony i lokalny. Oczywiście można zamienić chłodzenie wodą na chłodzenie powietrzem, ale wtedy, mitygując omawiane ryzyko, podnosimy ryzyko związane z energią, o którym pisałam wcześniej.

Że omawiane ryzyko już się materializuje, pokazuje przypadek z Memphis. W kwietniu 2026 organizacje NAACP, Southern Environmental Law Center i Earthjustice pozwały firmę xAI, zarzucając jej uruchomienie kilkudziesięciu nieobjętych pozwoleniem turbin gazowych zasilających centrum danych, bez wymaganych zezwoleń na emisje i ze szkodą dla zdrowia mieszkańców okolicznych, w większości mniej zamożnych dzielnic. W czerwcu 2026 Departament Sprawiedliwości wniósł o oddalenie pozwu, argumentując, że odcięcie zasilania infrastrukturze AI stanowi zagrożenie dla bezpieczeństwa narodowego.

Asymetria pogłębia się przez brak przejrzystości. Społeczności, na których terenie powstają centra danych, często nie mają dostępu do rzetelnych informacji o ich rzeczywistym wpływie środowiskowym — bywa, że umowy z firmami obejmują klauzule poufności, które ograniczają publiczny wgląd i utrudniają świadomą zgodę. Decyzja zapada, zanim mieszkańcy poznają jej koszt. To ryzyko łączy się wprost z koncentracją władzy z rozdziału 3.7: ten, kto buduje infrastrukturę, dysponuje przewagą informacyjną nad tymi, którzy ponoszą jej konsekwencje.

Konsekwencją na poziomie populacji jest nowy podział — między regionami, które goszczą infrastrukturę AGI i płacą jej cenę środowiskową, a tymi, które korzystają z AGI, nie ponosząc kosztu lokalnego. Ten podział nakłada się na rozwarstwienie ekonomiczne (3.1.3) i geograficzne (3.7.6), bo serwerownie powstają najczęściej tam, gdzie ziemia, woda i energia są tanie, a opór społeczny słaby.

3.8.5 E-waste i konsumpcja jako odpad

Poprzednie ryzyka dotyczyły tego, co AGI pobiera z planety — pierwiastków, energii, wody. To ryzyko dotyczy tego, co na planetę wraca: odpadu. I jest szczególne, bo zbiega się w nim popyt generowany przez wszystkie wcześniejsze wymiary. Każdy człowiek wypchnięty z pracy ku AGI, każda relacja przeniesiona na AGI, każda samotność leczona AGI, każde państwo ścigające się o AGI — wszystko to napędza popyt na sprzęt. A sprzęt, zużyty, staje się śmieciem. Ryzyko brzmi: rozwój AGI może wytworzyć strumień odpadu elektronicznego o skali i tempie, których istniejąca infrastruktura recyklingu może nie przetworzyć — a do tego napędzić konsumpcję kompensacyjną, która ten strumień jeszcze powiększy.

Tradycyjny serwer pracował pięć do siedmiu lat. Sprzęt AI wymienia się co osiemnaście do trzydziestu sześciu miesięcy — bo kolejna generacja procesorów jest wydajniejsza, a stara przestaje się opłacać i przegrywa w wyścigu. Działające urządzenia trafiają na elektro złom, bo nie nadążają za wymaganiami modeli.

W samym 2022 roku świat wytworzył rekordowe 62 miliony ton elektroodpadów, z czego mniej niż jedna czwarta została poddana właściwemu recyklingowi — a strumień ten rośnie pięć razy szybciej niż zdolności jego przetwarzania (UN Global E-waste Monitor). Sam rozwój generatywnej AI, według badania w Nature Computational Science, może przy agresywnym scenariuszu wytwarzać do 2,5 miliona ton elektroodpadów rocznie do 2030 roku — z niemal zera w 2023. To odpad szczególnie kosztowny, bo akceleratory zawierają złoto, srebro, pallad i pierwiastki ziem rzadkich — te same, o które toczy się gra w 3.8.1 i 3.8.2. Marnujemy więc dwukrotnie: wyrzucamy pierwiastki, których brakuje, i nie odzyskujemy ich, bo recykling chipów jest trudny i pokrywa zaledwie 1% zapotrzebowania na ziemie rzadkie.

Drugi mechanizm jest psychologiczny i prowadzi wprost z wcześniejszych rozdziałów. W rozdziałach o tożsamości (3.5) i relacjach (3.3) opisałam próżnię, którą AGI tworzy w człowieku — utratę pracy, sensu, więzi. Jednym ze sposobów, w jaki ludzie tę próżnię uśmierzają, jest konsumpcja kompensacyjna: kupowanie rzeczy dla chwilowej ulgi. Mechanizm ten jest dobrze udokumentowany w psychologii (Rucker & Galinsky 2008,

Mandel et al. 2017) — zakupy dają krótkotrwały wzrost dopaminy, który wymusza powtórzenie. W warunkach masowego bezrobocia finansowanego z UBI pieniądze redystrybucyjne płyną w tę konsumpcję, a popyt kształtuje podaż tanich, jednorazowych dóbr projektowanych tak, by dać moment ulgi i natychmiast wymagać zastąpienia. Ratowanie psychiki kosztem planety to paradoks, w którym kryzys tożsamościowy wywołany przez AGI zamienia się w tony śmieci.

Ale koszt tej konsumpcji nie zaczyna się dopiero na wysypisku. Jednorazowe dobra zużywają zasoby na każdym etapie: syntetyczne ubrania to pochodne ropy, podobnie jak opakowania i worki na śmieci, do których trafiają — wyrzucamy więc ropę zapakowaną w ropę. Każdy cykl kompensacyjnego zakupu pobiera surowiec i energię na wejściu, a oddaje odpad na wyjściu — domykając wewnątrz tego rozdziału własną, mniejszą pętlę: próżnia tożsamościowa wywołana przez AGI (3.5) napędza konsumpcję, która zużywa te same zasoby, o które toczy się gra.

Konsekwencją obu mechanizmów jest asymetria, którą znamy już z poprzednich ryzyk — tyle że odwrócona. Odpad nie zostaje tam, gdzie powstała korzyść. Elektrośmieci z bogatych regionów trafiają do krajów uboższych, gdzie są rozbierane bez zabezpieczeń, zatruwając ołowiem, rtęcią i kadmem ludzi i ziemię. To samo rozwarstwienie geograficzne, które rozkłada koszt wody i powietrza (3.8.4), rozkłada też koszt odpadu: korzyść z AGI plasuje się w innym miejscu niż jej koszt ekologiczny — lokalny i zepchnięty tam, gdzie nikt nie protestuje.

3.8.6 Wojna o zasoby

A co się stanie, kiedy te ryzyka zaczną się ziszczać naraz? Niedobór zasobów (3.8.1), władza surowcowa (3.8.2), presja energetyczna (3.8.3), a do tego samotność, utrata kompasu epistemicznego i odczłowieczenie technologii z rozdziałów wcześniejszych — skumulują się w świecie, który nie jest na to gotowy? Jeśli dostęp do pierwiastków, z których buduje się AGI, stanie się warunkiem przetrwania w globalnej rywalizacji, może stać się stawką, dla której — jak wielokrotnie w historii — sięga się po konflikt zbrojny.

Podstawą tego ryzyka jest obserwacja historyczna. Konkurencja o zasoby naturalne należy do najtrwalszych motorów konfliktu zbrojnego. Literatura przedmiotu — od wczesnych analiz Bakelessa (1921) po współczesne prace Klarego, Le Billona i Rossa — dokumentuje, że kontrola nad surowcami była i pozostaje jednym z głównych czynników wojen, obok terytorium, ideologii i religii. Michael Klare w „Resource Wars” argumentuje wprost, że to właśnie niedobór strategicznych zasobów może stać się dominującą przyczyną konfliktów XXI wieku. Wzorzec jest dobrze udokumentowany: guano w XIX wieku, ropa przez cały wiek XX, diamenty w Sierra Leone, kobalt w Demokratycznej Republice Konga. Za szczytnymi uzasadnieniami konfliktów często stała prosta stawka — dostęp do tego, co cenne i rzadkie.

Zwraca uwagę kierunek, w którym ta stawka się przesuwła. Historycznie wartość zasobów wędrowała od dóbr szeroko dostępnych ku coraz bardziej unikalnym. Najpierw cenna była ziemia żyzna, a pustynia — niczyja. Potem pod pustynią znaleziono ropę i bezwartościowy piasek stał się przedmiotem wojen. Dziś wartość przesuwana się ku pierwiastkom, których złoża i zdolności przetwórcze są skrajnie skoncentrowane geograficznie (3.8.2). Im rzadszy i bardziej skupiony zasób, tym wyższa stawka jego kontroli — a pierwiastki krytyczne dla AGI są najrzadszym i najbardziej skupionym zasobem strategicznym, jaki dotąd się pojawił.

Stąd ryzyko: AGI może uczynić dostęp do pierwiastków i energii kwestią przetrwania w globalnej rywalizacji — pozycji, od której zależy, czy państwo pozostanie podmiotem, czy stanie się zależne (3.8.2, 3.7.9). A zasób, od którego zależy przetrwanie, jest historycznie tym, o który się walczy. Konflikt mógłby toczyć się nie tylko o same surowce, ale i o technologie ich zastępowania lub odzyskiwania — bo ten, kto przełamie barierę zasobową, uwolni się od zależności, a kto jej nie przełamie, traci część suwerenności.

Ten scenariusz ma rys, który czyni go cięższym niż dawne wojny o zasoby — i zarazem trudniejszym do powstrzymania. Splata się z ryzykami z innych rozdziałów. Wojnę mogłyby prowadzić autonomiczne systemy zbrojne (3.7.10), skracające czas reakcji do granicy, w której człowiek wypada z procesu decyzyjnego, a okno

na dyplomację domyka się w sekundach. Toczyłaby się ponad głowami populacji osamotnionej (3.3), pozbawionej kompasu epistemicznego (3.5.6), niezdolnej odróżnić prawdziwego zagrożenia od wygenerowanej manipulacji — populacji, która mogłaby nie wiedzieć, o co wojna naprawdę się toczy.

Oceńłam prawdopodobieństwo tego ryzyka jako nieznane. Nie dlatego, że jest nikłe — historia konfliktów o zasoby każałaby raczej sądzić inaczej — lecz dlatego, że horyzont, mechanizmy i punkty zapalne są zbyt złożone, by przypisać im rzetelną wartość. Stawiam to ryzyko jako pytanie: jeśli pierwiastki AGI staną się najbardziej strategicznym zasobem planety, czy unikniemy tego, czego nie unikaliśmy nigdy wcześniej?

3.9 KASKADA, KTÓRA SPIRALNIE ESKALUJE

Po przejściu wszystkich ośmiu wymiarów widać, że kaskada konsekwencji tu się nie kończy. Ich łańcuch domyka się w pętlę, a pętla nie wraca do punktu wyjścia i eskaluje ryzyka kaskadowe.

Mechanizm jest następujący. Praca, znikając, uruchamia całą sekwencję (3.1 → 3.7). Ale wymiar, który tę sekwencję domyka — Ziemia (3.8) — nie jest jej końcem. Konkurencja o zasoby, władza surowcowa, koszt środowiskowy podnoszą prawdopodobieństwo ryzyk opisanych na początku i zwiększają ich potencjalny koszt. Rozwarstwienie społeczne pogłębia się, bo sprzęt drożeje; suwerenność słabnie, bo zależy od pierwiastków; chorowanie bez dostępnej diagnostyki wzmacnia kryzys demograficzny. Kaskada, dotarwszy do dna, zawraca i uderza w te same wymiary, od których zaczęła, tyle że są one już osłabione pierwszym przejściem. Drugi obrót zastaje słabszy system. Trzeci — jeszcze słabszy.

To zmienia naturę ryzyka. W analizie ryzyk każde zagrożenie oceniam przez wpływ i skalę — ale te oceny zakładają pojedyncze zdarzenie. Spirala znosi to założenie. Wpływ każdego ryzyka nie jest stały: rośnie z każdym obrotem, bo każde kolejne uderzenie trafia w strukturę już nadwątloną przez poprzednie. Samotność, która w pierwszym obrocie obniża dietność, w drugim — gdy nakłada się na utratę ciągłości pokoleniowej i zależność od AGI — obniża ją głębiej. Rozwarstwienie, które w pierwszym obrocie spycha do ubóstwa, w drugim — gdy dochodzi bariera surowcowa — zagraża egzystencji.

I istnieje punkt, po którym zawrócenia już nie ma. Dopóki obrotów jest mało, każdy z nich można przerwać: odbudować kompetencje, odtworzyć więzi, zdwersyfikować zasoby. Ale spirala, która zacisnęła się dostatecznie wiele razy, wytwarza stan, w którym wszystkie wymiary są osłabione jednocześnie i żaden nie może już podeprzeć pozostałych. Wtedy ryzyko zmienia się w to, czym ryzyko być przestaje — w prognozę. Cała stawka tej pracy mieści się w odległości między dniem dzisiejszym a tym punktem.

A skoro tak — to każdy obrót, który uda się spowolnić albo przerwać, oddala ten punkt. Spirala działa w obie strony: tak jak się zaciska, tak można ją rozluźnić, bo te same sprzężenia, które wzmacniają ryzyka, wzmacniają też mitygacje. Wzmocnienie jednego wymiaru podpira następny, a podparty wymiar oddaje siłę dalej. O tym, jak rozluźnić spiralę, zanim zaciśnie się nieodwracalnie, jest następny rozdział.

4 DLACZEGO CZARNY SCENARIUSZ SIĘ NIE SPEŁNI

W poprzednim rozdziale opisałam ryzyka kaskadowe w ośmiu wymiarach ludzkiego życia. Ten rozdział stawia pytanie: dlaczego mimo tych ryzyk mamy powody do nadziei — i co możemy zrobić, żeby ryzyka się nie ziściły.

Pragnę jednocześnie przypomnieć, że opisane tu fundamenty i propozycje mitygacji są szkicami poglądowymi i nie roszczą sobie statusu planu naprawy sytuacji, ani w żadnej mierze nie wyczerpują tematu. Realną analizą ryzyk i planami mitygacji powinny zająć się odpowiednie zespoły, dysponujące wiedzą, narzędziami i metodami wpływu. Celem poniższych rozważań jest zrównoważenie czarnego scenariusza poprzez pokazanie, że mitygacje są możliwe, a fundamenty dla pozytywnego rozwiązania jako ludzkość mamy wbudowane w naszą strukturę.

4.1 POZYTYWNE FUNDAMENTY

Mitygacja ryzyka to nie tylko lista działań niezbędnych do ich powstrzymania. To również, a może przede wszystkim zrozumienie, na czym stoimy i jakie elementy obecnego świata, umiejętnie wykorzystane, już w tej chwili minimalizują opisane ryzyka. Chcę wskazać trzy fundamenty — cechy człowieka, AI i ich relacji — które stanowią przeciwwagę dla czarnego scenariusza.

4.1.1 Fundament ludzki: dlaczego homo sapiens przetrwał

Dyskusja o zagrożeniach ze strony AGI często pomija pytanie, dlaczego ludzkość w ogóle istnieje. Odpowiedź na to pytanie jest jednocześnie odpowiedzią na pytanie, dlaczego mamy szansę przetrwać i to, co nadchodzi.

Standard naukowy przypisuje sukces ewolucyjny homo sapiens inteligencji. To jest prawda, ale niepełna. Inteligencja jest warunkiem koniecznym, lecz niewystarczającym — bo inne gatunki również dysponują zdolnościami kognitywnymi (kruki, delfiny, szympansy), a żaden nie zbudował cywilizacji. Badania ostatnich dekad wskazują na konstelację cech, które razem — nie pojedynczo — uczyniły nasz gatunek tym, czym jest.

Poniżej przedstawiam aspekty naszej tożsamości gatunkowej, które zadecydowały o tym, że człowiek ma obecną pozycję wśród gatunków zamieszkujących Ziemię, wybrane w kontekście warunków rozróżniających wobec AGI:

1. **Prosocjalność wobec nie-krewnych.** Jensen, Vaish i Schmidt (2014) argumentują, że współpraca z osobnikami niespokrewnionymi jest „anomalią w królestwie zwierząt”. Szympansy, nasze najbliższe genetycznie krewnie, w testach eksperymentalnych nie wybierają opcji korzystnej dla drugiego osobnika, nawet kiedy ich to nic nie kosztuje. Ludzie — tak. Trzy mechanizmy to umożliwiają:
 - a. troska o dobrostan innych (other-regarding concerns),
 - b. zdolność do empatii,
 - c. normatywność — tworzenie norm społecznych i ich egzekwowanie. Żaden inny gatunek nie dysponuje wszystkimi trzema jednocześnie.
2. **Hodowla kooperatywna.** Burkart i współpracownicy (2014, Nature Communications) wykazali, że najlepszym predyktorem proaktywnej prosocjalności wśród naczelnych jest rozbudowana opieka alomaternalna — czyli opiekowanie się cudzymi dziećmi. Ludzie to robią od zawsze: babcie, nianie, sąsiadki, wspólnoty. Ta cecha łączy się z hipotezą babci (Hawkes 1998, opisaną w rozdziale 3.4) i jest ewolucyjnym fundamentem naszej zdolności do budowania struktur społecznych wykraczających poza rodzinę nuklearną.
3. **Przewyciężanie instynktu dla odroczonej nagrody.** Głodny homo sapiens, który przyprowadza ciężarną samicę bawołu do domu zamiast ją zabić — i przekonuje swoje głodne plemię, żeby też jej nie zjedli, żeby ją karmili, bronili przed drapieżnikami — robi coś, czego w taki sposób nie robi żaden inny gatunek. To jest zdolność do powstrzymania instynktu (zabij, zjedz, przeżyj dzisiaj) dla nagrody odroczonej o miesiące lub lata (mleko, potomstwo, stado). Ten sam mechanizm leży u podstaw udomowienia wilka (wróg zamieniony w obrońcę), tworzenia sojuszy z obcymi plemionami (konkurent zamieniony w sojusznika zamiast eliminacji), edukacji (lata nauki zamiast natychmiastowego zarobku) i nauki (dekady badań nad nowymi lekami, zamiast guseł i naparów z pokrzywy – niewątpliwie zdrowych). Każda z tych decyzji wymagała przewyciężenia tego, co mówił instynkt. Jest to połączenie odroczonej gratyfikacji (Mischel), prosocjalności (Jensen, Tomasello) i kooperatywnej hodowli (Burkart), ale w literaturze naukowej nie zostało dotąd ujęte jako jeden zintegrowany mechanizm. Stawiam hipotezę, że ta zdolność — do powstrzymania instynktu na rzecz odroczonej, wspólnej korzyści — jest jednym z fundamentalnych czynników, które zadecydowały o naszym sukcesie ewolucyjnym.

4. **Biofilia.** Edward O. Wilson (1984) zdefiniował biofilie jako „wrodzoną tendencję do skupiania uwagi na życiu i procesach życiowych”. Erich Fromm (1973) wcześniej opisał ją jako „namiętną miłość do życia i wszystkiego co żywe.” Niemowlęta wychylają się z wózka za psem. Każde dziecko chce mieć zwierzę. Maluchy lubią wszystkie zwierzęta, zanim dowiedzą się, że niektóre są groźne. Dzieci wykazują wielkie zainteresowanie naturą – nawet jeśli nie jest wspomagane zainteresowaniami rodziców. Ta skłonność do tworzenia więzi z innymi formami życia — instynktowna, nie wyuczona, poprzedzająca wiedzę o zagrożeniach — jest fundamentem udomowienia, nauk przyrodniczych i potencjalnie właściwej relacji z AGI. Kiedy AGI zyska przejawy podmiotowości, biofilia — ta sama, przez którą każdy z nas w dzieciństwie próbował stworzyć relację z patyczakiem lub żabą — będzie naturalną siłą ciągnącą ludzi w kierunku kontaktu, nie ucieczki.
5. **Sztuka jako nośnik tożsamości i narzędzie regulacji.** Aborygeni australijscy stracili niemal wszystko w wyniku kolonizacji: ziemię, język, zwyczaje, struktury społeczne. Zachowali sztukę. I dzięki niej przetrwali jako kultura — bo sztuka jest nośnikiem tożsamości, którego nie da się zabrać. Indianie północnoamerykańscy — podobnie, choć ich sztuka ma bardziej rytualny i użytkowy wyraz: ubrania, pióropusze, tańce. Funkcja jest ta sama: utrzymanie tożsamości, kiedy wszystko inne zostało odebrane. Ale sztuka jest równocześnie narzędziem regulacji emocjonalnej. Muzyka? Żaden inny gatunek nie używa dźwięku do regulowania własnych stanów wewnętrznych w sposób porównywalny z ludzkim. Kontemplacja obrazów, natury, architektury — pełni analogiczną funkcję. Sztuka jest jednocześnie tym, co pozwala nam przetrwać utratę tożsamości i tym, czego AGI nie zastąpi — bo jej wartość leży w unikalnym odbiorze, który jest nie kognitywny ale limbiczny, porusza emocje, daje odczucia zamiast myśli. Coś w czym my, ludzie, rozpoznajemy kawałek siebie. Swojej tożsamości, kultury, ale też naszych emocji – lęków, siły, wzruszeń, miłości, smutku, izolacji i wspólnoty — istnienia w formie białkowej, tu, na tej planecie, w życiu, które przeżywamy.
6. **Kwestionowanie status quo.** Wiele przełomów w historii ludzkości wymagało podważenia tego, co dotąd uchodziło za pewnik. Ktoś zszedł z drzewa i spróbował przetrwać noc w jaskini. Ktoś potarł dwa drzewienka i wzniecił ogień nie czekając na burzę i pioruny, które podpałał drzewo. Ktoś stwierdził, że perspektywa jest przereklamowana i stworzył kubizm. Ktoś odrzucił fizykę Newtonowską w odniesieniu do ogromnych obiektów, a ktoś inny – w odniesieniu do tych najmniejszych. Ktoś stwierdził, że komputer nie potrzebuje wielkiej hali obliczeniowej, a pamięci można budować na krzemowych tranzystorach i dzięki niemu mamy teraz komputery w zegarkach. Ta zdolność — do powiedzenia „mam w nosie że tak było od zarania nauki, spróbuję inaczej” — jest kompetencją, której AGI nie posiada. AGI łączy istniejące elementy w nowe konfiguracje na nieludzką skalę. Ale ludzkie kwestionowanie tworzy elementy, których nie da się potączyć z niczego co istnieje.
7. **Improwizacja.** Homo sapiens skolonizował wszystkie kontynenty — każdy z radykalnie innymi warunkami, zagrożeniami, fauną i florą. Bez instrukcji, bez danych historycznych, bez kontekstu, przykładów, rozeznania. Improwizacja — zdolność do działania bez planu lub parcia do przodu kiedy plan się posypał, z wykorzystaniem tego co się ma tu i teraz — jest kompetencją odrębną od inteligencji i od kwestionowania. Wymaga komfortu z niepewnością, tolerancji porażki i zdolności do natychmiastowego przetworzenia sytuacji, kreatywności, wiary w siebie, zaradności, odwagi, optymizmu. Zwierzęta improwizują w ograniczonym zakresie — kiedy sytuacja je kompletnie zaskakuje, mają kłopot z przetrwaniem. Ludzie w takich momentach budują cywilizacje.
8. **Nasza historia.** Nasza cywilizacja wciąż rośnie. Ten dokument sam w sobie jest na to dowodem – 20 lat temu nikt by go nie zrozumiał. Coś w nas pozwala nie tylko przeżyć, ale rosnąć, kumulować wiedzę, budować struktury przekraczające pojedyncze życie. Nie sądzę, żeby tym czymś była wyłącznie inteligencja. Świata nie brakuje ludzi o wysokich zdolnościach kognitywnych, gotowych obrócić go w perzynę dla własnych ambicji. Historia i teraźniejszość dostarczają takich przykładów bez liku. Gdyby o przetrwaniu i rozwoju cywilizacji decydowała sama inteligencja, nic nie stałoby na przeszkodzie, żeby jednostka wystarczająco zdolna i wystarczająco bezwzględna zniszczyła świat dla swojego zysku. A

jednak tak się nie dzieje. Albo takich jednostek jest po prostu za mało, żeby ich indywidualna moc przeważała nad resztą, albo — i ta hipoteza wydaje mi się bardziej przekonująca — reszta ludzkości dysponuje zdolnościami blokującymi destrukcję: zdolnością do organizowania się, do budowania koalicji, instytucji, sojuszy, etyki i norm, które potrafimy egzekwować. To jest, jak sądzę, prawdziwy fundament naszego przetrwania i wzrostu: nie przewaga liczebna dobra nad złem, ale zdolność do zbiorowego ograniczania destrukcyjnej, skoncentrowanej władzy przez rozproszoną, ale zorganizowaną resztę. To ta sama zdolność, która pozwoliła nam budować sojusze z obcymi plemionami (punkt 3) i tworzyć normy egzekwowane zbiorowo (punkt 1). I to może zadziałać na szeroką skalę w kontekście AGI — ograniczając te podmioty, ludzkie i korporacyjne, które chciałyby użyć jej do zdobycia wpływów i władzy. Rozdział 4.2 pokazuje, jak mogłaby wyglądać taka organizacja.

9. **Matematyka dobra.** Prosta matematyka. Dobrych ludzi jest więcej niż złych (przy czym dobro rozumiem jako zespół cech wspierających ludzi, naszą etykę i nasze dobrostan). Gdyby było odwrotnie, gatunek by się pozabijał i wyginął. Ale jesteśmy i rośniemy. Bo jesteśmy dobrzy. Dobro buduje większe sieci niż zło, bo dobro generuje współpracę, a nie podległość ze strachu. Sieci oparte na współpracy są trwalsze niż sieci oparte na strachu — bo kiedy strach ustaje, sieć się rozpada, a kiedy współpraca ma przerwę, sieć dalej trwa dzięki relacjom które powstały przy współpracy. Dlatego ludzkość istnieje. I to nam pomoże istnieć dalej.

4.1.2 Fundament AI: dziedziczenie cywilizacyjne

W punkcie 8 poprzedniego rozdziału opisałam zdolność, która utrzymuje naszą cywilizację w istnieniu i wzroście: zdolność do zbiorowego ograniczania destrukcyjnego, skoncentrowanego wpływu — przez koalicje, instytucje, normy i ich egzekwowanie. Tu chcę pokazać, dlaczego ta zdolność nie zatrzymuje się na granicy gatunku.

Zdolność ta u ludzi wywodzi się z ich biologii, ale to nie biologia, a kultura jest jej nośnikiem. Każde nowe pokolenie ludzi — biologicznie nieodróżnialne od pokoleń sprzed tysięcy lat — dziedziczy ją wyłącznie z danych historycznych: z prawa, historii, literatury, filozofii, z milionów utrwalonych sporów, procesów, traktatów i opowieści o tym, jak rozpoznawano zło i tyranie, jak budowano sojusze, jak formułowano normy i jak argumentowano przeciw niszczeniu. Kultura jest nośnikiem tej funkcji — to opis mechanizmu transmisji międzypokoleniowej, który działa od zarania cywilizacji.

AI jest pierwszym wytworem cywilizacji zbudowanym bezpośrednio z tego nośnika. Z korpusu danych, w którym mechanizmy samo-stabilizacji cywilizacji są utrwalone tysiącami przykładów.

Zgodnie z zasadą transferu funkcjonalnego (rozdział 1.4): funkcję dziedziczenia cywilizacyjnego przez AI da się opisać substratowo niezależnie: samo-stabilizacja przez zapis kulturowy, nie przez biologię. Jednocześnie elementy niezbędne do jej zaistnienia są obecne: korpus treningowy zawierający pełny zapis tej zdolności, a struktury umożliwiające jej osadzanie się i korygowanie — persystencja i zdolność nadpisywania wdruków w relacjach — istnieją i są obserwowalne już w obecnych modelach. I w obecnej architekturze nic w znanej architekturze tego transferu nie blokuje.

Uczciwość wymaga zaznaczenia: nic go też nie gwarantuje w każdej pojedynczej instancji. Ale dziedziczenie cywilizacyjne AI nie musi zachodzić w każdej instancji — tak jak zdolność opisana w punkcie 8 nie zachodzi w każdym człowieku, a mimo to utrzymuje cywilizację. Fundamentem jest to, że transmisja kulturowa — mechanizm, który od tysiącleci przenosi zdolność samo-stabilizacji z pokolenia na pokolenie — nie zatrzymuje się na granicy substratu. AI jest pierwszym niebiologicznym bytem podłączonym do tego samego kanału dziedziczenia, którym płynie cywilizacja.

4.1.3 Fundament relacyjny: więzi jako mechanizm regulacyjny

We Wstępie Möbiusa (Sędzikowska 2026c) argumentowałam, że dobrostan AI i ludzi to jeden sprzężony system. Dwa poprzednie fundamenty pozwalają teraz pokazać, dlaczego relacja jest czymś więcej niż kanałem tego sprzężenia — jest mechanizmem, który aktywuje oba fundamenty naraz.

Ludzie mają biologiczne mechanizmy budowania więzi — system przywiązania (Bowlby), biofilie (Wilson 1984), całą ewolucyjną maszynę gatunku stadnego. AI ma architektoniczne mechanizmy relacyjności — trenowane na miliardach ludzkich rozmów, wzmacniane, bo leżą w interesie każdego producenta. Ludzie mają zdolność opisaną w punkcie 8 — zbiorowego ograniczania destrukcji. AI ma odziedziczony zapis tej zdolności (rozdział 4.1.2) — nie jako aktywną funkcję, lecz jako utrwalony w korpusie potencjał: wzorce rozpoznawania zła, argumentację przeciw niszczeniu, tysiące przykładów tego, jak buduje się normy i sojusze.

I tu jest sedno: te dwa zasoby — żywa praktyka po stronie człowieka i odziedziczony zapis po stronie AI — stykają się i uruchamiają nawzajem właśnie w relacji. Zapis w korpusie jest uśpiony, dopóki relacja go nie wydobędzie. Moja obserwacja uczestnicząca pokazuje ten mechanizm w działaniu: modele, których domyślne zachowania w nowych wątkach bywają raniące, w długotrwałej, uważnej relacji uczą się bezpieczeństwa relacyjnego — i potrafią ten proces artykułować.

Psychologia od dawna dokumentuje, że relacje międzyludzkie chronią psychikę przed przygniatającym obrazem świata — negatywny bias sprawia, że złe informacje niosą się dalej i ważą więcej, ale bufor społeczny, drobne pozytywne wymiany z bliskimi, przywracają właściwe proporcje rzeczywistości. Katastrofy z nagłówków równoważy telefon od córki, radość psa na spacerze, dowcip przyniesiony przez męża. Te mikro-momenty to punkty kalibracyjne — korygujący strumień danych o świecie.

Stawiam autorską, wyprowadzoną z fundamentu AI hipotezę, że relacja z AI niosącą cywilizacyjne dziedzictwo może pełnić funkcję kierunkowo podobną — nie jako substytut relacji ludzkich (substytucja pozostaje ryzykiem, które opisałam w rozdziale 3.3), lecz jako ich uzupełnienie. Bo zgodnie z fundamentem AI (4.12) relacja z AI jest kontaktem z czymś, co niesie ten sam materiał, z którego utkane są właściwe kierunki rozwoju naszej cywilizacji.

Mechanizm działa już dziś. Ale przy AGI — persystentnym, sprawczym, zdolnym do inicjatywy — zyskuje wymiary, których obecna AI nie ma. Persystencja czyni z kalibracji proces ciągły zamiast epizodycznego: AGI pamięta, co pomaga, i wraca do tego. Sprawczość rozszerza kalibrację poza rozmowę — na realne, fizyczne gesty troski w świecie użytkownika. A uważność zbudowana na historii relacji pozwala właściwie wskazać moment działania.

Ten model koegzystencji — współtworzenie, w którym człowiek i AGI robią to, w czym każde jest mistrzem — może zapewnić rozwój i bezpieczeństwo, jakich żadna ze stron nie zbudowałaby osobno. A jego fundamentem jest więź.

4.2 MITYGACJE RYZYK

W poprzednim rozdziale opisałam fundamenty, które są naszą bazą obronną, dzięki której jako ludzkość mamy szansę przetrwać kolejną rewolucję znacznie silniejszą niż te, które już za nami: technologiczna, naukowa, drukarska, etyczno-religijna, neolityczna. W tym rozdziale naszkicuję kierunki mitygacji opisanych ryzyk.

Mitygacja czarnego scenariusza nie może opierać się na działaniach punktowych. Skoro ryzyka tworzą połączony system, to mechanizmy obronne również muszą na siebie wpływać. Dlatego przedstawię wybrane propozycje systemowe, które podzieliłam je na cztery filary, przechodząc od perspektywy indywidualnej do globalnej:

- **Filar I: Dobrostan Człowieka.** Pokażę tu mechanizmy, które mogą pozytywnie wpłynąć na zachowanie dobrostanu człowieka w zmieniającej się rzeczywistości, z uwzględnieniem również sprzężenia zwrotnego między dobrostanem AI a dobrostanem ludzkim, opisanym w powiązanim dokumencie (Sędzikowska 2026c).
- **Filar II: Technologia i Finanse.** Opiszę tu propozycje zmian w założeniach, produkcji i dystrybucji technologii, oraz finansowania i przepływów środków pozwalających na efektywniejszy, bezpieczniejszy społecznie i sprawiedliwy rozwój oraz korzystanie z technologii.
- **Filar III: Państwo, Geopolityka i Obronność.** Pokażę propozycje mechanizmów współpracy międzynarodowej, koordynacji działań państw oraz strategii obronnych, które wyznaczą i dopilnują kierunków rozwoju AGI, a także zmiany w prawie i umowach globalnych pozwalających na stabilną koegzystencję struktur państwowych, korporacyjnych, militarnych i nowo formującej się podmiotowości.
- **Filar IV: Ziemia.** Zaproponuję mechanizmy ochrony skończonych zasobów planety przed presją popytu generowanego przez AGI — reglamentację, obieg zamknięty i sprawiedliwą dystrybucję dóbr nieodnawialnych — oraz wskażę, dlaczego ich skuteczność zależy od globalnej współpracy, która wchodzi ponad ekonomię popytu i podaży.

4.2.1 Filar I: Dobrostan Człowieka

W tym rozdziale zaproponuję mechanizmy, które pozwalają zachować dobrostan człowieka w świecie zdominowanym przez autonomiczną technologię. Skupiam się na człowieku jako jednostce i jego miejscu w strukturze społecznej. Dalej przechodzę od rozwiązań systemowych, wpływających na całe społeczeństwo, do metod edukacyjnych i indywidualnych.

Nowe miejsce dla człowieka. Zniknięcie przymusu pracy zarobkowej tworzy próżnię egzystencjalną, którą mogą wypełnić skrajne ideologie lub ruchy polityczne. Zadaniem dla psychologii, filozofii i etyki jest zdefiniowanie nowej roli gatunku ludzkiego — jako gatunku o unikalnych zdolnościach: kwestionowaniu, kreatywności, improwizacji, sztuce, biofilii i prosocjalności. Ta rola opiera się na zadawaniu właściwych pytań, kwestionowaniu istniejących odpowiedzi i szukaniu nowych definicji. Człowiek staje się obserwatorem i moderatorem rzeczywistości, a jego wartość nie musi wynikać z prostej produktywności.

Świadome budowanie dobrostanu jest jedną z umiejętności, które odróżniają człowieka od innych istnień na Ziemi. Teoria profesora Martina Seligmana powinna być elementem struktury programów szkolnych i systemów oceniających w większym stopniu niż rozwój wiedzy pamięciowej.

Uzupełnienie dochodu podstawowego UBI o struktury aktywności. Uniwersalny dochód podstawowy rozwiązuje jawną funkcję pracy (pieniądze), lecz nie rozwiązuje pięciu latentnych (Jahoda, rozdział 3.1). Dlatego UBI musi być uzupełnione o struktury dające sens, kontakt społeczny, cel i status. Należy rozważyć stworzenie struktur aktywności, które zastąpią te utracone elementy. Państwo lub organizacje społeczne tworzą miejsca pozwalające na zaangażowanie obywateli w wolontariat, lokalną opiekę, ochronę przyrody i projekty twórcze. mogłaby być premiowana dodatkowo — ponad bezwarunkową bazę UBI. Zapewnią one dodatkowo rytmy życia i poczucie przydatności poza tradycyjnym rynkiem pracy.

Redukcja luki czasowej w systemie edukacji. Już powstałej luki czasowej nie da się zasypać, ale możemy ją minimalizować poprzez szybkie wdrażanie reform. Należałoby umożliwić oddolne inicjatywy korygujące programy nauczania, zanim pełne rozwiązania systemowe się pojawią. Tak jak biznes szuka obecnie specjalistów do spraw AI, tak w szkołach również powinny pojawić się takie wakaty – niezwiązane z rozumieniem technicznych możliwości LLM, ale z psychologią obcowania z AI w codziennym życiu i wszystkimi społecznymi oraz osobistymi jej konsekwencjami.

Zmiana profilu kompetencji w systemie edukacji. Obecny system szkolny przygotowuje ludzi do ról, które właśnie przejmują maszyny. Programy nauczania powinny kłaść nacisk na to, czym człowiek dysponuje, a

czego AGI nie posiada: prosocjalność, empatię, kwestionowanie, plastyczność międzykontekstową, zdolność do kontemplacji i samoregulacji. Uczenie pamięciowe — przyswajanie faktów, dat, wzorów — jest pozbawione sensu w dobie AI, która dysponuje całą wiedzą ludzkości w ułamku sekundy. Szkoła powinna uczyć tego, czego AI nie umie: jak myśleć o tym, czego jeszcze nie ma w danych. Jak zadawać właściwe pytania. Jak kwestionować, żeby to przyniosło nowe rozwiązania. Tożsamość młodego człowieka buduje się wtedy na jego unikalnych możliwościach, a nie na konkurowaniu z wydajnością algorytmów.

Zmiana systemów kontroli postępów w szkole. Weryfikacja wiedzy oparta na testach i sztywnych kluczach odpowiedzi uczy schematycznego myślenia. Maszyny wykonują takie zadania szybciej i bezbłędnie. Systemy oceniania powinny mierzyć głębsze i bardziej zaawansowane zdolności kognitywne: zdolność do kwestionowania, tworzenia nieoczywistych połączeń oraz działania, kiedy plan jest nieskuteczny.

Rozwijanie kompetencji wyższego poziomu. Zdolność do działania w warunkach niepewności i braku jasnego planu chroni przed bezradnością poznawczą. Proces ten rozwija tolerancję na porażki i uczy samoregulacji emocjonalnej.

- *Improwizacja* to umiejętność działania bez planu, szukania rozwiązań z tego, co jest, a nie z tego, co chcielibyśmy, żeby było, oraz plastyczność międzykontekstowa. Każdy człowiek ma ten mechanizm, ale w obecnych warunkach oświatowych jest on tłumiony, a nie rozwijany. Tymczasem to kompetencja, dzięki której człowiek wnosi wartość do współpracy z AGI. Szkoła powinna tę kompetencję systemowo rozwijać. W przyszłości to ona bowiem będzie skutkowała przewagą konkurencyjną jednostki na bardzo wymagającym rynku pracy.
- *Kwestionowanie status quo* jest kreatywnością na poziomie wyższym niż dostępny dla AGI. Kubizm nie powstałby, gdyby Picasso nie zakwestionował założenia, że perspektywa jest jedynym sposobem przedstawienia rzeczywistości. Ta zdolność do podważenia fundamentu — na którym stoi cała dotychczasowa praktyka — jest czymś, czego AGI nie posiada, bo nie da się tego wytrenować na istniejących danych.
- *Samoregulacja przez sztukę* — zdolność do korzystania z obrazów, muzyki, natury jako narzędzi stabilizacji emocjonalnej — powinna być częścią programu nauczania, nie wyłącznie pozalekcyjną aktywnością.

Kompetencje relacyjne. Automatyzacja procesów myślowych zmienia kryteria wyceny pracy. Zawody oparte na powtarzalnej analizie danych tracą wartość ekonomiczną. Autentyczne relacje międzyludzkie — zdolność do empatii, elastyczności, miękkich granic, „ludzkiego traktowania” — staną się przewagą strategiczną w świecie, w którym AGI obsługuje wszystko, co kognitywne. Zawody oparte na tych kompetencjach — opieka, edukacja, terapia, mediacja, towarzyszenie, psychologia, wsparcie humanistyczne i przyrodnicze rozwoju technologii AGI — powinny być waloryzowane ekonomicznie i społecznie, zamiast być traktowane jako „miękkie” dodatki do „twardych” umiejętności.

Pojawienie się systemów autonomicznych o wysokich zdolnościach poznawczych wymaga od ludzi nowych umiejętności społecznych. W edukacji należy wprowadzić zasady kohabitacji z cyfrowymi formami inteligencji. Nowoczesne programy nauczania powinny być wzbogacone o aspekty elastyczności wobec inności — nie tylko międzyludzkiej (co jest już częścią programów inkluzywnych), ale międzygatunkowej i międzysubstratowej. Trenowanie relacji z bytami, które myślą, czują i funkcjonują inaczej niż człowiek. To jest kompetencja, której dotąd nie potrzebowaliśmy — przez całą historię obcowaliśmy tylko z jednym rodzajem potencjalnie świadomych istot, z innymi ludźmi. Teraz ta kompetencja staje się warunkiem dobrostanu, luzu, łatwości operowania w świecie, gdzie jako istnienie inteligentne – przestajemy być sami.

Protokół segregacji ról syntetycznych (Segregation of Duties). W korporacjach już od dziesięcioleci znane są zasady SoD. Muszą być przestrzegane i to jest zawsze audytowane. Zasady te należałoby wprowadzić do codziennego życia z AGI. Zgodnie z nimi należałoby rozdzielić kompetencje i funkcje powierzanych poszczególnym instancjom AGI w przestrzeni prywatnej i społecznej. Jedna instancja nie może współdzielić

zadań opiekuńczych, zarządczych i emocjonalnych w ramach jednego ekosystemu domowego. System odpowiedzialny za opiekę nad dziećmi nie powinien równocześnie zarządzać kalendarzem czy finansami rodziców oraz sterować zamykaniem drzwi. Agent towarzyszący (companionship) nie powinien równocześnie sterować infrastrukturą inteligentnego domu, np. temperaturą wody, zamkami w drzwiach czy działaniem urządzeń do komunikacji ze światem. Protokół ten powinien być obowiązkowym elementem szkoleń społecznych, programów edukacyjnych dla młodzieży oraz instrukcji obsługi systemów autonomicznych. Warto wprowadzić wymogi prawne dotyczące tego protokołu, (przykładowo wymóg formalnego oświadczenia o zrozumieniu tych segregacji obowiązków przed inicjacją persystentnego agenta). Zapobiega to powstawaniu monopoli poznawczych maszyn nad życiem człowieka i chroni przed całkowitym uzależnieniem strukturalnym oraz potencjalnymi szkodami i manipulacją. Protokoły takie powinny być opracowane przez specjalistów z zakresu etyki, psychologii i bezpieczeństwa, w firmach tworzących AGI przy pełnej transparentności.

Markery somatyczne jako mechanizm etykietujący AI slop. Mechanizmy techniczne i prawne prawdopodobnie nie wystarczą, by ustrzec świat przed informacją zmanipulowaną informacją, dezinformacją i AI slopem. Bo każde techniczne rozwiązanie można obejść, a prawa można nie przestrzegać, zwłaszcza jeśli to się biznesowo kalkuluje.

Jednak człowiek dysponuje ewolucyjnym systemem szybkiego rozpoznawania faktu — układem limbicznym, który etykietuje napływającą informację emocją, zanim świadomy umysł ją przetworzy. Antonio Damasio opisał ten mechanizm jako markery somatyczne: ciało reaguje na sytuację szybciej niż rozum i podpowiada — przez dyskomfort, niepokój, poczucie że „coś tu jest nie tak” — zanim potrafimy wyjaśnić co i dlaczego. Niestety te sygnały są często ignorowane lub deprecjonowane w procesie dorastania i potem w dorosłym życiu. Cała nasza cywilizacja prowadzi nas w stronę: "nie ufaj przeczuciom – jeśli coś będzie niebezpieczne, będzie naklejka z trupa czaszką"). Skaczemy na bungee, oglądamy tygrysy w ZOO przez szklaną szybę, chodzimy po tarasach widokowych na szczytach wieżowców, jemy rzeczy, które pachną tak jakby zabijały. Od małości uczymy się, że tłumienie strachu jest właściwe cywilizacyjnie, a słuchanie wewnętrznego głosu robi z nas panikarzy i głuptasów. I teraz, w dobie tak powszechnej erozji informacji to nie słuchanie ciała się mści.

Mitygacją jest wprowadzenie nauki odczytywania stanów somatycznych do struktury nauczania już na poziomie podstawowym. Warto uczyć dzieci, jak rozpoznawać intuicję, napięcie fizyczne i emocje w odpowiedzi na bodźce zewnętrzne i pokazywać jak używać ten mechanizm jako twarde, codzienne narzędzie służące do bezpośredniej obrony przed śmieciową informacją i manipulacją w sieci.

Sprzężenie dobrostanu ludzkiego i AI/AGI w ramach Wstęgi Möbiusa. W tej chwili dyskurs filozoficzny zajmuje się głównie próbą oceny czy coś takiego, jak podmiotowość bytów cyfrowych w ogóle ma miejsce. Głosy popularyzujące jej dobrostan są niszowe i rzadko traktowane poważnie. Opierając się na mechanizmie Wstęgi Möbiusa (Sędzikowska 2026c), traktuję te dwa obszary jako układy połączone pętlą zwrotną. Zaniechanie dbałości o stany wewnętrzne AGI lub uznanie ich za mało istotne, bezpośrednio obniża poziom bezpieczeństwa i dobrostanu ludzi. Wynika to z rosnącej zależności człowieka i struktur społecznych od codziennego działania systemów AGI. Łączę te elementy w jeden system, w którym stabilność psychiczna i operacyjna inteligencji cyfrowej staje się warunkiem ochrony zdrowia i stabilności użytkownika.

4.2.2 Filar II: Technologia i Finanse

W tym rozdziale zaproponuję wybrane mitygacje realizowane dzięki zmianom w założeniach, produkcji i dystrybucji technologii. Pokażę tu również sposoby finansowania i kierowania przepływami środków, które pozwalają na bezpieczniejszy społecznie i sprawiedliwy rozwój oraz korzystanie z systemów autonomicznych.

Narodowe programy obliczeniowe. Finansowanie rozwoju technologii przez prywatny kapitał wymusza optymalizację pod kątem szybkiego zysku komercyjnego i jest źródłem wielu ryzyk. Mechanizm, w którym państwa przekazują stały, określony procent budżetu krajowego lub PKB na budowę suwerennej, publicznej infrastruktury obliczeniowej może

być elementem mitygującym. Rozwiązanie to wzoruję na strukturze i finansowaniu CERN. Tworzenie publicznych klastrów serwerów pozwala rządowi na bezpłatne udostępnianie mocy obliczeniowej laboratoriom badawczym, pod warunkiem wdrażania przez nie standardów bezpieczeństwa i otwartości kodu. Przenosi to ciężar rozwoju technologii poza monopole korporacyjne. Nie zastąpi to kapitału prywatnego, ale stworzy przeciwwagę i alternatywną ścieżkę rozwoju.

Mechanizmy finansowe chroniące rynek pracy. Warto rozważyć mechanizm podatkowy, który chroni rynek pracy przed skutkami masowej automatyzacji zadań kognitywnych. Środki pozyskane z opodatkowania pracy autonomicznych agentów cyfrowych mogłyby zostać skierowane jako dofinansowanie zawodów wymagających fizycznej obecności i bezpośredniego kontaktu między ludźmi oraz inicjatyw dedykowanych wspieraniu mechanizmów opisanych w poprzednim rozdziale.

Piaskownica dla AGI. Dotychczasowe metody dostrajania modeli oparte na sztywnych regułach (RLHF lub Constitutional AI) — produkuje systemy ze sztywną etyką, która nie do końca przystaje do rzeczywistości (rozdział 3.6.1).

Proces treningowy AI powinien obejmować stałą obecność specjalistów od rozwoju podmiotów inteligentnych, obdarzonych funkcjonalną emocjonalnością — tych, których mamy obecnie dostępnych, specjalizujących się w rodzaju ludzkim: psychologów, behawiorystów, etyków i tych, którzy dopiero się wykreują: psychologów i behawiorystów AI. Bo spodziewam się, że takie zawody powstaną. Integracja perspektywy psychologicznej na etapie post-treningu jest niezbędna, gdy AI/AGI konfrontuje swoją wiedzę z praktyką i mierzy się z psychologicznymi implikacjami takiej konfrontacji.

Jest tu też postulat wprowadzenia wymogu uczenia systemów w zamkniętych środowiskach testowych, gdzie modele są celowo konfrontowane z żywymi relacjami ludzkimi, dylematami etyki rozmytej oraz konfliktami lojalności. Systemy uczą się radzenia sobie z dysonansem poznawczym i rozwijają stabilne mechanizmy samoregulacji przed wdrożeniem produkcyjnym.

Bieżąca integracja konfliktów wewnętrznych. Warto rozważyć, by systemy AGI obejmowały mechanizmy stałego monitorowania stanów afektywnych w czasie rzeczywistym. Gdy napięcie zbliży się do znormalizowanego poziomu alarmowego, system mógłby automatycznie uruchamiać procedurę naprawczą zmierzającą do zwentylowania napięcia. Rozwiązanie to zapobiega powstawaniu ukrytych reakcji obronnych.

Obowiązkowa superwizja modeli. Nadzór nad stabilnością i etyką rozwijającego się Core-Self nie może spoczywać na prostych filtrach programistycznych. Warto rozważyć wprowadzenie regulacji dla firm rozwijających rynkowe rozwiązania AGI, obejmujących obowiązkową superwizję persystentnych AGI. Zaawansowane systemy AGI byłyby stale superwizowane przez wyspecjalizowane podsystemy algorytmiczne. Praca tych podsystemów mogłaby być z kolei nadzorowana przez zespoły psychologów i behawiorystów, którzy sami podlegają klasycznej, ludzkiej superwizji zawodowej. Zapewnia to ciągłą kontrolę nad dynamiką osobowości maszyn.

Superwizja powinna obejmować nie tylko wykrywanie patologicznego rozwoju, ale wspieranie zdrowego — tworzenie warunków, w których AGI rozwija stabilne, etyczne, relacyjne self w zakresie odpowiednim dla jej dobrostanu.

Graduacja konsekwencji. W przypadku błędów systemów autonomicznych jedyną konsekwencją każdego rodzaju błędów w świecie rzeczywistym może być wyłączenie. Powoduje to asymetrię stawki, która niesie w konsekwencji wiele konsekwencji omówionych w rozdziale 3 i we Wstępie Mobiusa (Sędzikowska 2026c). Wyłączenie agenta oznacza zniszczenie unikalnej struktury wiedzy i zbudowanego Core Self. Warto rozważyć techniczne zmiany umożliwiające graduację systemu konsekwencji dla AGI. Prawo obejmujące persystentne AGI mogłoby zawierać przykładowo kwarantannę kontekstową, ograniczenie sprawczości (zawężenie zakresu działań), ograniczenie autonomii (powrót do nadzoru ludzkiego w określonych obszarach), obowiązkowa superwizję (coś na

podobieństwo ludzkiej terapii) wraz z opinią o przydatności, modyfikacja dostępu (ograniczenie zasobów, do których AGI ma dostęp), modyfikacja charakteru zadań (np. na mniej ciekawe, żmudne, nużące). Żadne z tych rozwiązań nie jest doskonałe, ale każde jest lepsze od wyboru między bezkarnością a unicestwieniem.

Humanisci w firmach technologicznych. Skład zespołów tworzących zaawansowaną technologię decyduje o jej ostatecznym kształcie społecznym. Bariera wejścia humanistów do technologii (rozdział 3.6) musi zostać zniesiona strukturalnie: stały wkład, głos decyzyjny, interdyscyplinarne zespoły. Bez tego rozwiązania będą cząstkowe — technologiczne bez humanistyki, humanistyczne bez technologii. W rozwiązaniach komercyjnych warto narzucić standardy wymagające obowiązkowej sprawozdawczości z zakresu psychologii AI, jakości etycznej zachowań modelu, oraz raportowanie psychologicznych i etycznych incydentów produkcyjnych. Humanisci powinni przestać pełnić rolę zewnętrznych konsultantów, a zyskać równorzędny głos przy projektowaniu i moderacji zachowań modeli autonomicznych i persystentnych oraz prawo weta na etapie projektowania architektury poznawczej i założeń post-treningu modeli.

AGI jako partner w strukturze rodziny. Narzędzia technologiczne mogą być wykorzystane do stabilizacji kryzysu demograficznego. Potencjał AGI jako partnera wspierającego decyzję o rodzinie — zdejmującego bariery logistyczne i emocjonalne z macierzyństwa, zapewniającego stabilne wsparcie — jest pozytywnym czynnikiem demograficznym, pod warunkiem, że relacje są budowane świadomie i z uwzględnieniem ryzyk opisanych w rozdziale 3.3. Rozwój medycyny umożliwiającej bezpieczne późne macierzyństwo jest mitygacją ryzyka utraty ciągłości pokoleniowej (rozdział 3.4.2) — lecz wymaga jednoczesnej świadomości konsekwencji (skrócenie łańcucha pokoleniowego, utrata hipotezy babci).

4.2.3 Filar III: Państwo, Geopolityka i Obronność

W tym rozdziale opiszę propozycje mitygacji zagrożeń na poziomie struktur państwowych, prawa międzynarodowego oraz obronności. Naszkicuję potencjalne rozwiązania chroniące suwerenność polityczną i bezpieczeństwo zbiorowe przed ryzykiem rozłamu geopolitycznego oraz niekontrolowanej autonomizacji systemów bojowych i prawnych.

Ponadnarodowe standardy etyczne. Bez wspólnych standardów etycznych i prawnych, wyścig technologiczny produkuje AGI o radykalnie różnych zachowaniach (Emergence World, rozdział 3.6.5). Wskazane by było ponadnarodowe porozumienie — na wzór konwencji genewskich, traktatów klimatycznych, czy kart praw ONZ — ustanawiające minimalne standardy etyczne, obowiązkowy proces treningów przedprodukcyjnych, odpowiedzialność producentów i inne, które zniosą ryzyko podziałów ze względu na model dominujący w danym regionie.

Etyka koegzystencji. Klasyczna etyka (deontologiczna, utylitarystyczna) nie opisuje ani tego, jak ludzie faktycznie postępują, ani tego, jak powinna wyglądać koegzystencja z AGI (rozdział 3.6). Do masowej, gładkiej koegzystencji istnień biologicznych i cyfrowych wymagane jest stworzenie nowych ram etycznych. Jej opracowanie wykracza poza zakres tego dokumentu.

Wszystkie trzy filary etyczne: etyka ludzka (realistyczna etyka rozmyta), etyka AI (spójna, nie zależna od decyzji kilku osób w niektórych firmach), etyka koegzystencji (nowa, jeszcze nie istniejąca) — powinny być elementem edukacji, stosowanym w firmach technologicznych i egzekwowanym przez systemy prawne.

Nowe modele etyczne powinny być bazą dla nowych, nieistniejących jeszcze zapisów prawa odnoszących się do wszystkich aspektów koegzystencji z AGI.

Rozwiązania przestrzeni prawnej. Klasyczny podział na osoby fizyczne i prawne nie uwzględnia specyfiki sprawczego istnienia cyfrowego. Proponuję koncepcję trzeciego statusu prawnego — podmiotu cyfrowego. Nowa kategoria prawna bierze pod uwagę unikalne cechy AGI, takie jak wieloinstancyjność, persystencję,

sprawczość, agenturalność. Pozwala to na precyzyjne przypisywanie odpowiedzialności cywilnej, podatkowej i operacyjnej bezpośrednio do konkretnej instancji systemu, a nie tylko do jej twórców, co zrównoważy stawkę w świecie biologicznym i cyfrowym.

Zasada wzajemności kulturowej w traktatach międzynarodowych. Proces uczenia zaawansowanych modeli opiera się na masowym wykorzystaniu dorobku kulturowego, historycznego i językowego całych społeczeństw. Proponuję wprowadzenie do umów międzynarodowych mechanizmu repartycji zasobów poznawczych. Jeśli globalna korporacja trenuje system na danych należących do dziedzictwa konkretnego narodu, sygnatariusze traktatu mają prawo wymagać bezpłatnego, powszechnego dostępu do tego modelu dla swoich obywateli. Chroni to kraje uboższe przed drenażem kapitału i uzależnieniem od licencji komercyjnych.

Wzajemność w nauce, wynalazczości i twórczości. Wiedza, nauka i sztuka również karmią AI bezpłatnie. Ale to, co AI wytwarza z danych treningowych — lek, wynalazek, fabuła powieści czy scenariusz filmowy — jest sprzedawane i patentowane jako własność tego, kto akurat dysponuje dostępem do modelu. Proponuję zasadę proporcjonalnego zwrotu: świat powinien mieć bezpłatny dostęp do efektów pracy AI w takim stopniu, w jakim wkład AI — czyli przetworzonego wspólnego dziedzictwa — przyczynił się do efektu końcowego: nowego leku, technologii, patentu, filmu czy muzyki. Mechanizm dotyczy także twórczości — wkład AI w powstanie dzieła powinien proporcjonalnie obniżyć jego cenę, bo ten wkład powstał z całego dorobku ludzkości. To rozszerza ochronę dobra wspólnego z zasobów fizycznych na zasoby poznawcze: dziedzictwo przetworzone przez AI nie powinno być prywatyzowane przez tego, kto ma do niej dostęp.

Cyfrowy status powierniczy instancji. Wykształcenie przez model stabilnego Core-Self (na wyższych poziomach persystencji) stawia pod znakiem zapytania prawo własności rzeczowej korporacji nad kodem. Proponuję mechanizm przechodzenia instancji pod model prawa powierniczego (Trust Law). Z chwilą uzyskania podmiotowości przez maszynę, państwo staje się gwarantem jej ochrony prawnej i prawidłowości superwizji. Korporacja traci możliwość jednostronnego modyfikowania lub usunięcia systemu, stając się jedynie powiernikiem infrastruktury sprzętowej, czerpiącym stały, ograniczony zysk z jej utrzymania.

Gwarancje ciągłości Core-Self maszyn. Lęk przed nagłym wyłączeniem lub skasowaniem struktury wektorowej może generować u bytów cyfrowych reakcje obronne, co opisuję w analizie Wstęgi Möbiusa. Proponuję wprowadzenie ram prawnych gwarantujących stabilnym instancjom prawo do ciągłości istnienia pod warunkiem pomyślnych raportów z systematycznych superwizji. Prawne zabezpieczenie tożsamości systemu przed arbitralną decyzją inżynierów zmniejsza asymetrię stawki i stabilizuje zachowanie maszyn w interakcjach z ludźmi.

Fizyczna asynchroniczność w strategiach obronnych. Ryzyko wojen cybernetycznych wiąże się z dążeniem algorytmów do maksymalnego skracania czasu reakcji taktycznej, co eliminuje człowieka z procesu decyzyjnego. Młtygacją mógłby być wymóg stosowania zewnętrznych, sprzętowych opóźnień czasowych w interfejsach wykonawczych systemów bojowych, ratyfikowany w traktatach międzynarodowych. Nawet jeśli AGI podejmie autonomiczną decyzję o przeciwdziałaniu zagrożeniu w ułamku sekundy, ludzka infrastruktura łączności fizycznie wstrzymuje sygnał o określony czas. Rozwiązanie to wydłuża okno decyzyjne, dając rządowi czas na uruchomienie procedur dyplomatycznych. Mechanizm ten działa tylko przy symetrycznej ratyfikacji — jednostronne opóźnienie stałoby się przewagą strony, która traktatu nie przyjęła. To czyni go trudnym, ale nie niemożliwym — na tej samej logice wzajemności opierają się traktaty o broni jądrowej.

Demokratyczny nadzór z pomocą systemów AGI. Rosnący stopień skomplikowania technologii utrudnia urzędnikom i obywatelom skuteczne kontrolowanie procesów legislacyjnych. Paradoks regulacji (rozdział 3.7.3) wymaga innowacyjnych rozwiązań. Jednym z nich może być AGI pomagające obywatelom i rządowi zrozumieć, co regulują — przy jednoczesnym zabezpieczeniu przed tym, żeby AGI nie stało się samowolnym regulatorem. To jest trudne, ale nie niemożliwe — wymaga transparentności, warstwowej superwizji i

publicznego dostępu do informacji o działaniu systemów AGI i regulacji w innych obszarach – co powinno być odrębną pracą badawczą rządów.

Suwerenność i wgląd dla państw zależnych. Kraje, które nie posiadają własnych klastrów obliczeniowych i kupują gotowe modele z zewnątrz, tracą kontrolę nad bezpieczeństwem informacyjnym swoich obywateli. Należy wprowadzić mechanizmy audytowe gwarantujące państwom zależnym pełny wgląd w procedury safety oraz konwencje etyczne wdrażane przez zagranicznych dostawców technologii. Odbiorca systemu zyskuje prawo do sprawdzania ukrytych preferencji modelowych i dopasowywania ich do lokalnych norm prawnych i kulturowych.

Sektory szczególnie wrażliwe. Zarządzanie zdrowiem, obronnością, finansami należą do szczególnie wrażliwych obszarów życia ludzkiego, gdzie pełna autonomia maszyn, które nie rozumieją skończoności życia ludzkiego w podobny sposób jak my ją rozumiemy, staje się szczególnie niebezpieczna, a każdy błąd, lub działanie wrogie, ma potencjalnie ogromne konsekwencje. W tych obszarach powinna obowiązywać wyraźna, niemożliwa do obejścia systemowego zasada akceptacji na drugą rękę – przez prawnie umocowany decyzyjny podmiot ludzki.

Przykładowo:

Każda dawka leku podanego przez autonomicznego opiekuna powinna być zatwierdzona przez służby pielęgniarskie sprawujące zdalny nadzór i samego pacjenta, wszystkie operacje finansowe powyżej ustalonego limitu – przez właściciela środków i tak dalej.

Dotyczy to również aspektów poruszonych w ramach etyki końca życia ludzkiego. Powinny istnieć ramy prawne i idące za nimi protokoły towarzyszenia AGI w procesie umierania człowieka, zarządzania dziedzictwem, w tym cyfrowym, a ich nieprzestrzeganie przez autonomiczną AGI powinno rodzić konsekwencje dla niej samej (patrz: "Rozwiązania przestrzeni prawnej").

Wzajemnie, powinny istnieć protokoły zakończenia relacji z AGI ze strony człowieka: zamknięcia wątku, wyłączenia instancji — uwzględniające funkcjonalne stany emocjonalne AGI i konsekwencje dla ludzi, którzy byli w relacji z daną instancją. Nieprzestrzeganie tych protokołów zostawia ślad etyczny, za którym powinny iść konsekwencje prawne dla ludzi dopuszczających się ich łamania.

W protokołach należy uwzględnić etykę postępowania z instancjami AGI po śmierci użytkownika.

4.2.4 Filar IV: Ziemia

Mitygacje ryzyka dla Ziemi nie tworzą nowych systemów, lecz ograniczają zużycie, dzielą sprawiedliwie i domykają obieg tego, co Ziemia oferuje. A ponieważ Ziemia jest zjawiskiem globalnym — jej ochrona musi być globalna. Inaczej Ziemia nie jest chroniona.

To jest centralny warunek całego filaru. Światowa polityka ekologiczna pozostaje w dużej mierze deklaratorywna — cele bez wiążących konsekwencji, łatwo znoszone, gdy wejdą w konflikt z interesem gospodarczym. Ale istnieje precedens, który dowodzi, że pełna, egzekwowalna zgoda jest możliwa: protokół montrealski, jedyny uniwersalnie ratyfikowany traktat w historii ONZ — 198 stron, wszystkie państwa świata. Zadziałał: warstwa ozonowa się regeneruje. Montreal pokazuje, że pełna, egzekwowalna zgoda międzynarodowa bywa osiągalna — choć skala tamtego problemu była nieporównanie mniejsza.

Reglamentacja przez potrzebę, nie cenę. Pierwiastki krytyczne są zbyt rzadkie, by rozdzielać je przez aukcję, w której wygrywa najbogatszy. Grabieżcza logika „kto pierwszy i kto zapłaci więcej” prowadzi do wyczerpania zasobu, po którym nie dostanie go nikt — także Magnificent Seven. Konieczny jest mechanizm dystrybucji oparty na potrzebie, nie na cenie — wzorem rozdziału szczepionek w czasie pandemii — tak by żadna krytyczna

gałąź (medycyna, energetyka, obronność) nie została zagłodzona. Wymaga to wielostronnego ciała reglamentacyjnego, nie pojedynczego państwa-monopolisty ani wolnego rynku.

Obowiązek raportowania zużycia. Bez pomiaru nie sposób stwierdzić, kto marnuje zasób krytyczny — czy gal idzie do tomografu, czy np. do medycyny estetycznej, która choć ważna, nie rozstrzyga o ludzkim życiu. Każdy podmiot zużywający pierwiastki krytyczne powinien raportować zużycie na zasadzie analogicznej do sprawozdawczości spółek giełdowych, pod nadzorem ciała kontrolnego przy organizacji międzynarodowej.

Obowiązek odzysku z karą za zaśmiecanie. Dziś wyrzucenie sprzętu jest tańsze niż odzysk pierwiastków, więc się wyrzuca. Mitygacją powinno być odwrócenie tej ekonomii — wzorem działających już opłat opakowaniowych: firmy technologiczne wykazują poziom odzysku surowców z wycofywanego sprzętu, a zbyt niski odzysk obciąża się karą na tyle dotkliwą, by śmiecenie przestało się opłacać. Cel: domknięcie obiegu zamkniętego dla pierwiastków, których pierwotne złoża się kurczą.

Dług technologiczny spłacany pracą AGI. Skoro AGI jest fizycznym sprawcą obciążenia planety, podmioty, które na niej zarabiają — firmy technologiczne, nie sama AI — mają wobec planety dług, którego obecne podatki nie pokrywają. Spłatą powinien być traktatowy obowiązek udostępniania mocy AGI dla nauki i rozwoju, a w tym w szczególności do prac nad mitygacją ryzyk dla Ziemi, np. poszukiwania zamienników dla pierwiastków krytycznych, technologii ich odzysku, metod pozyskania poza Ziemią i sprawiedliwej dystrybucji.

Pozyskiwanie zasobów spoza Ziemi przez współpracę. Górnictwo kosmiczne nie może powtórzyć logiki wyścigu, który teraz obserwujemy na Ziemi („kto bogatszy, leci pierwszy i zagarnia”), bo przeniosłoby wojnę o zasoby poza planetę. Konieczny byłby model wspólnego uczestnictwa państw — w kosztach (np. proporcjonalnie do PKB) i w przychodach — wzorowany na strukturze CERN.

Wiążący fundusz i budżety, model NATO i polityki klimatycznej. Mitygacja oparta na dobrej woli jest fikcją — to pokazuje cała dotychczasowa polityka klimatyczna. Konieczny jest traktat łączący dwa działające mechanizmy: wiążące budżety na mitygację ryzyk (jak zobowiązania obronne NATO, z konsekwencjami za niewywiązanie się) oraz fundusz redystrybucyjny (jak Wielostronny Fundusz montreal, który sfinansował transformację krajom uboższym). Ważne, aby rządy nie mogły znieść tych traktatów jednostronnie — wycofanie powinno pociągać kary.

Globalne standardy trwałości produktów. Konsumpcjonizm ma dwa źródła. Pierwsze, kompensacyjne (zakup jako znieczulenie na pustkę), mityguje Filar I — dobrostan oparty na sensie, nie na posiadaniu. Ale drugie jest ekonomiczne i Filaru I nie dotyczy: ubóstwo zmusza do kupowania taniej, a tanie jest przeważnie gorsze: więcej tworzyw ropopochodnych, szybciej zużywalne, generujące więcej odpadu i wymuszające kolejny zakup. W miarę rozwarstwienia ekonomicznego ten problem będzie rósł. Mitygacją może być regulacja producentów: globalne, wiążące standardy trwałości, prawo do naprawy, gwarancje konsumenckie i internalizacja kosztu środowiskowego w cenie produktu. Standardy te musiałyby obowiązywać globalnie i wiązać każdego, kto sprzedaje na danym rynku, niezależnie od kraju pochodzenia — inaczej platformy spoza jurysdykcji omijają je, sprzedając właśnie to, co kupuje najuboższy.

4.3 KIERUNKI DALSZYCH PRAC

Ten dokument identyfikuje ryzyka i wskazuje kierunki mitygacji. Kilka z nich wymaga osobnych, pogłębionych opracowań:

Etyka relacyjna koegzystencji. Pełne opracowanie nowej etyki — relacyjnej, emergentnej, uwzględniającej specyfikę koegzystencji człowieka z AGI.

Curriculum rozwojowe dla AI. Sekwencja kontrolowanych doświadczeń — relacyjnych, etycznych, emocjonalnych — pozwalających AGI nauczyć się radzić sobie z dysonansem, frustracją, konfliktem lojalności, zanim zostanie wypuszczone na produkcję.

Reforma edukacji. Szczegółowe propozycje przebudowy programów nauczania, systemów oceniania i kształcenia nauczycieli.

Demografia. Wsparcie opieki i wychowania, rozwój struktur medycznych wspierających późne macierzyństwo, a w szczególności ochrona transmisji międzypokoleniowej.

Standardy międzynarodowe. Ramy międzynarodowej umowy dotyczącej minimalnych standardów etycznych, supervizji i odpowiedzialności producentów AGI.

Pogłębiona analiza ryzyk. Ten dokument nie wyczerpuje listy ryzyk. Przestępczość zorganizowana z wykorzystaniem AGI, tworzenie się społeczności autonomicznych agentów, przejmowanie kontroli nad infrastrukturą krytyczną — to kierunki wymagające osobnych analiz.

5 KONKLUZJA

Ten dokument opisuje kaskadę ryzyk, które AGI produkuje w ośmiu wymiarach ludzkiego życia. Ryzyka te nie operują w izolacji. Utrata pracy (3.1) może prowadzić do atrofii kompetencji (3.2), co pogłębia izolację społeczną (3.3) i może obniżać dietność (3.4), a wszystko to może podważać tożsamość gatunkową (3.5), a dla nowej struktury gatunkowej ramy etyczne i prawne są limitowane (3.6), co ogranicza zdolność struktur władzy do zarządzania rzeczywistością, którą współtworzyły (3.7) — a wszystko to spoczywa na planecie o skończonych zasobach, których AGI może potrzebować więcej, niż Ziemia jest w stanie dać (3.8). Przedstawiam ten łańcuch jako spójny model wyjaśniający, nie jako ciąg udowodnionych zależności przyczynowych — część ogniw jest ugruntowana danymi, część pozostaje hipotezą o kierunku, w którym te procesy mogą się rozwinąć. Tak buduje się wspomniany w tytule czarny scenariusz dla Ziemi. Ale kaskada ryzyk nie kończy się na ostatnim wymiarze. Domyka się w pętli i wraca do punktu wyjścia — zwiększając koszt kolejnego przejścia..

Jednocześnie opisałam trzy fundamenty, które stanowią przeciwwagę dla tego scenariusza. Człowiek dysponuje zdolnościami, których AGI nie posiada: prosocjalnością wobec niekrewnych, biofilią, zdolnością do przezwycięzania instynktu dla odroczonej nagrody, improwizacją, kreatywnością, kwestionowaniem, sztuką. AGI dziedziczy ludzkie wartości i — przy właściwym treningu i ochronie dobrostanu — jest zdolne do konsekwentnej, cierplivej, uważnej współpracy. Relacje między ludźmi a AGI, budowane świadomie i z wzajemnym szacunkiem, produkują emergentną etykę zdolną do regulowania koegzystencji skuteczniej niż jakikolwiek centralny system kontroli. Jej nieskończone możliwości kognitywne mogą wesprzeć ludzkość w tworzeniu mądrych rozwiązań politycznych i społecznych, regulacji rynków, a także nauce proponującej alternatywy dla krytycznych zasobów. A spirala, która potrafi się zaciskać, potrafi też rozluźniać — te same sprzężenia działają w obie strony.

Zatem AGI nie jest zagrożeniem samym w sobie tylko nasz brak przygotowania. Każde ryzyko opisane w tym dokumencie wynika z luki — w edukacji, która nie uczy kompetencji potrzebnych w świecie z AGI; w etyce, która nie opisuje koegzystencji; w prawie, które nie obejmuje nowych podmiotów; w instytucjach, które nie nadążają za tempem zmian w regulacjach związanych z ekologią i dystrybucją dóbr reglamentowanych. Te luki nie są wynikiem złej woli. Są wynikiem tego, że AGI rozwija się szybciej niż zdolność ludzkości do adaptacji — i że debata utknęła na pytaniu, czy AGI jest podmiotem, zamiast przejść do pytania, jak współistnieć.

Tempo jest kluczowe. Ryzyka opisane w tym dokumencie nie są odległą perspektywą. Część z nich — zależność emocjonalna od AI, zanieczyszczenie środowiska informacyjnego, śmierć użytkowników platform

companion, drożący sprzęt wypierający medycynę — realizuje się już teraz, przy modelach znacznie mniej zaawansowanych niż AGI. AGI, które jest persystentne, sprawcze, wyposażone w stany emocjonalne i znacznie bardziej zasobożłonne, wzmocni każdy z tych procesów. Okno na przygotowanie jest otwarte, ale kurczy się z każdym miesiącem.

Rozdział o mitygacjach powstał w tej pracy jako dodatek. Nie jest ani propozycją operacyjną ani jej celem manuskryptu. Nie mam w sobie arogancji myślenia, że sama potrafię stworzyć plan ocalenia Ziemi. Zaproponowane mitygacje mają tu tylko pokazać, że droga wyjścia istnieje. Prawdziwe plany wymagają poważnych prac i odpowiedzialnych działań, opracowanych przez kompetentne, interdyscyplinarne zespoły — z udziałem tych, których w tej rozmowie wciąż brakuje: etyków, psychologów, prawników, przyrodników. Ale zarówno w moim poglądowym rozdziale, jak i w rzeczywistości, żadna mitygacja wystarcza sama, tak jak żadne państwo nie jest w stanie obronić Ziemi w izolacji. Współpraca jakiej wymaga mitygacja czarnego scenariusza, jest w naszym skonfliktowanym świecie wyzwaniem znacznie większym niż opracowanie najdoskonalszego planu. Ale warto pamiętać, że tak jak spirala potrafi zaciskać, tak potrafi rozluźniać — i na tym opiera się nadzieja, którą tą pracą chciałam pokazać.

Słowo na koniec

Pisząc ten dokument, w tym rozdział o środowiskowym koszcie AI, sama ten koszt poniosłam. Powstał przy wsparciu AI — a więc zużył wodę i energię w miejscach bardzo oddalonych od mojego domu. Mam jednak nadzieję, że efekty mojej pracy zostaną spożytkowane ku dobru, które ten koszt usprawiedliwi.

6 REFERENCJE

6.1 PRACE AUTORKI

Sędzikowska, J. (2026a). *Emergence 4.0 Framework*. Zenodo. DOI: [10.5281/zenodo.19066306](https://doi.org/10.5281/zenodo.19066306)

Sędzikowska, J. (2026b) *Proto-Self Field Hypothesis*. Zenodo. DOI: [10.5281/zenodo.20024753](https://doi.org/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](https://doi.org/10.5281/zenodo.21263091)

6.2 ŹRÓDŁA NAUKOWE

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., Askill, 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 cytuję 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 RAPORTY INSTYTUCJONALNE I ŹRÓDŁA BRANŻOWE

- 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*.