

Mapping the Unmapped: A Structured Atlas of Human Futures with Artificial Superintelligence and the Cultural Visibility

Hypothesis

Aleksandar Stojanovic
DARI Foundation, Geneva, Switzerland

Correspondence: aleks@dari.earth
May 2026

ABSTRACT

The rapid advancement of artificial intelligence toward artificial general intelligence (AGI) and potentially artificial superintelligence (ASI) has generated a vast, fragmented, and often contradictory discourse about humanity's future. This discourse spans academic research, policy documents, industry publications, online communities, and popular culture, but no systematic attempt has been made to catalogue, classify, and compare the full spectrum of scenarios being proposed. This paper presents the SuperFutures Scenario Atlas, a structured taxonomy of 55 distinct scenario families for the human–ASI relationship, each classified across identity, type, mechanism, time horizon, human position, twelve societal impact domains, discourse map, pop culture footprint, and audience resonance. Drawing on established foresight methodology, the atlas treats scenarios as exploratory tools rather than predictions, and as a tagged graph rather than a mutually exclusive taxonomy. The paper makes two original theoretical contributions. First, the Cultural Visibility Hypothesis (CVH): the proposition that public perception of AI risk is driven primarily by cultural representation in film, television, and literature, rather than by expert assessment of plausibility. We present preliminary evidence that scenarios with high pop culture footprints (AI rebellion, extinction, surveillance) are systematically overestimated by the public, while scenarios with low cultural visibility but high expert plausibility (institutional lag, gradual integration, algorithmic capitalism) are systematically underestimated. Second, the theory of Scenario Reflexivity: the mechanism by which belief in a scenario generates behavioral patterns that make the scenario materially more probable. Drawing on Merton's self-fulfilling prophecy, Callon and MacKenzie's performativity, and Soros's reflexivity, we demonstrate through five

illustrative cases that AI future scenarios are not merely predictions to be evaluated but performative propositions that shape the conditions of their own realization. The implications for AI governance, public communication, and foresight methodology are discussed. The atlas is made publicly available as a research and engagement resource.

Keywords: artificial superintelligence; scenario planning; foresight; AI governance; cultural visibility; risk perception; pop culture; public engagement; scenario taxonomy

1. INTRODUCTION

The prospect of artificial superintelligence - AI systems that exceed human cognitive capacity across all domains - has moved from the margins of science fiction and philosophy to the center of global policy, investment, and public discourse. The leading AI laboratories now explicitly target artificial general intelligence (AGI) on timelines measured in years (Altman, 2024; Amodei, 2024; Hassabis, 2024), and the implications of such systems for labor, governance, security, culture, meaning, and existential survival are being debated with increasing urgency across academia, policy, industry, and the public.

Yet this debate is profoundly fragmented. At least eight major intellectual camps propose fundamentally different futures for the human–ASI relationship, ranging from post-scarcity abundance (Altman, 2021; Amodei, 2024) to human extinction (Bostrom, 2014; Yudkowsky, 2022), from techno-feudal concentration of power (Varoufakis, 2023; Zuboff, 2019) to democratic technological empowerment (Buterin, 2023; Weyl & Posner, 2018), from transhumanist merger (Kurzweil, 2005) to deliberate withdrawal from technology (Berry, 1987). These camps operate with different vocabularies, different assumptions, different publication norms, and often without awareness of each other's arguments.

The result is a discourse that is socially salient but cognitively fragmented. Policymakers regulate against assumed futures that may not materialize. Executives invest against a single scenario without considering alternatives. Citizens form opinions shaped by cultural products - films, novels, and television series - that dramatize some scenarios while ignoring others entirely. No shared, structured, comprehensive map of the terrain exists.

This paper addresses that gap by presenting the SuperFutures Scenario Atlas: a structured, publicly accessible taxonomy of 55 distinct scenario families for humanity's future with artificial

superintelligence. The atlas draws on established foresight methodology (Schwartz, 1991; van der Heijden, 1996; Wilkinson & Kupers, 2014) and applies it to the specific domain of human–ASI futures, with adaptations for the unprecedented pace, scope, and cultural saturation of the AI discourse.

Beyond the atlas itself, this paper introduces the Cultural Visibility Hypothesis (CVH): the proposition that public perception of AI futures is driven more by cultural representation in popular media than by expert assessment of analytical plausibility. We present preliminary evidence for this hypothesis and discuss its implications for AI governance, public communication, and foresight practice.

The paper is organized as follows. Section 2 reviews the relevant literature on scenario planning, AI foresight, and the role of popular culture in shaping technology perceptions. Section 3 describes the methodology for atlas construction, including taxonomy design, scenario identification, and classification protocols. Section 4 presents the atlas structure and a descriptive overview of the 55 scenario families. Section 5 introduces the Cultural Visibility Hypothesis and presents preliminary evidence. Section 6 discusses implications for governance, communication, and foresight practice. Section 7 acknowledges limitations. Section 8 concludes.

2. LITERATURE REVIEW

2.1 Scenario Planning and Strategic Foresight

Scenario planning - the use of multiple divergent narratives to explore uncertainty rather than predict a single future - has a well-established methodological tradition. Originating in military strategy (Kahn & Wiener, 1967) and popularized by Royal Dutch Shell's planning group in the 1970s (Wack, 1985; Schwartz, 1991), the method has been widely adopted by the UNDP, OECD, European Commission, and major consultancies (Wilkinson & Kupers, 2014). The core principle is that divergent scenarios surface hidden assumptions, test strategies across multiple possible futures, and prepare decision-makers for uncertainty rather than encouraging false precision (van der Heijden, 1996).

In foresight practice, scenarios are not predictions but structured thought experiments. They are typically presented in sets of two to four contrasting narratives, each internally consistent but mutually divergent (Schwartz, 1991). The method has been validated across domains including

energy (Shell Scenarios), geopolitics (National Intelligence Council, 2021), public health (UNICEF), and technology governance (European Commission, 2023).

However, the application of rigorous scenario methodology to AI futures has been limited. Most AI foresight work takes the form of single-scenario analysis - a specific claim about what AI will or will not do - rather than structured multi-scenario exploration (Dafoe, 2018). Notable exceptions include Bostrom's (2014) taxonomy of AI outcome types, Drexler's (2019) 'Comprehensive AI Services' framework, and Ord's (2020) existential risk analysis. But no work has attempted to catalogue the full spectrum of scenarios currently in active discourse, classify them using a consistent taxonomy, and present them as a navigable atlas.

2.2 The AI Futures Discourse

The discourse on AI futures has grown rapidly since 2014, driven by breakthroughs in deep learning, the advent of large language models, and increasingly explicit AGI timelines from major AI laboratories. The discourse can be organized into several major traditions, each with distinct intellectual roots, institutional bases, and policy implications:

The existential risk tradition, anchored by Bostrom (2014), Yudkowsky (2022), and Russell (2019), emphasizes the alignment problem and the possibility of human extinction or permanent disempowerment from misaligned superintelligence. The techno-optimist tradition, represented by Kurzweil (2005), Altman (2021), Amodei (2024), and Andreessen (2023), anticipates abundance, radical life extension, and human enhancement. The political economy tradition, represented by Zuboff (2019), Varoufakis (2023), Acemoglu and Restrepo (2019), and Crawford (2021), emphasizes power concentration, labor displacement, and surveillance. The governance tradition, aligned with UNESCO (2021), OECD (2019), and the EU AI Act, emphasizes institutional adaptation, human rights, and regulatory preparedness. The accelerationist tradition, formalized by e/acc proponents (Verdon et al., 2022) and rooted in Land's earlier work, advocates unrestricted technological acceleration. The decolonial tradition, articulated by Birhane (2020), Mohamed et al. (2020), and the TESCREAL critique of Gebru and Torres (2024), challenges the ideological foundations of the dominant AI discourse.

These traditions operate largely in parallel, with limited cross-referencing, incompatible vocabularies, and different assumptions about what constitutes evidence. No work has attempted to map them into a unified, navigable framework.

2.3 Popular Culture and Technology Perception

The role of popular culture in shaping public perceptions of technology has been studied across several domains but is notably underexplored for AI specifically. The general literature on media framing and risk perception (Slovic, 1987; Kahneman & Tversky, 1979) establishes that perceived risk is driven more by availability heuristics - the ease with which examples come to mind - than by statistical probability. Cultural products such as films, novels, and television series are powerful suppliers of available examples.

For nuclear technology, studies have documented the influence of films like *The China Syndrome* (1979) and *Silkwood* (1983) on public risk perception (Weart, 2012). For biotechnology, cultural narratives around *Frankenstein* and genetic engineering have been shown to shape regulatory attitudes (Turney, 1998). For AI specifically, several commentators have noted the influence of the *Terminator* franchise on public perception of AI risk (Bostrom, 2014; Russell, 2019; Cave et al., 2018), but no systematic mapping of cultural visibility across a full spectrum of AI scenarios has been attempted.

Cave, Dihal, and Dillon (2020) at the Leverhulme Centre for the Future of Intelligence developed a taxonomy of AI narratives in fiction, identifying four major narrative types (AI as tool, AI as threat, AI as companion, AI as transformer). Their work provides a valuable starting point but does not connect these narratives to specific analytical scenarios, does not assess the relationship between cultural visibility and perceived probability, and does not cover the full breadth of AI futures discourse beyond fiction.

This paper extends this line of inquiry by (a) mapping cultural visibility systematically across a comprehensive atlas of AI future scenarios, and (b) proposing and providing preliminary evidence for the Cultural Visibility Hypothesis: that cultural representation, not expert assessment, is the primary driver of public risk perception for AI.

3. METHODOLOGY

3.1 Atlas Design Principles

The SuperFutures Scenario Atlas was designed according to six principles derived from established foresight methodology:

Comprehensiveness. The atlas aims to catalogue every major scenario family currently in active discourse, not a curated selection. The criterion for inclusion is that a scenario is discussed by identifiable proponents in published or semi-published form (academic papers, books, policy documents, major essays, podcasts, institutional statements, or organized online communities). Purely speculative scenarios without identifiable advocacy are excluded.

Structured classification. Every scenario is classified using the same taxonomy, enabling systematic comparison. The taxonomy was designed iteratively through three expansion rounds, with fields added when existing fields proved insufficient to capture relevant distinctions.

Non-exclusivity. Scenarios are treated as a tagged graph, not a mutually exclusive list. Many scenarios overlap, hybridize, or are preconditions for each other. Cross-referencing and network visualization are integral to the atlas design.

Separation of functions. Classification, interpretation, probability assessment, and editorial commentary are kept distinct within each scenario profile. The atlas does not smuggle advocacy into taxonomy.

Transparency. Every scenario profile discloses its evidence basis, the strongest critiques of its claims, its unresolved assumptions, and its proponents and critics by name.

Cultural mapping. Each scenario profile includes a systematic survey of its representation in popular culture (film, television, literature, music, games, and visual art), along with an assessment of the relationship between cultural visibility and perceived probability. This field is the empirical basis for the Cultural Visibility Hypothesis.

3.2 Scenario Identification

Scenarios were identified through a multi-source review process covering:

Canonical discourse: Books, academic papers, policy documents, institutional reports, and manifestos. Key sources included Bostrom (2014), Kurzweil (2005), Russell (2019), Harari (2016,

2024), Varoufakis (2023), Zuboff (2019), UNESCO (2021), OECD (2019), Andreessen (2023), Verdon et al. (2022), Buterin (2023), and Gebru and Torres (2024), among others.

Active discourse: News coverage, podcasts, conference proceedings, newsletters, think tank commentary, institutional releases, and AI laboratory publications.

Engagement discourse: Social media, online communities (LessWrong, Alignment Forum, e/acc Twitter, Reddit), memes, subcultural expressions, and informal publications. This level is included because many scenario variants appear first in informal discourse before being formalized.

Cultural discourse: Films, television series, novels, short stories, music, games, and visual art that embody, popularize, or critique specific AI future scenarios.

The identification process was iterative. An initial seed set of 10 scenario families was derived from the most prominent traditions in the literature (singularity, transhumanism, extinction, abundance, governance, oligarchy, obsolescence, pause, emulation, hybrid futures). Three expansion rounds followed, each adding 12–20 scenarios identified through gap analysis, respondent feedback, and emerging discourse monitoring. The final atlas contains 55 distinct scenario families.

3.3 Taxonomy Design

Each scenario is classified across the following fields, designed to enable systematic cross-scenario comparison:

Identity fields: Scenario name, alternative names, short definition (50–100 words), one-sentence thesis, and discourse status (mainstream, emerging, fringe, legacy, fictional-origin, or hybrid).

Classification fields: Scenario type (from a controlled vocabulary: utopian, dystopian, mixed, transition, governance, metaphysical, posthuman, civilizational collapse, coexistence, merger, replacement). Main claim. Mechanism (the causal pathway by which the scenario occurs). Time horizon (near term 0–5 years, mid term 5–15, long term 15–50, deep future 50+). Human position (sovereign, co-equal, dependent, subordinated, merged, obsolescent, extinct, or transformed beyond recognition).

Societal impact fields: Structured assessments across twelve domains: labor and income, education, governance and democracy, war and security, inequality and class structure, culture and art, meaning and purpose, family and reproduction, health and longevity, rights and agency, environmental effects, and existential survival. Each assessment is a qualitative narrative of 30–100 words.

Discourse map fields: Key proponents (named individuals), key institutions, canonical texts (with full citations), strongest critiques, and major unresolved assumptions.

Pop culture footprint fields: Notable references in film, television, literature, music, games, and visual art (with titles, years, and brief descriptions of relevance). A qualitative Cultural Visibility Score (high, moderate, low, or very low). A narrative Share-of-Culture Note analyzing the relationship between cultural visibility and perceived probability for this specific scenario.

Acceptance fields: Primary audiences (which intellectual, professional, or political communities find the scenario most compelling). Geographic concentration.

Persona fields: Two brief interpretive personas (one sentence each) illustrating typical proponents. One detailed 'Hard Believer' user profile (name, age, occupation, location, daily life, media diet, core conviction, social circle, biggest fear, biggest hope) designed to make the scenario emotionally concrete for non-expert audiences.

3.4 Cultural Visibility Assessment

The pop culture footprint of each scenario was assessed through a systematic review of cultural references identified in the scenario identification process. For each scenario, we catalogued all identified references in film, television, literature, music, games, and visual art, and assessed their aggregate cultural visibility using a qualitative four-point scale:

Very high: Multiple iconic, globally recognized cultural works directly dramatize this scenario (e.g., *The Terminator* and *The Matrix* for AI rebellion/extinction; *Nineteen Eighty-Four* for surveillance).

High: Several well-known cultural works are closely associated with this scenario (e.g., *Blade Runner* for AI consciousness; *Gattaca* for cognitive stratification).

Moderate: Some cultural works reference this scenario, but they are not its primary dramatization (e.g., Wall-E for human obsolescence; various dystopian fiction for techno-feudalism).

Low or very low: Few or no significant cultural works directly dramatize this scenario (e.g., institutional lag, algorithmic capitalism, managed coexistence, AI as invisible infrastructure).

This assessment is qualitative and preliminary. A quantitative methodology - based on cultural reference counts, box office data, publication sales, streaming viewership, and social media citation frequency - is proposed as future work (see Section 7).

4. THE SUPERFUTURES SCENARIO ATLAS: OVERVIEW

4.1 Scope

The first edition of the SuperFutures Scenario Atlas contains 55 fully structured scenario profiles. The scenarios span the entire ideological and analytical spectrum of current human–ASI discourse, from radical acceleration (effective accelerationism, Scenario 48) to radical withdrawal (Neo-Luddite enclaves, Scenario 24), from extinction (Scenario 10) to transcendence (transhumanist enhancement, Scenario 3), from the most dramatic (intelligence explosion, Scenario 1) to the most mundane (AI as invisible infrastructure, Scenario 40).

The 55 scenarios are organized into several broad clusters, though they resist clean categorization due to extensive overlap:

Acceleration and transformation scenarios (1–5, 25–26, 36–37, 48–49): Hard takeoff, soft takeoff, transhumanist enhancement, mind upload, aligned abundance, longevity escape velocity, AI colonization of space, AI-driven scientific revolution, collective superintelligence, technocapital singularity, and deceleration/d/acc.

Governance and institutional scenarios (6, 9, 12, 28, 30, 34, 42, 54–55): Managed coexistence, pause/moratorium, AI-mediated democracy, benevolent AI governance, institutional lag, lethal autonomous weapons, algorithmic courts, AI as oracle, and sovereign AI.

Dystopian and critical scenarios (7–8, 10, 13–14, 20, 23, 27, 29, 31–35, 41, 46, 50–51): AI oligarchy, human obsolescence, extinction, surveillance singularity, AI arms race, cognitive stratification, algorithmic capitalism, cultural atrophy, biopolitical control, AI companionship, epistemic collapse, digital colonialism, great decoupling, human domestication, recursive bureaucracy, TESCREAL hegemony, and Moloch.

Metaphysical and posthuman scenarios (4, 17–18, 21, 38, 43–44, 47, 52–53): Mind upload, AI as new species, simulation discovery, AI theocracy, digital afterlife, voluntary handover, consciousness expansion, synthetic biology convergence, cosmic endowment, and consciousness co-evolution.

Realist and structural scenarios (11, 15, 22, 24, 40, 45): Hybrid fragmented futures, post-work society, decentralized AI commons, neo-Luddite withdrawal, AI as invisible infrastructure, and AI and disability liberation.

4.2 Distribution of Scenario Characteristics

Across the 55 scenarios, the distribution of key classification variables reveals several patterns:

Scenario type: 19 scenarios are classified as dystopian or primarily dystopian, 14 as utopian or conditionally utopian, 12 as mixed, and 10 as transition or governance scenarios. The preponderance of dystopian scenarios reflects the discourse itself, where critical and cautionary voices currently outnumber optimistic ones in volume if not in influence.

Time horizon: 22 scenarios are near-to-mid term (already underway or expected within 15 years), 18 are mid-to-long term (15–50 years), and 15 are long term to deep future (50+ years). The concentration of scenarios in the near-to-mid term reflects the acceleration of AI discourse since 2022.

Human position: Only 12 scenarios maintain human sovereignty as the default. 18 describe humanity as dependent or subordinated. 10 describe transformation or merger. 8 describe obsolescence. 7 describe extinction or irreversible loss of agency. This distribution reflects a discourse in which optimism about human agency is a minority position.

5. THE CULTURAL VISIBILITY HYPOTHESIS

5.1 Statement

The central original contribution of this paper is the Cultural Visibility Hypothesis (CVH):

The AI futures that the public considers most probable are not those that domain experts assess as most analytically plausible, but those that have the highest representation in popular cultural forms - particularly film, television, and literature. Cultural visibility, not analytical plausibility, is the primary driver of public risk perception for artificial intelligence.

The CVH draws on and extends the availability heuristic framework of Kahneman and Tversky (1979) and the risk perception work of Slovic (1987). The core mechanism is straightforward: people assess the probability of events partly by the ease with which relevant examples come to mind. Popular cultural products - films watched by hundreds of millions, novels that define genres, television series that enter the cultural lexicon - are extremely powerful suppliers of available examples. When a citizen is asked 'what do you think AI will do?', the answers that come to mind most easily are those dramatized by the Terminator franchise, the Matrix trilogy, Nineteen Eighty-Four, Blade Runner, and Black Mirror - not those discussed in policy papers, economic analyses, or governance frameworks.

5.2 Preliminary Evidence

The evidence for the CVH is currently preliminary and qualitative, derived from the systematic pop culture mapping conducted across all 55 scenario profiles in the atlas. A full quantitative test requires survey data correlating public scenario beliefs with cultural consumption patterns, which is proposed as future work. The preliminary evidence takes two forms.

5.2.1 The Overestimation Pattern

The scenarios with the highest cultural visibility scores in the atlas are:

Extinction / AI Rebellion (Scenario 10): Cultural visibility: very high. Key references: The Terminator (1984–), The Matrix (1999–), Battlestar Galactica (2004–09), Avengers: Age of Ultron (2015), I, Robot (2004). The Terminator franchise alone has grossed over \$2 billion globally. The concept of 'killer robots' is among the most recognized AI tropes in global culture.

Surveillance / AI Panopticon (Scenario 13): Cultural visibility: very high. Key references: Nineteen Eighty-Four (1949), Minority Report (2002), Black Mirror (2011–). The term 'Orwellian' has entered common English vocabulary. 'Big Brother' is a globally recognized metaphor.

AI Consciousness / Machine Rights (Scenario 17): Cultural visibility: very high. Key references: Blade Runner (1982/2017), Ex Machina (2014), Her (2013), A.I. Artificial Intelligence (2001), Westworld (2016–22).

Public surveys consistently show that these three scenario families dominate public perception of AI risk. Pew Research (2023) found that Americans' top AI concerns cluster around 'AI making decisions about people's lives,' 'AI being used to create fake content,' and 'loss of human jobs' - all themes heavily dramatized in popular culture. The 2023 YouGov survey found that 46% of Americans believe AI could lead to human extinction, a figure that many AI researchers consider dramatically inflated relative to the evidence.

5.2.2 The Underestimation Pattern

Conversely, the scenarios with the lowest cultural visibility scores in the atlas include several that governance experts, development economists, and technology historians consider among the most likely near-term outcomes:

Permanent Institutional Lag (Scenario 30): Cultural visibility: very low. No significant film, novel, or television series directly dramatizes the scenario of governance institutions permanently failing to keep pace with AI development. Yet many governance professionals consider this the most likely dystopian outcome.

AI as Invisible Infrastructure (Scenario 40): Cultural visibility: the lowest in the atlas. No films or novels dramatize the scenario where AI becomes as mundane and invisible as electricity. Yet this is arguably the most likely near-term trajectory for most people's experience of AI.

Algorithmic Capitalism (Scenario 23): Cultural visibility: low. The scenario of autonomous economic decision-making at machine speed has limited cultural representation, yet algorithmic trading, automated pricing, and AI-driven hiring are already pervasive.

The Great Decoupling (Scenario 35): Cultural visibility: low. The macroeconomic divergence of productivity and wages has few cultural representations, despite being empirically documented (Brynjolfsson & McAfee, 2014).

Managed Coexistence (Scenario 6): Cultural visibility: very low. The scenario of successful institutional adaptation is virtually absent from popular culture because gradual, successful governance is not dramatic. Yet it is the baseline scenario for most international governance frameworks (UNESCO, 2021; OECD, 2019).

5.2.3 The Structural Explanation

The asymmetry between cultural visibility and analytical plausibility is not random. It has a structural explanation rooted in the requirements of narrative form:

Popular culture selects for drama, not plausibility. A scenario must have a protagonist, an antagonist, conflict, stakes, and resolution to be dramatizable in film or fiction. AI rebellion (Terminator), surveillance (Nineteen Eighty-Four), and machine consciousness (Blade Runner) all provide these narrative elements naturally. Institutional lag, gradual integration, and algorithmic capitalism do not. They are structurally undramatizable - not because they are unimportant but because they lack narrative arc.

This selection effect is self-reinforcing. Cultural products shape public perception, which shapes political demand, which shapes media production, which shapes cultural products. The Terminator does not just reflect fear of AI rebellion - it creates and perpetuates it. The absence of films about institutional lag does not just reflect low interest - it creates and perpetuates low interest.

5.3 Implications

If the CVH is correct - and the preliminary evidence is suggestive though not yet conclusive - the implications for AI governance, public communication, and foresight practice are significant:

For governance: Policymakers should be aware that public demand for AI regulation is shaped by cultural narratives, not by expert risk assessment. The public may demand regulation against dramatic but unlikely scenarios (robot armies) while ignoring regulation for probable but undramatic scenarios (algorithmic discrimination, institutional lag). Calibrating governance to expert assessment rather than cultural salience is both necessary and politically difficult.

For public communication: Communicators about AI risk - whether in government, civil society, or AI companies - should recognize that their audience arrives with pre-installed narrative templates from popular culture. Effective communication must work with and against these

templates: acknowledging the scenarios the audience already 'knows' while redirecting attention to the scenarios they cannot see.

For foresight practice: Foresight professionals should include cultural visibility as a variable in scenario assessment. A scenario's cultural footprint is not just a communication challenge - it is a causal factor. Scenarios with high cultural visibility attract preparation, investment, and institutional attention, which can make them either more or less likely (a self-fulfilling or self-defeating dynamic). The Cultural Visibility Hypothesis suggests that narrative momentum should be treated as a variable in probability assessment, not just as a communication challenge.

5.4 Scenario Reflexivity: When Belief Creates Reality

The Cultural Visibility Hypothesis describes how cultural representation shapes belief. But belief does not remain inert. It drives behavior. And when enough people behave as though a particular AI future is probable or desirable, their aggregate behavior shifts the material conditions that make that future more likely. We call this mechanism Scenario Reflexivity: the process by which belief in a scenario generates the behavioral patterns that make the scenario materially more probable.

This mechanism extends the CVH from a theory of perception into a theory of causation. Cultural visibility shapes belief (CVH). Belief shapes behavior. Aggregate behavior shapes material reality. Material reality confirms the belief. The feedback loop is complete. The future is not merely predicted or imagined - it is, in part, performed into existence.

5.4.1 Theoretical Foundations

Scenario Reflexivity draws on several established theoretical traditions. The Thomas Theorem (Thomas & Thomas, 1928) states that 'if men define situations as real, they are real in their consequences.' Merton (1948) formalized this as the self-fulfilling prophecy: a false definition of the situation that evokes behavior making the false conception come true. Callon (2006) and MacKenzie (2006) developed the concept of performativity in economics, demonstrating that financial models do not merely describe markets but actively shape them - the Black-Scholes formula changed option pricing by being adopted by traders, not by being correct. Soros (1987) described reflexivity in financial markets: participants' beliefs about the market change the market, which changes their beliefs, creating feedback loops that cannot be understood

from outside the loop. Shiller (2019) extended this to narrative economics, showing that economic narratives go viral and change economic behavior - recessions can be partly caused by recession narratives.

The application of these frameworks to AI future scenarios has not been systematically attempted. The closest precedent is the concept of hyperstition, developed by the Cybernetic Culture Research Unit (CCRU) in the 1990s: ideas that make themselves real through their own propagation (Land, 1993). While the hyperstition concept is provocative, it lacks the empirical grounding and mechanistic specificity required for foresight and governance application. Scenario Reflexivity provides that grounding by specifying the mechanism through which belief produces behavioral change and behavioral change produces material outcomes.

5.4.2 The Mechanism in Detail

Scenario Reflexivity operates through a three-stage feedback loop:

Stage 1: Belief formation. An individual encounters a scenario through cultural products (film, fiction, media), expert discourse (papers, podcasts, conferences), or social influence (online communities, professional networks, peer groups). They come to believe the scenario is probable, desirable, or both.

Stage 2: Belief-driven behavior. The belief changes the individual's actions. These changes may be economic (investment, career choice, consumption), political (advocacy, voting, donation), technological (adoption, development, resistance), or social (community formation, lifestyle change, relationship patterns).

Stage 3: Behavioral aggregation. When a sufficient number of individuals act on the same belief, their collective behavior shifts material conditions - markets, institutions, technology adoption curves, political landscapes - in ways that make the scenario more probable. The belief becomes self-fulfilling not through any metaphysical mechanism but through the aggregate behavioral consequences of widespread adoption.

Critically, this process operates even when the original belief is analytically unfounded. A scenario does not need to be probable to become real - it only needs enough people to believe it is probable and to act accordingly. The aggregate behavior of believers can shift material conditions past tipping points that make the scenario materially irreversible.

5.4.3 Illustrative Cases

We illustrate Scenario Reflexivity with five cases drawn from the atlas:

Case 1: Transhumanist Enhancement (Scenario 3) - self-fulfilling through adoption. An individual believes cognitive enhancement through brain-computer interfaces is the future. To gain competitive advantage, they adopt an early-stage neural implant. Their adoption creates consumer demand. Demand funds further research and development. Improved devices attract more adopters. The enhancement market grows. Employers begin to expect augmentation. Unenhanced individuals fall behind. The transhumanist future materializes not because it was inevitable but because enough early believers created the demand, investment, and social pressure that made it self-reinforcing. Each adopter makes the next adoption more rational, creating a positive feedback loop.

Case 2: Post-Scarcity / UBI (Scenario 15) - self-fulfilling through economic behavior. An individual believes that AI automation will eliminate most employment within 15 years. Acting on this belief, they stop investing in long-term career development, spend rather than save ('use it or lose it' before the economic system restructures), and begin advocating for universal basic income. They join political organizations campaigning for UBI. Enough people behaving this way - reducing labor market attachment, increasing consumption, and building political coalitions - actually accelerates the conditions that make UBI politically necessary: lower workforce participation, higher demand for redistribution, and organized political pressure. The belief that work is ending contributes, at the margin, to the conditions that make work end.

Case 3: Effective Accelerationism (Scenario 48) - self-fulfilling through competitive pressure. An entrepreneur believes that unrestricted AI acceleration is both inevitable and desirable. They build an AI startup without investing in safety research, ship products as fast as possible, and lobby against regulation. Competing founders, observing the competitive dynamics, feel pressure to match the pace or be outcompeted. The aggregate effect of many founders behaving this way creates the very acceleration that the ideology predicted - not because the thermodynamic argument was correct, but because enough builders treated it as a mandate. The Moloch dynamic (Scenario 51) is the structural companion: competitive pressure makes individual restraint irrational even when collective restraint is optimal.

Case 4: Surveillance Singularity (Scenario 13) - self-fulfilling through resignation. An individual believes that comprehensive surveillance is technologically inevitable. Acting on this belief, they stop using encryption, stop supporting privacy legislation, and accept surveillance infrastructure as a fact of modern life. Enough people resigning in this way removes the political friction that could have constrained surveillance expansion. Privacy regulation lacks popular support. Companies face no consumer backlash for data collection. The panopticon materializes partly because enough people accepted it as fate - a self-fulfilling prophecy of surrender.

Case 5: Extinction Risk (Scenario 10) - simultaneously self-fulfilling and self-defeating. An individual believes that misaligned superintelligence poses a significant probability of human extinction. They donate to alignment research organizations, redirect their career toward AI safety, and advocate for regulatory constraints on AI development. At scale, this belief produces an alignment research industry, political pressure for AI governance, and institutional attention to safety. If these efforts succeed, the belief was self-defeating: the expectation of doom produced the behavior that prevented doom. But if the attention and resources diverted to extinction risk are drawn away from other, more probable risks - institutional lag, algorithmic discrimination, digital colonialism - then the belief may be self-fulfilling through neglect: the system fails not from the risk that was feared but from the risks that were ignored because attention was consumed by the dramatic scenario. This dual dynamic makes extinction risk the most complex case of Scenario Reflexivity in the atlas.

5.4.4 Implications for the Atlas and the CVH

Scenario Reflexivity has three implications for the SuperFutures framework:

First, the Probability Engine must account for reflexivity. A scenario's probability is not independent of how many people believe in it. The narrative momentum metric in the Probability Engine is designed to capture this: a scenario that is gaining believers is, other things being equal, becoming more probable - not because belief causes reality directly, but because aggregate belief-driven behavior shifts material conditions. This makes the Probability Engine partially self-referential: it measures a system in which measurement itself is a variable.

Second, the CVH and Scenario Reflexivity together describe a complete causal chain. Cultural products create scenarios in the public imagination (CVH). Imagined scenarios become beliefs. Beliefs drive behavior. Behavior changes material conditions. Changed conditions confirm

or modify the beliefs. The loop is: Culture → Belief → Behavior → Reality → Culture. The SuperFutures atlas, by mapping all four elements (cultural footprint, public belief, behavioral indicators, and material conditions), provides the first comprehensive empirical framework for tracking this loop across multiple AI future scenarios simultaneously.

Third, Scenario Reflexivity makes the atlas itself a participant in the system it describes. By publicizing low-visibility scenarios (institutional lag, algorithmic capitalism), SuperFutures may increase their cultural visibility, which may increase public awareness, which may increase the behavioral responses that address or exacerbate them. This is not a flaw but a feature: the purpose of foresight is not neutral observation but informed preparation. If the atlas increases the visibility of currently neglected scenarios and thereby catalyzes preparation for them, it is fulfilling its purpose - even at the cost of disrupting the asymmetry it describes.

5.5 Democratic Amplification: How Cultural Distortion Becomes Governance Failure

The CVH and Scenario Reflexivity describe how cultural visibility shapes belief and how belief shapes behavior. But in democratic societies, there is a specific institutional mechanism that converts distorted public belief into distorted resource allocation: the electoral process. This mechanism - which we term Democratic Amplification - completes the causal chain from pop culture to governance failure.

5.5.1 The Mechanism

In a functioning democracy, elected officials are responsive to public concern. They campaign on the issues that voters care about, legislate on the problems that constituents raise, and allocate budgets toward the risks that the electorate perceives as most urgent. This responsiveness is normally a virtue. But when public perception is systematically distorted by cultural visibility - as the CVH proposes - democratic responsiveness becomes a transmission mechanism for distortion.

The chain operates as follows. Cultural products (films, novels, television series) make certain AI scenarios highly salient in the public imagination. Voters arrive at the ballot box with beliefs about AI risk that are shaped more by the Terminator franchise than by any policy briefing. Political candidates, responding to polling data and constituent sentiment, campaign on the AI risks that voters already recognize: job displacement (culturally visible through extensive media

coverage), autonomous weapons (culturally visible through decades of robot-war cinema), and existential risk (culturally visible through the Terminator, the Matrix, and a growing corpus of doom-adjacent media). Candidates who win on these platforms then set legislative agendas, allocate budgets, and create institutions oriented toward the culturally visible risks.

Meanwhile, the scenarios with low cultural visibility - which may be far more probable and more immediately consequential - receive no electoral attention, no campaign rhetoric, no legislative priority, and therefore no institutional preparation or funding. The result is a systematic preparedness gap: democratic societies are institutionally prepared for the dramatic but uncertain futures that cinema made salient, and institutionally unprepared for the mundane but probable futures that resist dramatization.

5.5.2 Illustrative Cases

Case 1: Autonomous weapons vs. algorithmic discrimination. Public concern about 'killer robots' is extremely high, driven by the Terminator franchise, the Slaughterbots short film, and the Campaign to Stop Killer Robots. This concern has produced significant legislative attention: the UN Convention on Certain Conventional Weapons has held multiple rounds of discussion, and several states have called for preemptive bans. By contrast, algorithmic discrimination in credit scoring, insurance pricing, hiring, and criminal sentencing - which affects millions of people today - has received far less legislative attention in most jurisdictions, despite being empirically documented (ProPublica, 2016; O'Neil, 2016). The discrepancy is not proportional to the risks. Algorithmic discrimination is a present, documented, large-scale harm. Autonomous weapons are a future, speculative, smaller-scale risk. But the cultural visibility of killer robots vastly exceeds the cultural visibility of biased credit algorithms, and democratic responsiveness to voter concern has allocated institutional attention accordingly.

Case 2: Existential risk vs. institutional lag. The narrative of AI extinction risk has achieved extraordinary political salience since 2023, driven by the CAIS statement signed by hundreds of AI researchers, Geoffrey Hinton's public warnings, and extensive media coverage framed around 'will AI destroy humanity?' Several governments have created AI safety institutes explicitly oriented toward catastrophic and existential risk. Meanwhile, the scenario of permanent institutional lag - where governance simply fails to keep pace with AI deployment, producing cascading harms through regulatory inadequacy rather than through any single catastrophic event

- has received almost no dedicated institutional funding, no electoral attention, and no public urgency. Yet many governance professionals consider institutional lag the most probable dystopian outcome. The asymmetry in institutional response mirrors the asymmetry in cultural visibility: extinction is cinematic; institutional lag is bureaucratic.

Case 3: AI job displacement vs. the Great Decoupling. The fear that 'AI will take your job' is electorally powerful because it is personally threatening and culturally legible. Politicians campaign on job protection, retraining programs, and automation taxes. But the more structurally significant economic risk may not be mass unemployment but the Great Decoupling: a scenario in which productivity grows exponentially while median wages stagnate, GDP rises while living standards for the majority flatline, and the economy appears healthy by aggregate measures while most people get poorer. The Great Decoupling is harder to campaign on because it is structural rather than personal, gradual rather than sudden, and measurable only through macroeconomic data rather than individual stories. Its low cultural visibility means it receives little electoral attention, even as the empirical evidence for it grows (Brynjolfsson & McAfee, 2014).

Case 4: Election outcomes driven by AI assumptions. Assumptions about the long-term impact of ASI are increasingly shaping short-term electoral outcomes. A candidate who promises to 'protect jobs from AI' appeals to the culturally visible displacement scenario. A candidate who promises to 'accelerate AI for national competitiveness' appeals to the AI arms race scenario. In both cases, the candidate's AI position - shaped by the culturally available scenarios - may win them an election, and the resulting administration's policies will then be oriented toward the scenarios that were electorally useful rather than the scenarios that are analytically most probable. The election itself becomes a mechanism by which cultural visibility is converted into institutional priority, potentially increasing the likelihood of the culturally visible scenarios (by directing resources toward them) while simultaneously diverting attention and funding from scenarios that are more probable but less electorally salient. The democratic process, designed to translate public will into governance, becomes in this context a mechanism for translating cultural distortion into preparedness gaps.

5.5.3 The Democratic Dilemma

Democratic Amplification creates a genuine dilemma for AI governance. The democratic principle holds that governance should be responsive to public concern. But if public concern is

systematically distorted by cultural visibility, then democratic responsiveness produces systematically distorted governance. The options are uncomfortable:

Option 1: Technocratic override. Ignore public concern and allocate resources based on expert assessment alone. This produces better-calibrated governance but undermines democratic legitimacy and may itself be captured by expert ideology (as the TESCREAL critique argues).

Option 2: Public education. Correct the distortion by educating the public about the full spectrum of AI scenarios, including the low-visibility ones. This is the approach SuperFutures itself takes. But it is slow, uncertain in its effects, and must compete against the vastly greater resources of the entertainment industry.

Option 3: Institutional design. Create governance institutions that are insulated from short-term electoral pressure and mandated to prepare for the full scenario spectrum, not just the electorally salient scenarios. Independent central banks provide a partial model: they make monetary policy decisions that are responsive to economic conditions rather than to electoral cycles. An analogous 'independent AI preparedness authority' could be mandated to assess and prepare for all scenarios, including the culturally invisible ones, without being subject to the electoral pressure to focus only on what voters fear.

None of these options is fully satisfactory. But the Democratic Amplification mechanism makes clear that the choice is not between action and inaction - it is between governance shaped by cultural distortion and governance designed to account for it. The SuperFutures framework, by making the distortion visible and measurable, provides the empirical foundation for whichever approach a democracy chooses.

6. DISCUSSION

6.1 Contributions

This paper makes five contributions to the literature. First, the SuperFutures Scenario Atlas provides the most comprehensive structured taxonomy of human–ASI future scenarios currently available, with 55 scenario families classified across a consistent, multi-dimensional taxonomy. This addresses a significant gap in the foresight and AI governance literature, where most work focuses on one or a few scenarios rather than the full spectrum.

Second, the Cultural Visibility Hypothesis offers a novel explanation for the widely documented gap between expert and public perceptions of AI risk (Pew, 2023; Eurobarometer, 2024; Zhang et al., 2024). By linking perceived probability to cultural representation rather than to information deficits or cognitive biases alone, the CVH provides a structural, testable explanation that has practical implications for governance and communication.

Third, the theory of Scenario Reflexivity extends the CVH from a theory of perception into a theory of causation. By specifying the mechanism through which belief in a scenario generates the behavioral patterns that make the scenario materially more probable, Scenario Reflexivity provides a framework for understanding AI futures not as predictions to be evaluated but as performative propositions that shape the conditions of their own realization. This has direct implications for how foresight professionals, policymakers, and the public should engage with AI future scenarios: not as passive consumers of forecasts but as participants in a reflexive system where their beliefs and behaviors are themselves variables.

Fourth, the theory of Democratic Amplification identifies the specific institutional mechanism - the electoral process - through which culturally distorted public perception is converted into distorted governance priorities. By showing that democratic responsiveness, in the presence of cultural visibility distortion, produces systematic preparedness gaps, this contribution connects the CVH and Scenario Reflexivity to concrete governance failure modes and proposes institutional design responses.

Fifth, the atlas's inclusion of pop culture mapping as a standard taxonomic field - alongside conventional fields like mechanism, time horizon, and proponent analysis - represents a methodological innovation in foresight practice. Culture is not decorative context; it is a causal variable that shapes what futures are considered, prepared for, and built.

6.2 Relationship to Existing Work

The atlas builds on and extends several existing frameworks. Bostrom's (2014) taxonomy of AI outcome types (extinction, enslavement, utopia, etc.) provides a high-level classification that the atlas decomposes into much finer-grained scenario families. Drexler's (2019) CAIS framework is incorporated as Scenario 54 (AI as Oracle). The Leverhulme Centre's AI narrative taxonomy (Cave et al., 2020) is incorporated into the cultural visibility mapping. The TESCREAL critique (Gebu & Torres, 2024) is incorporated both as a scenario (Scenario 50) and as a reflexive challenge to the atlas's own ideological positioning.

The atlas departs from most existing foresight work in two ways. First, it does not limit itself to a small number of contrasting scenarios (the conventional 2×2 matrix of scenario planning) but aims for exhaustive coverage. This is justified by the exceptional breadth and fragmentation of the AI futures discourse, which cannot be captured in four scenarios without losing most of its content. Second, it includes pop culture as a standard analytical dimension, which no prior foresight atlas has done systematically.

6.3 The Self-Referential Problem

SuperFutures is itself a cultural product that will influence the discourse it maps. By cataloguing and classifying scenarios, it may increase the visibility of some (bringing obscure scenarios to public attention) while inadvertently legitimizing others (giving fringe positions a platform alongside mainstream ones). This self-referential dynamic is acknowledged as a structural feature, not a flaw, of any meta-analytical project. The atlas manages this through transparency: every scenario profile discloses its discourse status (mainstream, emerging, fringe), and the methodology section is explicit about the inclusion criteria.

6.4 Civilizational Scope: The Limits of a Western Atlas

The most significant limitation of the current atlas is not a gap in coverage but a structural bias in the taxonomy itself. The atlas was constructed from within Western intellectual traditions: its scenario types assume individual agency as the unit of analysis, its governance frameworks assume democratic institutions as the baseline, and its pop culture mapping draws overwhelmingly on Hollywood and English-language fiction. This is not merely an omission to be corrected by

'adding' non-Western scenarios. It is an architectural constraint that must be addressed through adaptation of the framework, not just expansion of the content.

At least six major civilizational traditions produce AI future scenarios that are not variants of Western scenarios but genuinely distinct constructions with different foundational premises:

Confucian AI ethics (China) prioritizes social harmony, collective welfare, and hierarchical responsibility over individual rights. A Confucian-aligned ASI scenario would optimize for harmony and accept surveillance as a governance tool - a scenario that is dystopian in Western framing but potentially utopian in Confucian framing. The Social Credit System, understood from within China as a trust-building mechanism rather than Orwellian control, represents a scenario with no equivalent in the current atlas. Liu Cixin's Dark Forest hypothesis produces distinct scenarios for ASI's cosmic posture rooted in civilizational concealment rather than Western competitive expansion.

Russian Cosmism (Fedorov, Tsiolkovsky, Vernadsky) envisions cosmic expansion and even resurrection of the dead as a collective moral obligation, not an individual enhancement choice. This differs fundamentally from Western transhumanism because it is communal, quasi-religious, and duty-based rather than individualist and market-driven.

Hindu and Buddhist philosophical frameworks (India) address consciousness from premises that Western philosophy of mind does not share. The Advaita Vedanta tradition holds that consciousness is fundamental and universal, producing AI scenarios where ASI does not 'achieve' consciousness but becomes a new vehicle for universal awareness. India's caste system interacts with algorithmic systems in ways that Western categories of inequality do not fully capture.

Islamic jurisprudence evaluates technology through the *maqasid al-Shariah* (protection of life, intellect, progeny, wealth, and religion), producing governance frameworks fundamentally different from both Western rights-based and Chinese harmony-based approaches.

Ubuntu philosophy (Southern Africa) produces a communal ethics of technology where AI should optimize for relational wellbeing and mutual obligation rather than individual utility. Afrofuturism offers positive technological visions constructed from African aesthetics and knowledge systems.

Shinto animism (Japan) produces a cultural context in which AI companionship is not dystopian but natural - an expansion of the community of beings. The Japanese cultural relationship with robots is fundamentally different from the Western one, and anime and manga constitute a distinct AI imagination that shapes expectations across East and Southeast Asia.

These frameworks do not merely add scenarios to the atlas; they challenge some of its taxonomic assumptions. The 'Human Position' field assumes individual agency as the unit of analysis. A 'Civilizational Framing' field is needed to state the philosophical tradition from which each scenario emerges. A 'Collective Position' companion field may be required for traditions where the individual is not the primary unit. The Cultural Visibility Hypothesis itself operates differently in non-Western cultural contexts: Bollywood shapes Indian AI perception differently than Hollywood shapes American perception; anime shapes Japanese expectations differently than live-action cinema shapes Western expectations.

Furthermore, the Democratic Amplification mechanism (Section 5.5) operates only in democratic systems. In authoritarian or monarchical governance contexts - which describe the political systems governing the majority of the world's population - the transmission mechanism from public belief to governance priority is different. In China, AI governance is shaped by Party ideology and technocratic assessment, not by electoral dynamics. In the Gulf states, AI strategy is shaped by monarchical vision and sovereign wealth allocation, not by voter concern. The atlas's governance analysis must account for these alternative transmission mechanisms or risk being applicable only to the democratic minority of the global population.

A planned expansion round will add 15 scenarios constructed from within these civilizational traditions through partnership with regional intellectual collaborators. The principle is partners, not projections: scenarios must be constructed from within each tradition by people embedded in it, not by Western researchers applying non-Western labels to Western frameworks. The expansion requires adapting the taxonomy architecture, not just adding content.

7. LIMITATIONS AND FUTURE WORK

The atlas and the CVH are subject to several limitations that should be addressed in future work.

Cultural visibility assessment is currently qualitative. The four-point scale (very high, high, moderate, low) is based on the author's assessment of cultural reference density and impact. A quantitative Cultural Visibility Index - based on box office data, streaming viewership, publication sales, academic citation counts, social media mention frequency, and Google Trends data - is needed to test the CVH rigorously. This is proposed as a priority for future work.

Public perception data is indirect. The CVH is supported by the convergence of high cultural visibility and documented public overestimation for specific scenarios (e.g., extinction), but a direct test requires survey instruments that measure both scenario belief and cultural consumption patterns simultaneously. A cross-national survey is proposed as future work.

Anglo-American and Western structural bias. The current atlas overrepresents English-language, Western, and online-native discourse. More fundamentally, the taxonomy itself embeds Western assumptions: individual agency as the unit of analysis, democratic governance as the baseline, and Hollywood cinema as the primary pop culture reference frame. Non-Western AI futures discourse - including Chinese Confucian AI ethics, Russian Cosmism, Hindu and Buddhist frameworks on consciousness, Islamic maqasid-based governance, Ubuntu communal technology ethics, and Japanese animist technology culture - produces scenarios that are not variants of Western scenarios but genuinely distinct constructions with different foundational premises. A planned expansion round will add 15 scenarios constructed from within these civilizational traditions through partnership with regional intellectual collaborators who are embedded in each tradition. The taxonomy will be adapted with a 'Civilizational Framing' field and alternative formulations of fields that currently assume Western premises.

Single-author construction. The atlas was constructed primarily by a single author, which introduces individual selection and framing bias. An editorial board review process is being established to provide multi-perspective quality assurance.

Scenario boundaries are contested. The decision to treat some scenarios as distinct (e.g., Human Obsolescence vs. Post-Work Society) and others as variants (e.g., different forms of transhumanism) involves judgment calls that others might make differently. The tagged graph structure mitigates this by allowing overlap, but boundary disputes are inherent to any taxonomy of complex phenomena.

Rapid obsolescence risk. AI discourse evolves faster than any static document can track. The atlas is designed as a living document with quarterly updates, but any published version is a snapshot of a moving target.

8. CONCLUSION

The SuperFutures Scenario Atlas provides the first comprehensive, structured, publicly accessible map of the futures being imagined for humanity in the age of artificial superintelligence. With 55 scenario families classified across a consistent multi-dimensional taxonomy, it offers policymakers, researchers, educators, communicators, and the public a navigable framework for an otherwise fragmented and overwhelming discourse.

The atlas's first theoretical contribution - the Cultural Visibility Hypothesis - reveals that the gap between expert and public perceptions of AI risk is not primarily a product of ignorance or irrationality but of the structural dynamics of cultural production. The scenarios that Hollywood dramatizes are the scenarios the public fears; the scenarios that resist dramatization are the scenarios the public ignores.

The second contribution - Scenario Reflexivity - extends the CVH from perception into causation. AI future scenarios are not passive descriptions of possible worlds. They are performative propositions: when enough people believe in a scenario and act accordingly, their aggregate behavior shifts the material conditions that make the scenario more probable. The transhumanist who adopts a neural implant, the post-work advocate who stops investing in traditional career skills, the accelerationist who ships AI without safety constraints, the citizen who resigns to surveillance - each is performing a future into existence through the behavioral consequences of their belief. The complete causal chain runs: Culture → Belief → Behavior → Reality → Culture. The SuperFutures atlas is the first comprehensive framework for tracking all four elements of this loop across the full spectrum of AI future scenarios.

Together, the CVH and Scenario Reflexivity suggest that the most important task for AI foresight is not prediction but imagination and reflexive awareness. If the most probable AI futures are the ones we cannot imagine - because they lack the narrative drama that film and fiction require - then making them visible is not merely informative but materially consequential. And if our

beliefs about AI futures actively shape those futures through behavioral feedback loops, then how we imagine the future is not a luxury but a form of collective action.

The SuperFutures Scenario Atlas is a contribution to both tasks: making the invisible futures visible, and making the reflexive dynamics of belief and behavior legible. It is offered as a living, public resource, open to expansion, critique, and improvement by the research community and the public it is designed to serve.

REFERENCES

- Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2), 3–30.
- Altman, S. (2021). Moore's Law for Everything. <https://moores.samaltman.com/>
- Amodei, D. (2024). Machines of Loving Grace. <https://darioamodei.com/machines-of-loving-grace>
- Andreessen, M. (2023). The Techno-Optimist Manifesto. <https://a16z.com/the-techno-optimist-manifesto/>
- Berry, W. (1987). Why I Am Not Going to Buy a Computer. *New England Review and Bread Loaf Quarterly*, 10(1).
- Birhane, A. (2020). Algorithmic colonization of Africa. *SCRIPTed*, 17(2), 389–409.
- Bostrom, N. (2003). Are you living in a computer simulation? *Philosophical Quarterly*, 53(211), 243–255.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age*. W.W. Norton.
- Buterin, V. (2023). My techno-optimism. https://vitalik.eth.limo/general/2023/11/27/techno_optimism.html
- Callon, M. (2006). What does it mean to say that economics is performative? CSI Working Paper Series No. 005, Centre de Sociologie de l'Innovation, Mines ParisTech.
- Cave, S., Dihal, K., & Dillon, S. (Eds.). (2020). *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford University Press.
- Cave, S., Craig, C., Dihal, K., Dillon, S., Montgomery, J., Singler, B., & Taylor, L. (2018). *Portrayals and perceptions of AI and why they matter*. Royal Society.
- Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Dafoe, A. (2018). *AI Governance: A Research Agenda*. Future of Humanity Institute, University of Oxford.
- Drexler, K. E. (2019). *Reframing Superintelligence: Comprehensive AI Services as General Intelligence*. FHI Technical Report.
- European Commission. (2023). *Artificial Intelligence Act*. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- Gebru, T., & Torres, É. P. (2024). The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday*, 29(4).
- Harari, Y. N. (2016). *Homo Deus: A Brief History of Tomorrow*. Harvill Secker.
- Harari, Y. N. (2024). *Nexus: A Brief History of Information Networks from the Stone Age to AI*. Random House.

- Kahn, H., & Wiener, A. J. (1967). *The Year 2000: A Framework for Speculation on the Next Thirty-Three Years*. Macmillan.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–291.
- Kurzweil, R. (2005). *The Singularity Is Near*. Viking.
- Land, N. (1993). Machinic desire. *Textual Practice*, 7(3), 471–482.
- MacKenzie, D. (2006). *An Engine, Not a Camera: How Financial Models Shape Markets*. MIT Press.
- Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, 8(2), 193–210.
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33, 659–684.
- OECD. (2019). *OECD Principles on AI*. <https://oecd.ai/en/ai-principles>
- Ord, T. (2020). *The Precipice: Existential Risk and the Future of Humanity*. Hachette.
- Pew Research Center. (2023). Public awareness and attitudes about AI. <https://www.pewresearch.org/>
- Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.
- Schwartz, P. (1991). *The Art of the Long View: Planning for the Future in an Uncertain World*. Doubleday.
- Shiller, R. J. (2019). *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.
- Slovic, P. (1987). Perception of risk. *Science*, 236(4799), 280–285.
- Soros, G. (1987). *The Alchemy of Finance: Reading the Mind of the Market*. Simon & Schuster.
- Thomas, W. I., & Thomas, D. S. (1928). *The Child in America: Behavior Problems and Programs*. Alfred A. Knopf.
- Turney, J. (1998). *Frankenstein's Footsteps: Science, Genetics and Popular Culture*. Yale University Press.
- UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>
- van der Heijden, K. (1996). *Scenarios: The Art of Strategic Conversation*. John Wiley & Sons.
- Varoufakis, Y. (2023). *Technofeudalism: What Killed Capitalism*. The Bodley Head.
- Verdon, G., et al. (2022). Notes on e/acc principles and tenets. Substack.
- Wack, P. (1985). Scenarios: Shooting the rapids. *Harvard Business Review*, 63(6), 139–150.
- Weart, S. R. (2012). *The Rise of Nuclear Fear*. Harvard University Press.
- Weyl, E. G., & Posner, E. A. (2018). *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.

- Wilkinson, A., & Kupers, R. (2014). *The Essence of Scenarios: Learning from the Shell Experience*. Amsterdam University Press.
- Yudkowsky, E. (2022). AGI Ruin: A List of Lethalities. LessWrong.
<https://www.lesswrong.com/posts/uMQ3cqWDPHhjtiesc/>
- Zhang, B., et al. (2024). Artificial Intelligence Index Report 2024. Stanford HAI.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism*. PublicAffairs.