Perspective of Sociotechnical Imaginaries

2026 Blue Fellowship Showcase Poster
Mohamed Debbagh*†, PhD Student

"Any sufficiently advanced technology is indistinguishable from magic." - Arthur C. Clarke

"I think the level of danger of technologies is directly proportional to how close they are approximated to magic. With AI, it's very easy to bring it close to magic, right? So it's easier to distance oneself from understanding the socioeconomic dimension underneath" - Anonymous perspective (Tricase et al., 2025)

Technological Sublime

Technological Sublime describes the sentiment of overwhelming awe, wonder, and fear towards technological achievements, in which technology becomes an object of reverence giving it a quasi-religious or mystical stature in public life. It is often studied through the politics of perception in industrial society, as an emotional and cultural configuration that emerges from new social and technological conditions, with each new configuration, partially displaces earlier forms of the sublime.

Role on Nationalism & Societal Identity: It shapes ordinary understandings of progress, virtue, morality, and destiny, often producing a sense of inevitability around technological development (atomic power competition, the space race, and the contemporary "LLM race").

Role on Industrial Society and Development: It is frequently publicly experienced as a demonstration of progress in a variety of industrial sectors through promises of improvement of societal welfare through technological advancements, often reducing complex sociotechnical realities into simplified narratives and decoupling technological development from its social contexts, underlying mechanisms, and power structures. In this sense, it plays a significant role in national AI strategies, shaping policy tools, public discourse, and more recently, the allocation of resources to AI.

Role in Fiction, Narratives, and World Building: It provides a basis for imagining futures as projections of current sociotechnical systems and assumptions about their inevitable trajectory. These sociotechnical imaginaries often appear in fiction as industrial utopias that promise ideal outcomes through technological advancement, or as cyberpunk dystopias that reflect societal fears about exploitation, power, inequality, and the consequences of technological dominance.

Sociotechnical Imaginaries

Industrial Utopia

An optimistic vision of the future driven by technological advancements in various industrial sectors, especially health and automation. Grounded by the idea that "technological advancements and automation will free us from work" through the maximization of production efficiency, resource abundance, and eradication of diseases. Proponents of this sociotechnical imaginary have positive outlooks on technological determinism and is typically envisioned by tech companies, venture capitalists, and technocrats.

In works of fiction Industrial Utopias are often depicted as a facade which masks the implicit problems with themes of discursive closure, mass surveillance, environmental damage, exploited labor (or hidden under class), and/or loss of individuality.

Aesthetic Themes: big factories, megacities, centralized infrastructure and surveillance networks, control over natural resources, clean, polished surfaces, mass transit arteries, engineered simulated environments.

Axiom spaceship from Pixar's WALL-E

Washington DC 2054 from Minority Report

Cyberpunk Dystopia

A pessimistic vision of the future, imagined as hypercapitalism intensified by technological advancements, where highly advanced technology exists alongside deepening inequality and social collapse. Grounded in the idea that technological progress does not inherently produce social progress, cyberpunk instead imagines "high tech, low life" world where automation and digital infrastructure maximize profit, control, and consumption while public institutions erode. Cyberpunk highlights the anxieties of current sociotechnical systems destructive potential over the environment, democracies and human cognition imagined by new wave science-fiction and punk movements originating in the 80's.

In works of fiction, cyberpunk dystopias are depicted as a pervasively integrated system with little possibility for deliberate strategies of social transformation, giving rise to underground rebellious/ resistance groups.

Aesthetic Themes: megacities governed by mega-corporations, surveillance, heavy neon signage, high-tech with decaying infrastructure, underground subcultures, contrasts between elites and street life.

Neo Tokyo from Akira (1988)

Los Angeles 2019/2049 from Blade Runner

Solarpunk

An alternative hopeful vision for the future emerging from recent grassroots cultural movements, solarpunk challenges both the technological pessimism of cyberpunk and the techno-deterministic solutionism of industrial utopian imaginaries. Solarpunk perspectives typically originate from environmental activism, online open-source communities, and post-capitalist thinkers, with ideas grounded in a critique of contemporary socioeconomic systems that promote endless growth, extractive production, and concentrated corporate control. Instead, it argues for regenerative development, community ownership, and equitable access to resources.

In works of fiction, solarpunk futures are often presented as societies rebuilt in the aftermath of ecological and social collapse. Rather than portraying technology as inherently liberatory or inherently oppressive, solarpunk frames it as a situated tool, most valuable when it is transparent, repairable, and democratically governed, supporting decentralized renewable energy, circular material flows, and human-scale infrastructures that coexist with ecological systems.

Aesthetic Themes: decentralization, cities overtaken by natural ecosystem, community cooperation, high-tech integrated with nature, renewable/regenerative infrastructure, soft-futurism.

Valley of the wind from Nausicaä of the Valley of the Wind

† Building 21 Research Institute, 651 Sherbrooke St W, Montreal, Quebec H3A 1E3
* McGill University. Department of Bioresource Engineering, 21111 Lakeshore Rd, Sainte-Anne-de-Bellevue, Quebec H9X 3V9
Contact email: mohamed.debbagh@mail.mcgill.ca

Contextualizing the Social Failures of AI Systems

Mohamed Debbagh*†, PhD Student

Food for thought: Who develops AI systems and how are they currently informed on social values?

The AI Alignment Problem

With the growing sense of the technological sublime surrounding the rapid development and integration of AI systems in daily life, particularly Large Language Models (LLMs) and generative models, AI alignment has become a topic of interest in both academia and public discourse. The broad literature on this topic seeks to address two main questions: (1) "How can society remain sustainable as AI systems become increasing capable?" and (2) "Which values should these systems be aligned with?"

Assumptions and Approach : Underlying these questions is an implicit assumption, AI systems may achieve a level of autonomy, cognition, and normative decision-making that could diverge from socially desirable outcomes, resulting in social failures. This anthropomorphization of AI as encapsulated intelligent tools, or agents, provides a framing for guiding the principled judgement process of AI systems by ascribing a set of values for the agentic systems to learn from, typically operationalized through policies. Accordingly, alignment is currently understood as a policy-mediated process, in which abstract values are translated into enforceable constraints that shape AI behavior across different societal layers. Thus, the majority of literature primarily focuses on approaches for determining relevant value sets, operationalizing these values as policy, and developing mechanisms for policy enforcements within these AI systems.

Values as Policies

Fundamental Values & Moral Norms (Philosophy)

These values are grounded in the fundamental principles of human existentialism. They can be decomposed into core human aims, including survival, sustainable intergenerational continuity, social coexistence, education, and the pursuit of truth.

Sources of value conflict: Value pluralism, moral dilemmas and tragic trade-offs, individual and collective tensions, epistemic uncertainty and moral hegemony, context-dependence of moral meaning.

International Norms (Governance)

These values are derived from fundamental moral norms and grounded in historical context and international governance frameworks to formalize shared standards for acceptable conduct across states, institutions, and societies. They can be decomposed into core aims including the protection of human rights, preservation of peace and security, promotion of international cooperation, maintenance of rule-based order, and reduction of harm through accountability.

Sources of value conflict: normative pluralism across states, conflicts in sovereignty, security vs civil liberties, unequal power and representation in global governance, enforcement asymmetries, dominance through militarism and might is right politics

Conventional Norms (Institutions & Industry)

These values are regulated by international norms and oversight and grounded in internal philosophies, professional standards, and market-driven expectations of acceptable conduct. They can be decomposed into core aims, including efficiency and scalability, reliability and safety, profitability and growth, risk management and compliance, reputation and trust, and the maintenance of organizational stability and competitive advantage.

Sources of value conflict: profit vs public welfare, optimization for measurable metrics over human outcomes, short-term performance pressures versus long-term responsibility, misaligned incentive stakeholders, and institutional normalization of harm

Human-defined value set

Human dignity, autonomy, safety, fairness, equity, solidarity, social welfare, truth, responsibility...

AI constitution & training

Humanity-Aligned AI

Regulatory policy

Human rights, non-discrimination, privacy, transparency, accountability, due process, safety, proportionality...

Policy interpretation & training

Regulation-Aligned AI

Institutional policy

Productivity, performance, cost-effectiveness, innovation, user satisfaction, brand protection, security, reliability, legal defensibility...

Policy interpretation & training

Industry-Aligned AI

† Building 21 Research Institute, 651 Sherbrooke St W, Montreal, Quebec H3A 1E3
* McGill University. Department of Bioresource Engineering, 21111 Lakeshore Rd, Sainte-Anne-de-Bellevue, Quebec H9X 3V9
Contact email: mohamed.debbagh@mail.mcgill.ca

Critique on the Value Alignment Discourse

2026 Blue Fellowship Showcase Poster
Mohamed Debbagh*†, PhD Student

"AI is neither artificial nor intelligent. Rather, artificial intelligence is both embodied and material, made from natural resources, fuel, human labor, infrastructures, logistics, histories and classifications. (...) due to the capital required to build AI at scale and the ways of seeing that it optimizes AI systems are ultimately designed to serve existing dominant interests. In this sense, artificial intelligence is a registry of power" - Kate Crawford, 2021, Atlas of AI

Challenging the Claims of AI Alignment

Much of the AI alignment discourse treats social failure of AI systems as a problem of value specification and enforcement, rather than one of contested values, power, and institutional incentives. As such the following claims posed by alignment discourse are challenged:

1) ***Alignment is primarily a technical control problem:*** AI alignment can be primarily addressed through technical methods such as objective specification, training procedures, and constraint enforcement
2) ***Human values are learnable and operationalizable:*** Human values can be inferred from data and feedback and operationalized into AI decision-making (e.g., AI constitutions, policies, reward functions)
3) ***Institutional governance can constrain rising misalignment risk:*** As AI capability increases, misalignment risk rises, but it can be constrained through regulation and policy deployment within current institutional structures.

Re-contextualizing AI as a Tool: Here we challenge the assumption, reinforced through the framing of the technological sublime, that AI systems should be understood as anthropomorphized agents with autonomy, cognition and normative decision making capacities that may diverge from human value sets. Not because such divergence is impossible, but because we contextualize AI systems as technology, applied knowledge embodied in tools, systems, and processes. As such, the critical question becomes not only whether AI will deviate from "human values," but where this technology is developed, deployed, and maintained, and by whom. From this perspective, alignment is not merely an abstract technical objective, but a socio-technical outcome shaped by the institutions that fund and operationalize AI, the infrastructures that sustain it, and the policies that govern its use.

Examining Some Social Failures of Present AI Systems

Labor

The Guardian
'You feel like you're in prison': workers claim Amazon's surveillance violates labor law

AP News
In Hollywood writers' battle against AI, humans win (for now)

Futurism
CEO of Palantir Says AI Means You'll Have to Work With Your Hands Like a Peasant

Fortune
Ford CEO Jim Farley: Blue-collar labor shortages are hampering AI data center expansion, reshoring

CBC
Amazon cutting 14,000 corporate jobs to spend more on AI

Information

Al Jazeera
Mother of Elon Musk's child sues his AI company over Grok deepfake images

PBS
Musk's Grok AI faces more scrutiny after generating sexual deepfake images

NPR
Elon Musk's AI chatbot, Grok, started calling itself 'MechaHitler'

The Guardian
Cuomo condemned over racist AI ad depicting 'criminals for Zohran Mamdani'

The Guardian
Chris Cuomo mocked for response after falling for deepfake AOC video

Surveillance & IP

The Independent
Inside the AI surveillance firm whose data is being fed to ICE

Gizmodo
Palantir CEO Says a Surveillance State Is Preferable to China Winning the AI Race

CTV News
Microsoft reduces Israel's access to cloud and AI products over reports of mass surveillance in Gaza

Global News
News publishers' copyright lawsuit against OpenAI to go ahead in Ontario

CTV News
'Terrified me': Winnipeg musician claims her voice was used by AI

CBC
Canadian music industry leaders ask government for copyright protections against AI

Environmental Sovereignty

Reuters
Exclusive: Greenland 'Freedom City'? Rich donors push Trump for a tech hub up north

Popular Science
Billionaires dream of building utopian techno-city in Greenland

ABC News
How the White House and governors want to fix AI-driven power shortages and price spikes

The Guardian
US and EU critical minerals project could displace thousands in DRC – report

Axios
The war for minerals, oil and AI
Venezuela and Greenland hold many of the critical minerals essential to AI and defense technology — and therefore future global dominance.

Critiques to each Claim

(1) Alignment is primarily a technical control problem.

As technical researchers approach AI alignment through an engineering perspective, it is often framed as a problem of optimization and control. This framing is frequently predicated on implicit assumptions that: (a) AI systems can be treated as relatively bounded tools whose behavior can be studied independently of broader normative and social context; and (b) misalignment primarily reflects objective misspecification, such that improved reward/loss functions and training regimes will yield aligned behavior.

We challenge these assumptions by arguing that AI systems cannot be meaningfully separated from the social contexts, as evidenced by recurring controversies surrounding AI-driven social failures. Rather than asking only how to align AI systems to human values, we must also ask which actors and institutions AI systems are already aligned to in practice, and how those institutional priorities may diverge from collective societal well-being.

(2) Human values are learnable and operationalizable.

While value-learning approaches suggest that human values can be inferred from data and feedback, this assumption overlooks the extent to which training data is shaped by systematic bias, uneven representation, and historically embedded power relations. Even when values appear learnable in controlled settings, model behavior can diverge in deployment, particularly when systems are optimized for measurable proxies (engagement:shocking, polarizing, or emotionally provocative content may be rewarded because it captures attention). Moreover, scaling value inference often depends on expansive data collection and the normalization of unrestricted data access and surveillance, raising ethical concerns about consent, ownership, and whose voices are ultimately captured and prioritized. As a result, operationalizing values through data may reproduce dominant norms and institutional incentives rather than reflecting a genuinely plural or universally shared moral consensus.

(3) Institutional governance can constrain rising misalignment risk.

This claim assumes that regulatory systems can respond with sufficient speed, scope, and enforcement capacity to shape AI deployment. However, governance is embedded within political economy, where economic forces shape decision-making power, a dynamic reinforced by the post–World War II shift toward neoliberal economic governance. Regulation therefore often functions less as a universal constraint than as a negotiated boundary shaped by powerful actors, lagging behind technological development and fragmented across jurisdictions. Moreover, AI expansion has become a key component in geopolitical competition and resource dependence, linking alignment to broader dynamics of energy demand, supply-chain control, and critical mineral extraction. As a result, institutional governance may constrain certain harms while failing to impose meaningful limits on dominant corporations and elites, raising doubts about whether regulation alone can reliably manage misalignment at scale.

† Building 21 Research Institute, 651 Sherbrooke St W, Montreal, Quebec H3A 1E3
* McGill University. Department of Bioresource Engineering, 21111 Lakeshore Rd, Sainte-Anne-de-Bellevue, Quebec H9X 3V9
Contact email: mohamed.debbagh@mail.mcgill.ca