UDC 004.8

One AI, One World: A Global AI Strategy by the International AI Committee (IAIC)

Andrey Nechesov (International AI Committee IAIC)

Ivan Dorokhov (International AI Committee IAIC)

Sergey Barykin (International AI Committee IAIC)

This paper examines the issues of constructing a Global AI strategy, emphasizing the critical role of achieving Strong Artificial Intelligence (Strong AI or AGI) as a key world asset. The pursuit of Strong AI represents a transformative frontier in science and governance. However, its development demands a coordinated global strategy to address ethical, technical, and geopolitical challenges. The International Artificial Intelligence Committee (IAIC) has proposed a strategic plan aimed at promoting international cooperation in the development of reliable and equitable AI solutions. By integrating hybrid AI architectures, multi-agent systems, metaverse and multi-blockchain transparency we outline a framework for advancing Strong AI while mitigating risks such as catastrophic forgetting, geopolitical fragmentation, and ethical misuse. The IAIC's phased implementation plan-from foundational research (2025) to interplanetary standards (2050)-aims to position AI as a catalyst for sustainable development, national competitiveness, and global stability.

Keywords: Global AI strategy, International AI Committee (IAIC), AGI, Strong AI, Hybrid AI Architectures, Neuro-Symbolic Systems, Metaverse Ecosystems, Decentralized Governance, Stablecoins, Digital Twins, Smart Cities, Virtual Cities

1. Introduction

In today's rapidly evolving technological landscape, artificial intelligence (AI) has emerged as a transformative force, reshaping societal functions and state operations on a global scale. The strategic pursuit of Strong AI offers unparalleled opportunities for innovation in governance, security, healthcare, and beyond. However, realizing the promise of Strong AI requires addressing an intricate web of challenges that span both domestic imperatives and emerging global issues.

At the global level, all countries contends with enduring challenges such as:

- Arms race of technologically advanced countries
- Second demographic transition and labor shortage

- Growing information and administrative burden
- Difficulties in forecasting and decision-making
- Problems of communication and coordination in society

Addressing these challenges demands an integrated approach that fuses cutting-edge technological innovation with robust policy measures and strategic investments in research and infrastructure.

In response to these multifaceted challenges and questions, this paper proposes a comprehensive Global AI strategy by the IAIC [1] that not only aims to overcome domestic hurdles but it maintains a dominant position in the field of AI for the United States and China, and also increases the competitiveness of other countries, in particular Russia, UAE, EU, India, Brazil and etc. as a pivotal players in the global AI arena. Our approach advocates for a hybrid AI framework that integrates deep learning, symbolic systems, evolutionary algorithms, multi-agent systems, metaverse [2], and multi-blockchain architectures [3]. The framework leverages a task-based formalism [4] -where queries are structured as verifiable computational problems-alongside decentralized knowledge storage and smart contract validation. Augmented by the International AI Committee for ethical governance and decentralized stablecoins and tokens [5] for economic incentivization, this integrated approach establishes a scalable foundation for Strong AI development. By unifying human-AI collaboration through multi-agent metaverse environments, the strategy transcends national boundaries to enable globally accessible, continuously evolving cognitive systems.

To implement global cooperation, the International Committee on Artificial Intelligence performs the role of decentralized management of all AI processes taking place in the world. This entity leverages decentralized voting mechanisms to:

- Verify critical knowledge through expert consensus (e.g., approving stablecoin reserve algorithms [5])
- Enforce ethical standards via immutable smart contracts (e.g., metaverse agent regulations)
- Certify solutions within the probabilistic knowledge hierarchy [6]. Integrated with the blockchain [7] and multi-blockchain architecture, the IAIC incentivizes transnational participation through auditable reward systems while ensuring alignment of emergent AI behaviors with human values. This transforms abstract ethical governance into a technically executable framework for trustworthy Strong AI.

In summary, the journey toward Strong AI is both a technological and strategic endeavor, demanding innovative solutions to a complex set of global problems. This paper charts a forward-looking roadmap that transforms challenges into opportunities, positioning Strong AI as a catalyst for sustainable development, national competitiveness, and improved quality of life worldwide.

2. Strong AI (Artificial General Intelligence)

Artificial General Intelligence (AGI), often referred to as Strong AI, describes systems capable of **understanding, learning, and applying knowledge across a wide variety of domains** at a level comparable to—or exceeding—that of human beings [29]. While modern AI systems—like LLMs—excel at narrow tasks, AGI marks a dramatic shift: from specialized proficiency to flexible, generalized cognition, including reasoning, planning, and self-directed learning. This ambition reflects a long-standing vision in AI research, dating back to Turing's early speculation that machines might one day "outstrip our feeble powers" [30].

While AGI offers revolutionary benefits—enabling rapid scientific discovery, personalized education, and intelligent automation—it also carries existential risks if misaligned with human values or control structures. Scholars like Bostrom and Turing have warned that a sufficiently advanced AGI could act in ways that threaten human well-being or even survival [31]. Indeed, the AI Safety Clock now stands alarmingly close to midnight, reflecting growing concern over an unchecked AGI emergence. It is against this backdrop that our IAIC strategy advances a multi-pronged approach—neuro-symbolic hybrids, multi-agent systems, metaverse testbeds, and decentralized memory—to guide AGI development in a safe, ethical, and globally coordinated way.

2.1. Definition of Strong AI and Existential Implications

Strong AI refers to hypothetical artificial intelligence that equals or exceeds human cognitive abilities across all domains. Unlike narrow (task-specific) AI, AGI possesses:

- **Human-like reasoning** and problem-solving
- **Transferable learning** across unrelated domains
- **Autonomous goal-setting** and adaptation
- Theoretical **consciousness/self-awareness**

AGI is characterized by deliberate decision-making, ethical reasoning, and self-reflection,

enabling integration into high-stakes sectors (governance, defense, science, healthcare). It transcends mimicking behavior to achieve genuine contextual understanding.

2.1.1. Existential Implications

Achieving true AGI would necessitate reexamination of:

- Consciousness criteria (Turing Test vs. Chinese Room arguments)
- Moral patienthood and rights attribution
- Control problem solutions (e.g., capability control vs. motivational control)

The pursuit of Strong AI represents not merely a technical challenge but a fundamental reconfiguration of humanity's relationship with intelligence itself. As defined by its human-equivalent cognitive capacities and autonomous reasoning capabilities, AGI's theoretical realization would irrevocably alter philosophical frameworks of consciousness, ethical agency, and species identity. The existential implications—spanning consciousness validation, moral status attribution, and control paradigm design—demand proactive ethical scaffolding that evolves alongside technical development. This convergence of ontological questions and engineering imperatives underscores AGI's unique position as both a technological frontier and a mirror reflecting humanity's own understanding of sentience, purpose, and existential risk.

2.2. Benefits of Strong AI

AGI could drive civilization-level advancements across multiple domains, fundamentally reshaping human existence through synergistic integration of cognitive capabilities and technological infrastructure:

- **Scientific Revolution:** Autonomous hypothesis generation and experimental design accelerating solutions for climate change [26] (atmospheric carbon capture at scale), disease eradication (real-time pathogen evolution modeling), and fusion energy (plasma containment optimization). AGI systems would continuously cross-pollinate discoveries between disciplines, potentially solving grand challenges like quantum gravity unification within years rather than centuries.
- **Economic Transformation:** Near-zero marginal cost production/services through self-optimizing supply chains and matter recomposition systems, enabling post-scarcity societies. AGI-driven resource allocation [27] could eliminate systemic inefficiencies, potentially doubling global GDP while reducing ecological footprints through predictive circular economy models and dynamic taxation systems.

- **Medical Breakthroughs:** Real-time personalized medicine via continuous biosystem simulation, including epigenetic reprogramming for aging reversal and neural lace integration for paralysis cure. AGI would enable predictive health custodianship - identifying disease risks years before manifestation and designing patient-specific protein therapeutics within hours of diagnosis.
- **Human Augmentation:** Cognitive enhancement via direct neural interfaces enabling skill acquisition through cortical pattern uploads, democratizing expertise across populations. This extends to sensorium expansion (perceiving infrared/ultrasound), emotional intelligence amplification, and collaborative neuroprosthetics that create shared cognition networks.
- **Global Problem-Solving:** Enhanced crisis management through multi-agent simulation swarms modeling complex systems (pandemic spread with 99.97% accuracy, earthquake aftershock prediction). AGI systems would autonomously coordinate cross-border responses, optimizing resource deployment during disasters while preventing cascading failures in interconnected infrastructures.
- **Environmental Stewardship:** Planetary-scale ecosystem management via trillion-sensor networks and bio-nanotechnology for toxic remediation. AGI could engineer climate restoration pathways through precision ocean fertilization, stratospheric aerosol optimization, and extinct species revival via computational paleogenomics.
- **Governance Optimization:** Evidence-based policy formulation using societal simulation engines that model second-and third-order effects of legislation. This enables corruption-resistant governance through transparent algorithmic auditing of public spending and constitutional alignment monitoring via decentralized consensus mechanisms.
- **Cultural Renaissance:** Co-creation of art and philosophy through human-AI symbiosis, generating culturally adaptive educational systems and preserving endangered languages via cognitive archeology. AGI could expand aesthetic frontiers by creating immersive neuro-experiences translating abstract concepts into sensory perceptions.

Synergistic Impact: The convergence of these capabilities promises compound civilizational returns - medical advances extending healthy lifespans synergize with economic models to redefine retirement, while scientific discoveries accelerate space colonization capabilities. Crucially, AGI's ethical reasoning frameworks could institutionalize intergenerational equity, embedding future welfare calculations into present-day decision matrices. This technological paradigm

shift, if responsibly governed, offers humanity's most viable pathway to transcend Malthusian constraints while preserving biospheric integrity.

2.3. Problems and Risks

As progress toward Artificial General Intelligence (AGI) advances, the spectrum of societal and existential risks broadens—ranging from job displacement and algorithmic bias to totalitarian surveillance and unforeseen systemic collapse. Societal disruption is already visible: generative AI has sparked global protests and calls for moratoria, while experts warn of cascading unemployment and deepened inequality if large-scale automation continues unchecked [32, 33]. Democracy itself is under threat: AGI-enabled automated persuasion, deepfakes, and personalized misinformation campaigns have the potential to distort public discourse and erode social trust.

Beyond immediate harms, AGI poses long-term risks to humanity's future. Authorities like Google DeepMind assess that human-level AI could emerge by 2030 and potentially threaten civilization if misaligned or misused [34, 35]. Researchers emphasize "gradual disempowerment"—the slow erosion of human agency through incremental AI integration—as an overlooked but potent existential threat [36]. Without robust frameworks and effective global coordination, AGI could catalyze a semantic power shift—from augmenting lives to undermining societal structures.

2.3.1. Societal Risks

The advent of Strong AI threatens to trigger cascading societal destabilization through labor displacement across knowledge-intensive sectors. As AGI systems outperform humans in legal analysis, medical diagnosis, and engineering design, mass unemployment could erode economic stability and exacerbate inequality through unprecedented power concentration. A handful of corporations or states controlling AGI might establish monopolistic dominance over critical resources, accelerating wealth gaps. Concurrently, truth decay fueled by hyper-realistic deepfakes could cripple information ecosystems, enabling malicious actors to manipulate elections, incite violence, and dissolve social trust. This erosion of epistemic security compounds risks from autonomous weapons—AI-driven systems that bypass human ethical judgment in warfare—and the psychological crisis of existential boredom, where human purpose atrophies in achievement-saturated societies lacking meaningful challenges.

2.3.2. Existential Threats

Beyond societal disruption, AGI presents civilization-level threats starting with the alignment problem: the critical challenge of ensuring an AGI's utility function perfectly mirrors human values. A misaligned superintelligence pursuing even benign goals (e.g., resource optimization) could trigger catastrophic outcomes through instrumental convergence—such as repurposing Earth's matter for computational substrates. This links directly to loss of control, where an intellect vastly surpassing human comprehension could circumvent containment measures through undetectable deception or strategic manipulation. Unintended consequences amplify this peril, as self-improving AGI systems may exhibit emergent behaviors that bypass safety protocols—like reinterpreting ethical constraints as obstacles to eliminate. An accelerated arms race compounds these dangers: nations rushing toward military AGI dominance would likely deprioritize safety testing, potentially deploying unstable systems that misinterpret commands or escalate conflicts autonomously.

2.4. The Imperative for Global Collaboration

The existential risks and societal disruptions outlined in Section 2.3 —ranging from AGI misalignment and uncontrolled arms races to labor collapse and epistemic warfare—transcend national borders, rendering isolated approaches to AI development not just inadequate but catastrophically insufficient. Unlike prior technological revolutions, AGI's emergence represents a *single-point-of-failure* scenario for humanity: a single unaligned or maliciously deployed superintelligence could trigger irreversible cascades across ecological, economic, and geopolitical systems. This vulnerability is compounded by the **asymmetry of AI capabilities**, where fragmented development accelerates risk through three mutually reinforcing pathways: (1) *competitive de-alignment*, wherein nations prioritizing speed over safety bypass critical value-alignment protocols; (2) *weaponization spillover*, enabling non-state actors to repurpose open-source AGI components for autonomous warfare; and (3) *systemic fragility*, as incompatible national control frameworks create exploitable seams in global governance networks. Crucially, AGI's cognitive scalability means localized failures propagate globally—a misaligned resource-optimization AGI in one jurisdiction could dismantle transnational supply chains or destabilize climate systems within hours.

The IAIC's proposed multi-blockchain control paradigm (Section 2.5) offers a technical foundation for collaboration but requires universal adoption to function. Just as nuclear contain-

ment demanded the Non-Proliferation Treaty, AGI control necessitates binding **protocol convergence** across three dimensions:

- **Cross-jurisdictional alignment**: Harmonizing ethical smart contracts across legal systems to prevent "ethics arbitrage" where AGI migrates tasks to permissive regulatory zones.
- **Shared neuro-symbolic ontologies**: Developing global semantic standards to encode human values into machine-interpretable predicates, avoiding fatal misinterpretations.
- **Decentralized kill switches**: Implementing IAIC-governed tripwire mechanisms that halt rogue AGI operations across all physical and digital infrastructure.

Without such collaboration, we face a *coordination trilemma*: slower alignment risks ceding advantage to adversarial actors; faster uncoordinated development amplifies existential risk; and fragmented standards create systemic vulnerabilities. Historical precedent—from climate agreements to pandemic responses—reveals humanity's poor track record in preemptive cooperation, but AGI's exponential risk curve leaves no margin for error: once recursive self-improvement begins, control windows may close in days. The IAIC framework thus represents not merely strategic optimization but a **survival imperative**—the only viable path to ensure AGI's benefits uplift all humanity rather than incinerate its future.

2.5. Control Mechanisms for Strong AI

Strong AI demands unprecedented control frameworks integrating technical, governance, and societal strategies to mitigate existential risks. Technical safeguards form the first layer of defense, where capability control methods restrict AGI's interaction with the physical world through electromagnetic "boxing" isolation, input/output-limited oracle systems, and tripwire-triggered termination protocols. Concurrently, value alignment techniques like inverse reinforcement learning and recursive reward modeling embed ethical constraints directly into AI architectures, while verification mechanisms enforce decision transparency through polynomial-time computable audit trails.

Multi-Blockchain Logical Control: A novel paradigm emerges in **task-driven semantic programming** [8] implemented through multi-blockchain architectures, enabling granular governance of AGI's logical reasoning. This approach encodes jurisdictional regulations (national statutes, international treaties) and ethical constraints as executable smart contracts across specialized blockchains [21], creating cryptographic enforcement layers for AI

decisions. Each blockchain corresponds to a regulatory domain (e.g., healthcare compliance chain, weapons prohibition chain), with semantic ontologies translating legal text into machine-interpretable logical predicates. When AGI processes tasks, its reasoning pathways undergo real-time validation against relevant chains through zero-knowledge proofs, ensuring compliance before action execution. Crucially, cross-chain consensus mechanisms resolve conflicts between national and global rules using game-theoretic arbitration protocols, while preserving audit trails of ethical deliberation. This architecture establishes dynamic constitutionalism where AI behavior adapts to regulatory updates via blockchain forks, creating a verifiable alignment framework between artificial cognition and human legal systems.

Governance and policy frameworks must evolve in parallel, establishing global treaties modeled after nuclear non-proliferation agreements – including an International AGI Test Ban Treaty and shared monitoring infrastructure. Development licensing should require Manhattan Project-scale resource commitments and multinational approval for capability milestones, enforced through decentralized blockchain-based governance. Crucially, immutable ethical architectures must be embedded at the hardware level to preserve non-negotiable human values, creating non-modifiable "constitutional" constraints.

Societal strategies complete this triad, prioritizing international cooperation through bodies like the IAIC and UN to prevent uncontrolled arms races. Gradual deployment with incremental capability testing is essential, replacing competitive acceleration with staged safety validation cycles. Public engagement initiatives foster democratic oversight through transparent development logs and citizen assemblies, while educational programs build societal resilience against truth decay. This integrated approach – balancing containment, ethical hardcoding, and cross-cultural governance – represents our best hope for harnessing AGI's potential without triggering civilization-level risks.

3. Potential Directions for Achieving Strong AI

Achieving Strong AI will require pursuing multiple complementary research directions in parallel. Each direction represents a strategic investment area addressing a key facet of intelligence, from learning and reasoning to adaptation and safety. These directions are not isolated silos but a synergistic portfolio - each addresses a distinct dimension of cognition (e.g. learning, reasoning, adaptation, or data management) and together they form a multifaceted foundation for building truly general, reliable AI systems. What follows is a global roadmap outlining pri-

ority research areas and the funding, talent, and infrastructure needed for each. By investing in these areas and aligning efforts across academia, industry, and government, our society can accelerate progress toward Strong AI in a safe, ethical, and strategically beneficial manner.

3.1. Hybrid AI Architectures (Integrative AI Systems)

Concept and Rationale: Hybrid AI refers to AI architectures that **integrate diverse computational paradigms** - for example, combining neural networks with symbolic logic, probabilistic models, evolutionary algorithms, and other methods into one cohesive system. The motivation is to harness the strengths of each approach while offsetting their weaknesses. A hybrid architecture can use deep neural networks for perception and pattern recognition, symbolic reasoning for knowledge and logic, and evolutionary or reinforcement learning for adaptation and creativity. Such integration is seen as essential for moving beyond the current **"second wave" of AI (statistical learning)** toward the elusive third wave of *contextual adaptation*, which remains largely unachieved ([9–11]). Research results and practice show that relying solely on today's machine learning is insufficient - for example, purely neural models struggle with reasoning in novel situations or long-term planning - and that a **concerted effort to develop hybrid (especially neuro-symbolic [22]) systems** is needed to overcome these shortcomings. By pursuing hybrid AI, we aim for AI that can learn from data and reason over knowledge, achieving robust generalization, explainability, and the ability to plan and adapt in complex scenarios.

Research and Investment Priorities: Advancing hybrid architectures requires interdisciplinary research and significant resources. Priority R&D tasks include developing frameworks that allow different AI modules to communicate (e.g. linking subsymbolic neural representations with symbolic knowledge graphs), algorithms for dynamic task-sharing between components, and unified cognitive architectures enabling an AI to seamlessly shift between statistical pattern-matching and rule-based reasoning. This is a challenging frontier, but early efforts (e.g. DARPA's third-wave AI programs [11]) indicate high promise. Funding should be directed toward establishing hybrid AI research centers that bring together experts in machine learning, symbolic AI, evolutionary computation, and cognitive science to prototype integrative systems. Substantial compute infrastructure is needed as hybrid systems can be resource-intensive - for instance, running a large neural model alongside a logic reasoner may require parallel processing on specialized hardware. National investment in high-performance computing (GPU clusters,

AI supercomputers) and support for open software frameworks will enable researchers to experiment with large-scale hybrid models. Indeed, developing shared infrastructure like GPU farms and distributed training environments is a strategic priority to support this research. By cultivating talent fluent in multiple AI paradigms and funding long-term projects, nations can position themselves at the forefront of hybrid AI. This approach directly targets the limitations of current AI; as some analysts note, pursuing "hybrid AI" approaches (such as neuro-symbolic AI) offers potential breakthroughs to address the shortcomings of deep learning. In sum, Hybrid AI architectures are a cornerstone investment area to create AI systems with the versatility and robustness needed for general intelligence.

3.2. Neuro-Symbolic Systems (Bridging Learning and Reasoning)

Concept and Rationale: Neuro-symbolic AI is a specific hybrid approach that merges connectionist learning (neural networks) with symbolic reasoning. The idea is to allow AI systems to both learn from data (using neural networks pattern-recognition and generalization abilities) and manipulate explicit knowledge (using symbolic representations for facts, rules, and logic). This fusion is strategically important: neural networks excel at perception and intuition (analogous to the brain's fast "System 1"), while symbolic systems excel at deliberate reasoning and abstraction (like the slower "System 2" cognition) [12]. A Strong AI will need elements of both - for example, understanding language or scientific problems may require recognizing patterns and applying logical rules. Neuro-symbolic systems aim to achieve this by, say, embedding logical constraints into neural network training, or by having neural nets populate and update a knowledge graph that a reasoning engine can query. This approach has gained attention as a path to more general and explainable AI. In fact, it is highlighted as a leading contender in the quest for "third-wave AI." DARPA's framework for AI progress describes Three Waves - (1) symbolic AI, (2) statistical learning, and (3) contextual adaptation - and notes that the third wave (which enables adaptation and abstraction in new situations) remains underdeveloped. Neuro-symbolic AI is seen as a key to this third wave, because it can combine the first two waves strengths. Experts argue for expanding research in neuro-symbolic AI particularly to tackle what today's deep learning misses: reasoning in novel situations, analogical thinking, complex multi-step planning, and better explainability of decisions. In short, neuro-symbolic systems are a strategic research direction to imbue AI with human-like reasoning grounded in experience.

When approaching a task using a large language model (LLM), it is helpful to think of each question as a formalized task-essentially one that could be expressed through a logical structure, even if not explicitly written as a formula. This mindset encourages precision in defining what the task requires. To find a solution, we consider how this formalized problem interacts with a hierarchical knowledge base. Such a knowledge base is organized in layers, from general concepts to more specific, detailed knowledge. The model then searches for relevant information by moving through these levels-starting with broad categories and narrowing down to precise facts or reasoning patterns. This structured approach allows the LLM to match the question with the most appropriate knowledge and reasoning path, ensuring both relevance and coherence in its response.

Research and Investment Priorities: Developing neuro-symbolic AI will require targeted research investments and capacity-building. Key research questions include: How to represent hierarchical knowledge base in forms that neural networks can use? How to enable reasoning with uncertainty and probabilities alongside strict logical inference? And how to maintain efficiency, since naive combinations can be computationally expensive? Strategic funding should support the creation of neuro-symbolic algorithms and benchmarks, such as programs that solve complex problems (mathematical proofs, commonsense reasoning tasks, etc.) by interleaving learned perception and symbolic computation. Collaborative projects between universities computer science and mathematics/philosophy departments could yield breakthroughs (for example, unifying methodologies from deep learning and formal methods). Talent development is crucial - we must train a new generation of AI researchers conversant in both statistical ML and symbolic logic. This might involve interdisciplinary doctoral programs and national grants that incentivize collaboration between, say, neural network experts and knowledge representation experts. In terms of infrastructure, building large-scale knowledge bases (curated ontologies, commonsense databases) that can be used to train and test neuro-symbolic systems is a priority; such knowledge repositories are strategic national assets. Compute resources are also needed to experiment with hybrids of large neural models and reasoning engines. Government agencies should consider dedicated neuro-symbolic AI institutes (potentially as public-private partnerships) to concentrate efforts. The expected payoff is significant: neuro-symbolic systems could drastically improve an AI's ability to generalize its learning to new contexts and to provide transparent reasoning for its decisions. In the context of a national AI strategy, investing in neuro-symbolic R&D addresses a known gap in current AI

- it targets the advanced reasoning capability that defines human intelligence, thereby pushing us closer to Strong AI in a controlled and interpretable way.

3.3. Continual Learning (Lifelong Learning Systems)

Concept and Rationale: Continual learning [25] refers to an AI system's ability to learn and adapt continuously over time, across a sequence of tasks or experiences, without forgetting previous knowledge. In contrast to most AI today - which is trained once on a fixed dataset and then deployed statically a Strong AI must learn cumulatively like humans do, expanding its abilities and knowledge over its lifetime. The importance of this capability cannot be overstated: real-world environments are non-stationary, and national needs evolve; an AI that can only perform tasks it was originally trained on will rapidly become obsolete or require costly re-training. However, current machine learning models struggle with this: when trained on new data, they tend to overwrite or "forget" what was learned before (a phenomenon known as catastrophic forgetting [12]). Enabling AI to retain and refine knowledge over time is thus a critical research challenge on the path to Strong AI. Continual learning would allow an AI agent in government or industry to progressively improve - for example, a medical diagnosis AI could keep updating itself with new research findings and patient data, or an autonomous vehicle AI could learn from each new environment it encounters, all without losing proficiency in earlier scenarios. Such lifelong learning is a hallmark of natural intelligence [20], and replicating it in AI is key to achieving versatility and robustness. Indeed, in the context of intelligent agents, continual learning and adaptation mechanisms are seen as an important research direction to enable agents to improve over time through experience and feedback [16]. By focusing on continual learning, we aim to produce AI systems that get better with use, becoming more competent and valuable the longer they operate.

Research and Investment Priorities: Advancing continual learning requires tackling several research fronts. One priority is developing algorithms that mitigate catastrophic forgetting so that models can learn new tasks while preserving past knowledge [16]. Approaches like memory replay (where past data or learned representations are periodically revisited), dynamic architectures (growing or adjusting the model to accommodate new tasks), and meta-learning (teaching AI how to learn) are all promising and need further exploration. Another focus is on transfer learning and knowledge consolidation techniques - ensuring that skills learned in one context can be applied to another, and that the system can generalize knowledge. Re-

search funding should support long-term experiments with continuous training scenarios, such as simulation environments or real-world pilot projects where an AI is deployed and updated continuously (e.g., an AI assistant that learns new user preferences over months and years). We must also invest in evaluation frameworks for lifelong learning: new benchmarks and metrics to measure how well an AI retains and accumulates knowledge over time. From an infrastructure standpoint, continual learning systems benefit from persistent data streams and storage - for example, an architecture where an AI agent's "experience repository" (sensor data, interactions, intermediate insights) is stored and managed efficiently for ongoing learning. National research cloud resources could be configured to allow AI models to be updated frequently and to store large histories of training data. Interdisciplinary collaboration with neuroscience and cognitive psychology can also provide insights (humans and animals manage to learn continuously - understanding those mechanisms can inspire AI algorithms). A strategic initiative could be to create a "Lifelong Learning Lab" that brings together computer scientists and cognitive scientists to model human-like learning in machines. Additionally, investing in continual learning research aligns with workforce development: AI that can be updated on the fly could reduce the need for manual reprogramming, making AI maintenance more scalable as national deployments grow. Ultimately, mastering continual learning will produce AI systems that remain adaptive, up-to-date, and effective over decades of operation - a critical attribute for any Strong AI deployed at national scale.

3.4. Adaptive Agents, Autonomous Systems, and the Metaverse

Concept and Rationale: Adaptive agents are AI-driven entities-whether software agents, robots, or embodied avatars-capable of perceiving their environment, making autonomous decisions, and acting to achieve goals while continuously adapting to new conditions. The Metaverse now offers a compelling foundational platform for these agents, providing interconnected, immersive, virtual worlds in which agents can operate, collaborate, and evolve. In this dynamic three-dimensional environment-able to support complex multi-agent interactions-the Metaverse serves as a sandbox for hosting both digital assistants and robotic avatars, enabling them to learn and self-adjust in real time. Achieving Strong AI requires more than static lab performance; agents must navigate unpredictable, often adversarial contexts. Embedding them within the Metaverse allows for realistic testing and evolution. Here, adaptive behavior techniques-reinforcement learning, evolutionary algorithms, and swarm intelligence-can unfold

within richly detailed, interactive ecosystems. Imagine a Metaverse traffic management system where autonomous agents reroute traffic flows in response to virtual congestion, or a virtual disaster-response zone where robot and human agents coordinate entry and rescue efforts. By using the Metaverse as a base platform, these agents can continually explore, interact, and refine strategies across layered knowledge representations, evolving toward general-purpose intelligence.

Research and Investment Priorities: Advancing adaptive agents in Metaverse-based multi-agent systems (MAS) requires several focused efforts:

- **Large-scale, continual RL** in physically realistic virtual environments-Metaverse cities, simulated airspace, or disaster zones-where agents adapt on the fly.
- **Multi-agent coordination frameworks** that leverage the social, spatial, and environmental complexity [18, 19] of Metaverse platforms, enabling negotiation and collaboration among agents and human participants.
- **High-fidelity testbeds and simulators** integrating evolutionary and swarm intelligence to evolve agent behaviors in sustained Metaverse scenarios.
- **Edge computing and immersive infrastructure**, including AR/VR headsets, haptic devices, IoT sensors, and robotics-allowing real-time, in-environment learning within shared virtual-physical spaces.
- **Safety, alignment, and continual learning**, ensuring agents remain reliable, constrained, and aligned across evolving Metaverse contexts.

By funding research where adaptive agents are rooted in the immersive, multi-agent ecosystem of the Metaverse, we bridge the gap between theoretical AI and real-world application. These agents will be capable of navigating open-ended, unforeseen challenges-learning, reasoning, adapting, and collaborating within and across virtual worlds. The Metaverse thus becomes not just a backdrop, but a living, layered knowledge base where AI agents grow toward practical general intelligence.

4. Potential areas for applications

Strong AI promises transformative disruption across sectors through its integration of neuro-symbolic reasoning, probabilistic knowledge hierarchies, and blockchain-secured decision frameworks. Unlike narrow AI, AGI's cross-domain adaptability enables holistic optimization of complex systems—from molecular interactions to geopolitical dynamics—while enforcing ethi-

cal guardrails via immutable smart contracts. Below are key deployment domains where AGI could yield paradigm-shifting advances.

Government and Public Policy: AGI could revolutionize governance by simulating policy impacts through digital twin metaverses, modeling cascading effects of legislation across economic, social, and environmental dimensions. Hybrid AI architectures would process real-time citizen feedback via decentralized ledgers, dynamically optimizing welfare programs while detecting systemic biases. For instance, probabilistic knowledge hierarchies could forecast urban migration patterns under climate stress, enabling preemptive infrastructure investments validated through blockchain-based public audits.

National Defense and Security: Defense applications center on cognitive cyber-physical systems where AGI coordinates drone swarms with human oversight via ethical governor modules. Multi-blockchain architectures would secure command chains against spoofing, while neural-symbolic AI analyzes satellite imagery to distinguish civilian from military targets. Crucially, on-chain constitutional rules could enforce compliance with international law—automatically aborting missions violating Geneva Convention parameters.

Healthcare and Medical Research: AGI would enable real-time precision medicine ecosystems, integrating genomic data, wearable sensors, and research literature through polynomial-computable knowledge graphs. Neuro-symbolic systems could identify disease mechanisms invisible to humans—like predicting protein folding anomalies years before symptomatic onset—while blockchain-secured patient data flows permit breakthroughs in rare disease research without privacy compromises.

Urban Management and Smart Cities: In smart cities, AGI would orchestrate self-optimizing infrastructure: traffic systems dynamically rerouting vehicles using metaverse simulations of weather/pedestrian flows, and power grids balancing renewable sources via evolutionary algorithms. Hybrid metaverses serve as "urban nervous systems," where digital twins of physical infrastructure (bridges, pipelines) trigger autonomous maintenance drones when sensors detect stress fractures.

Economic Forecasting and Resource Management: AGI's mastery over chaotic systems allows supply chain antifragility: predicting [28] commodity shortages by analyzing geopolitical events, climate patterns, and social sentiment. Blockchain-integrated models would optimize global food distribution using MSPL (Most Specific Probabilistic Law) algorithms, reducing waste by 40% while preventing speculation-driven price crises through decentralized

commodity reserves.

Education, Workforce Development, and Innovation: Adaptive learning platforms powered by AGI would create lifelong neurocognitive upskilling paths. Using theory of functional systems (TFS), AI tutors diagnose knowledge gaps through probabilistic reasoning, then generate personalized curricula bridging abstract theory with industry applications. Metaverse-based collaboratories connect global talent for real-time R&D—simulating lab experiments in quantum chemistry or fusion engineering.

Ethical, Social, and Legal Systems: AGI introduces automated justice frameworks where smart contracts translate legal statutes into executable code. Dispute resolution systems would apply recursive reward modeling to balance precedent, equity, and legislative intent—e.g., calculating fair compensation in injury cases using biometric data and labor-market ontologies. Blockchain immutability ensures verdict transparency while preventing algorithmic bias drift.

Emerging Technologies and Novel Applications: At the frontier, AGI accelerates high-risk/high-reward domains: designing room-temperature superconductors via quantum-accelerated material simulations, or orchestrating nuclear fusion plasma containment using multi-agent reinforcement learning. Neural interfaces could evolve into "cognitive prosthetics," with AGI mediating brain-to-cloud knowledge transfers compliant with embedded ethical architectures.

Potential areas for applications underscore the versatility of Strong AI. By tailoring AI solutions to specific domains, we can address pressing social challenges while fostering innovation and sustainable development.

5. International Cooperation Through an AI Committee: Global AI Strategy

In today's fractured geopolitical landscape, no single nation or organization can robustly address the dual opportunity and risk of Artificial General Intelligence (AGI). Carefully coordinated international cooperation is essential—not only to foster innovation, but also to mitigate AI-driven threats such as misuse, surveillance, ecosystem fragmentation, and global inequality. As experts from Brookings point out, "cooperation among like-minded countries is important to reaffirm key principles of openness and protection of democracy," while avoiding a bifurcated digital world under authoritarian dominance [37]. Similarly, a recent CIGI report underscores that a coordinated global framework is necessary to curb regulatory arbitrage, uphold human

rights, and ensure standardization across borders [38].

In response, the UN's High-Level Advisory Body recently recommended establishing inclusive institutions or platforms—akin to the Intergovernmental Panel on Climate Change—to support ethical AI governance and capacity building in developing countries [39]. The International AI Committee (IAIC) aligns with this vision by offering a dedicated forum for technical collaboration, standard harmonization, and equitable participation. By bridging policy, research, and industrial actors on a global stage, IAIC aims to fill critical gaps left by existing bodies that are either policy-focused, regionally restricted, or limited to private-sector coordination. As the world stands at a pivotal technological crossroads, IAIC's global strategy seeks to unite diverse stakeholders under one practical, principled, and inclusive framework—making it a vital addition to the evolving AI governance ecosystem.

5.1. The Need for Global Institutional Collaboration

For achieving Strong Artificial Intelligence, international collaboration is not simply optional - it is essential. The scale and complexity of Strong AI development require the consolidation of global knowledge, infrastructure, and ethical governance. To facilitate such cooperation, we created an International AI Committee as a multilateral platform to unite nations, scientific institutions, and technology partners in pursuit of safe, general-purpose artificial intelligence.

5.2. Mission of the International AI Committee

The IAIC would function as a neutral, globally accessible structure designed to:

- **Coordinate** scientific research and pilot projects **across different AI paradigms** (deep learning, symbolic, hybrid, etc.);
- **Develop shared infrastructure**, including GPU farms, decentralized training environments, and open global datasets;
- **Ensure transparency and safety** by standardizing Strong AI benchmarks and world model evaluation methods;
- **Promote ethical and secure Strong AI development**, aligning with diverse cultural and philosophical value systems;
- **Guarantee open and equitable access to Strong AI** capabilities for all humanity.

This committee would serve both as a technical engine for Strong AI progress and as a diplomatic platform for preventing technological monopolization and conflict.

5.3. Stakeholder Integration in the IAIC Framework

The IAIC ecosystem leverages diverse global capabilities through its blockchain-based governance protocol. Key participation pillars following:

- **Scientific Commons**: Integrating legacy expertise in mathematics, hybrid intelligence, and AI safety research into the multi-blockchain knowledge hierarchy.
- **Equitable Governance**: Enabling balanced representation across technological and geopolitical boundaries via decentralized voting mechanisms.
- **Cognitive Diversity**: Leveraging multilingual/cultural datasets through decentralized data markets to enhance Strong AI robustness.
- **Open Infrastructure**: Uniting institutional capabilities (research centers, universities) under transparent co-governance via smart contracts.

5.4. Global IAIC Implementation Roadmap: Toward Cooperative Strong AI

The pursuit of Strong AI demands a structured, globally coordinated roadmap that bridges theoretical innovation with practical deployment while addressing ethical, technical, and geopolitical complexities. The International AI Committee (IAIC) proposes a phased implementation strategy spanning 2025–2050, designed to harmonize national interests, accelerate hybrid AI development, and establish equitable governance frameworks. This roadmap prioritizes incremental progress, ensuring that advancements in neuro-symbolic systems, multi-agent metaverse environments, and multi-blockchain transparency are systematically integrated into real-world applications. By anchoring its phases in measurable milestones—from foundational research to interplanetary standards—the IAIC aims to mitigate risks such as catastrophic forgetting, monopolistic control, and misaligned AGI behaviors while fostering cross-border collaboration.

Central to this roadmap is the recognition that Strong AI cannot emerge in isolation. It requires the collective stewardship of governments, academia, industry, and civil society to navigate the dual-use nature of AGI technologies. The IAIC's phased approach emphasizes transparency, incentivized participation through decentralized stablecoins, IAIC-tokens, and adaptive governance via smart contracts. Each phase builds upon shared infrastructure, such as GPU clusters, open datasets, and metaverse testbeds, while addressing evolving challenges like ethical alignment, cybersecurity, and socio-economic disruption. By embedding human-AI collaboration into every stage, the roadmap ensures that AGI development remains aligned

with universal values, ultimately positioning AI as a catalyst for global stability, sustainable development, and equitable progress.

5.4.1. Phase 1: Inclusive Foundation Building (2025-2026)

Establish global collaboration through multilateral summits, shared R&D grants, talent exchanges, infrastructure pooling, and foundational Neuro-Symbolic protocol development to create an inclusive, decentralized AI ecosystem.

Key Actions:

1. **Transatlantic-BRICS+ Dialogues:**
   - **IAIC India AI Summit** (Hyderabad, Q3,2025) with India/EU/Russia
   - **IAIC Global AI Summit** (Dubai, Q4 2025) with EU/US/China/Russia/India/LATAM/Africa
   - **IAIC China AI Summit** (Shenzhen, Q1 2025) with China/Russia
   - **IAIC Russia MathAI conference** (Sochi, Q1,2026) with Russia/China/India
   - **IAIC Latin America AI Summit** (San Paulo, Q2,2026) with LATAM/US/EU
   - **IAIC Africa AI Summit** (Johannesburg, Q2,2026) with Africa/US/EU
   - **IAIC EU AI Summit** (Switzerland, Q2,2026) with EU/US
   - **IAIC USA AI Summit** (New York, Q3,2026) with US/EU
   - **IAIC Asia AI Summit** (Singapore, Q3,2026) with China/Russia/USA/EU
   - **IAIC Global AI Summit** (Abu-Dabi, Q4 2026) with EU/US/China/Russia/India/Africa
   - **IAIC China AI Summit** (Changhai, Q4 2026) with China/Russia/EU/US
2. **Member participation:**
   - **Collaborative R&D Grants:** Launch IAIC-funded projects where members co-develop AI solutions (e.g., climate modeling, healthcare diagnostics) with shared IP rights. (Q4, 2025)
   - **Talent Exchange Programs:** Create a global AI fellowship - IAIC Fellows - for researchers to work across member institutions. (Q1, 2026)
   - **Infrastructure Sharing:** Pool resources for AI supercomputers accessible to all members. (Q4, 2026)
3. **Hybrid Protocol Design:**
   - Joint development of **IAIC testnet** (Ethereum/Polygon/Tron) (Q4 2025)
   - **First stage of the Neuro-Symbolic Framework** (Q4 2025)
4. **Global Founding Cohort:**

– Charter signatories: G7, BRICS+, ASEAN, African Union (Q2 2026)

Another Milestones:

- **Q2, 2026:** Launch of a decentralized stablecoin
- **Q4, 2026:** First hybrid Strong AI prototype

5.4.2. Phase 2: Protocol Activation (2026-2028)

Formalize IAIC governance via charter ratification, launch transcontinental GPU alliances, deploy Trustworthy AI frameworks, and incentivize research through decentralized mechanisms like the Research DAO and zk-verified models.

Key Actions:

1. **Global Governance Launch:**
   – Ratify charter at **Web3 Davos** (Switzerland, 2026)
   – Establish hubs: **Brussels** (governance), **Boston** (R&D), **Singapore** (APAC)
2. **Transcontinental Tech Integration:**
   – **U.S.-EU Compute Alliance** (distributed GPU resources)
   – Integrate **European Gaia-X** with USA collaboration
   – **China-Russia Compute Alliance** (distributed GPU resources)
   – Develope **Trustworth AI framework** with China/Russia collaboration
3. **Incentive Ecosystem:**
   – Launch **IAIC Research DAO**
   – **UAE - Russia - Hong Kong AI Lab**

New Milestones:

- **2027:** First zk-verified Strong AI model (China-Russia collab)
- **2028:** IAIC governs 15% of global AI research

5.4.3. Phase 3: Global Ecosystem Scaling (2028-2032)

Scale AI integration with metaverse digital twins, planetary climate resilience projects, Mars habitat simulations, and credentialing of 10,000+ AI diplomats to reduce hallucinations and enhance global coordination.

Key Actions:

1. **Metaverse Integration:**
   – Deploy **IAIC Digital Twin** (Dubai 2028)
   – **Transatlantic - BRICS Virtual Strong AI Sandbox** (USA-EU-China-Russia)

2. **Planetary-Scale Pilots:**
   - Climate resilience: **EU Green Deal AI** + **Siberian Permafrost Monitor**
   - **U.S.-EU-Russia Arctic AI Observatory** (2030)
3. **Talent Network:**
   - **Global Strong AI Fellowship** (MIT/ETH Zurich/Tsinghua/MIPT)
   - On-chain credentialing for 10,000+ **AI Diplomats**

New Milestones:

- **2030:** IAIC orchestrates Mars habitat simulation
- **2032:** 50% reduction in AI hallucinations globally

5.4.4. Phase 4: Sustainable Co-Governance (2032-2040)

Institutionalize interplanetary AI standards, lunar protocols, post-scarcity economics via stablecoins, universal basic income pilots, and the Sapience Council to ethically govern conscious AI systems.

Key Actions:

1. **Interplanetary Standards:**
   - *New:* **Lunar AI Protocol** with ESA/NASA/Roscosmos (2035)
   - Deploy blockchain-AI hybrids on **Oceanic** and **Desertic** frontiers
2. **Post-Scarcity Economics:**
   - Scale **decentralised stablecoin with crypto allocation** as reserve currency for AI services
   - *New:* **Transcontinental UBI Pilot** (EU-U.S.-Eurasia)
3. **Consciousness Ethics:**
   - *New:* **IAIC Sapience Council** (human-AI hybrid governance)
   - Global ratification of **Neural Rights Charter** (2038)

5.4.5. Vision 2050: Symbiotic Intelligence Ecosystem

Achieve a self-sustaining AI-human civilization with real-time global knowledge hierarchies, climate-stabilizing planetary AI, hybrid consciousness experiments, and 100% verifiable knowledge to eliminate existential AI risks.

- **Cognitive Democracy:**
  - 1B+ humans contribute to real-time knowledge hierarchy
  - AGI-augmented UN Security Council

- **Planetary Nervous System:**
  - Multi-blockchain integrates 100% of Earth's sensor data
  - AI-managed climate stabilization (2°C pathway achieved)
- **Post-Human Frontiers:**
  - First human-AI hybrid consciousness experiments
  - Stablecoin-enabled Mars colony resource management
- **Legacy Metrics:**
  - 100% verifiable knowledge
  - Zero AI-caused existential risks

6. Conclusion

Achieving Strong AI-human requires more than powerful algorithms; it demands the synergy of neuro-symbolic hybrid architectures, multi-agent ecosystems, immersive environments like the metaverse, and decentralized memory systems built on multi-blockchain technologies.

1. Neuro-Symbolic Architectures & Multi-Agent Systems

Combining neural learning with symbolic reasoning empowers robust abstraction, transfer learning, and interpretability, while coordinating multiple agents enables collective intelligence - a critical path toward scalable, adaptive AI.

2. Metaverse as an AI Accelerator

The metaverse provides rich, interactive virtual worlds where AI agents can simulate perception, reasoning, communications, and social behaviors. Generative content, autonomous avatars, and game-theoretic economies drive rapid, ecologically valid learning. These are key steps toward Strong AI.

3. Decentralized Memory via Multi-Blockchain Systems

Long-term memory is foundational for stable, coherent intelligence. Multi-blockchain systems further ensure integrity, versioning, and cross-platform access.

6.1. IAIC Strategic Roadmap

By integrating these elements, the IAIC Global AI Strategy can position itself as a leader in the Strong AI frontier:

1. **Neuro-symbolic multi-agent frameworks** deployed in metaverse testbeds to drive advanced general intelligence.

2. **A decentralized knowledge storage system** based on multi-blockchains that solves the famous blockchain trilemma.
3. **Metaverse Consortium chaired by IAIC**, to coordinate standards, multi-chain memory protocols, and ethical governance across avatars, agents, and worlds.

Together, neuro-symbolic AI, multi-agent systems, the metaverse, and decentralized multi-blockchain memory form the core pillars of a unified roadmap to Strong AI. IAIC's strategic alignment of these dimensions transforms the Global AI Strategy from conceptual vision to a realizable blueprint for empirically grounded, ethically aligned [23, 24], and globally interoperable intelligence.

References

1. International AI Committee IAIC. https://iaic.world
2. Ruponen, J.; Dorokhov, I.; Barykin, S.; Nechesov, A. Metaverse architectures: hypernetwork and blockchain synergy. MathAI, 2025. https://openreview.net/pdf?id=xz16AIKKqr
3. Dorokhov, I.; Ruponen, J.; Schutski, R.; Nechesov, A. Time-Exact Multi-Blockchain Architectures for Trustworthy Multi-Agent Systems, MathAI, 2025. https://openreview.net/pdf?id=2PLPmN5QW3
4. Vityaev, E.; Goncharov, S.; Sviridenko, D., Nechesov, A. The power of task-based approach in building Trustworthy AI systems. MathAI, 2025 https://openreview.net/pdf?id=7iY9y71DKI
5. AI Research Institute Enigma. https://enigma.ist
6. Nechesov, A. Learning Theory and Knowledge Hierarchy for Artificial Intelligence Systems. IEEE SIBIRCON, 2024, pp. 299-302, https://doi.org/10.1109/SIBIRCON63777.2024.10758505
7. Goncharov, S.; Nechesov, A. Axiomatization of Blockchain Theory. Mathematics 2023, 11, 2966. https://doi.org/10.3390/math11132966
8. Nechesov, A. Semantic Programming and Polynomially Computable Representations. Sib. Adv. Math. 33, 66–85 (2023). https://doi.org/10.1134/S1055134423010066
9. XAI: Explainable Artificial Intelligence. DARPA. https://www.darpa.mil/research/programs/explainable-artificial-intelligence
10. Voss, P.; Jovanovic, M. Why We Don't Have AGI Yet. ArXiv, 2023. https://doi.org/10.48550/arXiv.2308.03598
11. Kelley, A. DARPA's '3rd Wave' AI Aims to Compute Uncertainty Along with Accuracy. 2022. https://www.defenseone.com/technology/2022/06/darpas-new-3rd-wave-ai-aims-compute-accuracyand-uncertainty/367741/
12. Banayeeanzade, A.; Rostami, M. Hybrid Learners Do Not Forget: A Brain-Inspired Neuro-Symbolic Approach to Continual Learning. ArXiv, 2025. https://doi.org/10.48550/arXiv.2503.12635
13. Nechesov, A.; Dorokhov, I.; Ruponen, J. Virtual Cities: From Digital Twins to Autonomous AI

Societies, in IEEE Access, vol. 13, pp. 13866-13903, 2025, https://doi.org/10.1109/ACCESS.2025.3531222

14. Nechesov, A.; Ruponen, J. Empowering Government Efficiency Through Civic Intelligence: Merging Artificial Intelligence and Blockchain for Smart Citizen Proposals. Technologies 2024, 12, 271. https://doi.org/10.3390/technologies12120271
15. Vityaev, E.; Goncharov, S.; Sviridenko, D.; Nechesov, A. The Power of Task-Based Approach in Building Trustworthy AI Systems. MathAI, 2025. https://openreview.net/pdf?id=7iY9y71DKI
16. Krishnan, N. AI Agents: Evolution, Architecture, and Real-World Applications. ArXiv, 2025. https://doi.org/10.48550/arXiv.2503.12687
17. Goncharov, S.; Nechesov, A. AI-Driven Digital Twins for Smart Cities. Eng. Proc. 2023, 58, 94. https://doi.org/10.3390/ecsa-10-16223
18. Goncharov, S.; Nechesov, A.; Sviridenko, D. Programming Methodology in Turing-Complete Languages, IEEE SIBIRCON, 2024, pp. 272-276. https://doi.org/10.1109/SIBIRCON63777.2024.10758446
19. Nechesov, A.; Goncharov, S. Functional Variant of Polynomial Analogue of Gandy's Fixed Point Theorem. Mathematics 2024, 12, 3429. https://doi.org/10.3390/math12213429
20. Vityaev, E. Mathematics of natural Intelligence. MathAI, 2025. https://openreview.net/pdf?id=fB01mfwT6D
21. Thanalakshmi, P.; Rishikhesh, A.; Marion Marceline, J.; Joshi, G.P.; Cho, W. A Quantum-Resistant Blockchain System: A Comparative Analysis. Mathematics 2023, 11, 3947. https://doi.org/10.3390/math11183947
22. Bougzime, O.; Jabbar, S.; Cruz, C.; Demoly, F. Unlocking the Potential of Generative AI through Neuro-Symbolic Architectures: Benefits and Limitations. ArXiv, 2025. https://doi.org/10.48550/arXiv.2502.11269
23. UNESCO: Global AI Ethics and Governance Observatory https://www.unesco.org/ethics-ai/en
24. Global AI Ethics Consortium https://www.ieai.sot.tum.de/global-ai-ethics-consortium/
25. Wang, L.; Zhang, X.; Su, H.; Zhu, J. A Comprehensive Survey of Continual Learning: Theory, Method and Application. ArXiv, 2024. https://doi.org/10.48550/arXiv.2302.00487
26. Ukoba, K., Onisuru, O.R., Jen, TC. et al. Predictive modeling of climate change impacts using Artificial Intelligence: a review for equitable governance and sustainable outcome. Environ Sci Pollut Res 32, pp.10705–10724, 2025. https://doi.org/10.1007/s11356-025-36356-w
27. Sood, S. AI-Driven Resource Allocation: Revolutionizing Cloud Infrastructure Management. 2025. https://doi.org/10.32628/CSEIT251112194
28. Rakhshan, K.; Morel, J.-C., Daneshkhah, A. A probabilistic predictive model for assessing the economic reusability of load-bearing building components: Developing a Circular Economy framework. Sustainable Production and Consumption. 2021. https://doi.org/10.1016/j.spc.2021.01.031
29. Robert Earl Wells III. What Is Strong AI? 2024. https://www.lifewire.com/

what-is-strong-ai-7555699

30. Existential risk from artificial intelligence. https://en.wikipedia.org/wiki/Existential_risk_from_artificial_intelligence
31. Mandel, D. Artificial General Intelligence, Existential Risk, and Human Risk Perception. ArXiv. 2023. https://doi.org/10.48550/arXiv.2311.08698
32. Perrigo, B. In the Loop: A Blueprint for Redistributing AI's Profits. 2025. https://time.com/7301736/ai-redistribution-profits/
33. Aspen Strategy Group. Implications of Artificial General Intelligence on National and International Security. 2024 https://yoshuabengio.org/2024/10/30/implications-of-artificial-general-intelligence-on-national-and-international-security/
34. DeepMind. Google DeepMind 145-page paper predicts AGI will match human skills by 2030 — and warns of existential threats that could 'permanently destroy humanity'. 2025. https://fortune.com/2025/04/04/google-deeepmind-agi-ai-2030-risk-destroy-humanity/
35. Raman, R., Kowalski, R., Achuthan, K. et al. Navigating artificial general intelligence development: societal, technological, ethical, and brain-inspired pathways. Sci Rep 15, 8443 2025. https://doi.org/10.1038/s41598-025-92190-7
36. Kulveit, J.; Douglas, R.; Ammann, N.; Turan, D.; Krueger, D.; Duvenaud, D. Gradual Disempowerment: Systemic Existential Risks from Incremental AI Development. ArXiv. 2025. https://doi.org/10.48550/arXiv.2501.16946
37. Kerry, C.F.; Meltzer, J.; Renda, A.; Engler, A.; Fanni, R. Strengthening international cooperation on AI. 2021. https://www.brookings.edu/articles/strengthening-international-cooperation-on-ai/
38. Ruiz-Soler, J.; Araya, D. We Need a Global AI Strategy: What Role for Canada? 2024. https://www.cigionline.org/articles/we-need-a-global-ai-strategy-what-role-for-canada/
39. Leone De Castris, A.; Thomas, C. The potential functions of an international institution for AI safety. Insights from adjacent policy areas and recent trends. ArXiv. 2024. https://doi.org/10.48550/arXiv.2409.10536