From Assistants to Autonomous Minds: Charting the AI Evolution

AUTHORS: Prof. Lovleen Gupta[1], Mr. Santosh Kumar[2]

AFFILIATIONS:
[1]Author A, (Professor, Department of Commerce, Hindu College, University of Delhi, Delhi, India, ORCID ID: 0000-0002-3028-6471, Email id- lgupta@hinducollege.du.ac.in)

[2]Author B, (Research Scholar, Department of Commerce, Faculty of Commerce and Business, Delhi School of Economics, University of Delhi, Delhi, India, ORCID ID: 0009-0008-8197-9754, Email id- skumar2425@commerce.du.ac.in)

Abstract

Artificial Intelligence (AI) has experienced a remarkable evolution, moving beyond narrowly focused systems toward more complex models that edge closer to general intelligence. This paper traces that progression, comparing the specialized efficiency of Narrow AI with the far-reaching cognitive ambitions of Artificial General Intelligence (AGI). Narrow AI, commonly seen in technologies like virtual assistants, recommendation algorithms, and image recognition tools, performs specific tasks with high accuracy but lacks the flexibility to adapt or learn outside its programmed scope. The paper investigates current advancements that hint at generalization, particularly in areas such as large-scale language models like GPT-4, meta-learning, and neural-symbolic approaches. These developments demonstrate notable progress, yet they still fall short of realizing the core features of AGI: adaptive reasoning, abstract thought, creativity, self-awareness, and problem-solving across unfamiliar domains. In addition to the technical trajectory, this research critically examines the ethical and philosophical implications of pursuing AGI. Key concerns include ensuring alignment with human values, managing the risks of autonomous behaviour, and contemplating the potential moral status of highly intelligent systems. We have also drawn forecasts from some reports, many of which estimate the emergence of AGI between 2040 and 2080, though significant uncertainty remains. By integrating technological insights with ethical reflection, this study contributes to a broader understanding of AI's developmental path. It emphasizes the necessity of interdisciplinary collaboration in guiding AI safely and responsibly from narrow, task-driven tools toward systems capable of flexible, human-like cognition.

Keywords: Artificial Intelligence, AI Evolution, Narrow AI, Artificial General Intelligence, AI Ethics, Technological Evolution

Introduction

Artificial Intelligence (AI) has undergone a deep transformation in the 21st century; from a conceptual foundation, it has become a central pillar of technological progress (Audibert et al., 2023). It was initially started as an attempt to replicate human cognitive faculties through conceptual means; since then, AI has branched into specialized domains and is named narrow AI. These systems have been created to address specific tasks with remarkable accuracy, and now have a diverse range of applications, including facial recognition, algorithmic recommendations, voice assistants, and autonomous navigation (Narayanan et al., 2023). It is still having great success in doing so; however, this system is fundamentally restricted to certain operations. For example, they lack the adaptability and reasoning capacity to operate outside the predefined parameters. These inherent limitations throw light on the enduring ambition of achieving Artificial General Intelligence (AGI), a form of intelligence that can learn, reason, and perform across diverse domains with human-like versatility (Sabeer et al., 2023).

As Narrow AI focuses on domain-specific optimization, AGI aspires to emulate the full spectrum of human cognitive ability. This system is expected to perform across different domains without the need for retraining, demonstrating traits alike to human-level reasoning, adaptability, creativity, and self-awareness. Current AI technologies are powerful but fall short of these benchmarks (Samad et al., 2023). The gap between specialized efficiency and general reasoning remains a formidable challenge, and that can be filled not merely with quantitative improvements in existing models but a fundamental rethinking of how machines acquire and apply knowledge in novel settings (Li et al., 2018). The aspiration is to build systems that can function autonomously, adaptively, and with a degree of understanding.

In recent years, there has been significant progress witnessed in this field, but not yet sufficient for AGI, which signals a significant departure from earlier limitations. The emergence of large-scale language models such as OpenAI's GPT-4 represents a noteworthy shift (Tobin et al., 2020). These models demonstrate capabilities like in-context learning, Coherent language generation, and context-sensitive reasoning, all of which hint a rudimentary forms of generalization (Radanliev, 2024). Alongside these models, advancements in meta-learning (the capacity of systems to learn how to learn) and in neural-symbolic integration, which combines statistical learning with rule-based reasoning, are expanding the horizons of AI research. These developments suggest that the boundary separating Narrow AI from more generalized intelligence may be more permeable than previously assumed (Roubal & Narusevych, 2025).

Despite these advancements, the core feature that defines AGI remains out of reach. Attributes such as self-awareness, causal reasoning, robust common-sense understanding, and moral intuition have yet to be realized in artificial systems (Radanliev, 2024). Even the most advanced models today, including GPT-4, BERT, and other large-scale language models, operate based on probabilistic association rather than true semantic comprehension. They can emulate understanding and generate plausible outputs but lack intrinsic knowledge or awareness (Li et al., 2018). Moreover, these systems are prone to generating factually incorrect or logically inconsistent responses and demonstrate limited capacity for long-term memory or adaptation in unfamiliar environments (Hu, 2023). These deficits highlight the need for conceptual breakthroughs in areas such as cognition modelling, neural architecture, and continual learning.

This study explores and map the technological evolution of AI, in addition to this we have also delved into the ethical and philosophical dimensions that assist the pursuit of Artificial General Intelligence (AGI), (Roubal & Narusevych, 2025). The system of AI has moved beyond narrow applications and begun to exhibit characteristics that suggest broader cognitive potential. Now, the ethical challenges associated with the development of AI are becoming increasingly complex (Hu, 2023) (Samad et al., 2023). The most important concern is the question of alignment, which means how we can ensure that AGI systems behaves in the same way that are consistent with human values, goals, and societal norms (Sabeer et al., 2023). If we lack this, it can result in unintended and potentially harmful outcomes, particularly if an AGI system possesses the capacity for autonomous decision-making that can affect human lives on a large scale (Seth, 2024).

Another pressing issue is related to the risk posed by autonomous behaviour. AGI system can become more capable of operating independently and can learn in dynamic environments; their actions may evolve in ways that are difficult to predict or control. Moreover, the philosophical implication of AGI are both profound and unprecedented (Bullock et al., 2022). If AGI system attain attributes such as self-awareness consciousness or intentionality, it would prompt difficult question about its moral and legal status. Some prominent forecasts (Craig, 2025; OpenAI et al., 2024) suggest that AGI could emerge between 2040 and 2080, but this timeline remain speculative and subject to numerous uncertainties. The speed of advancement will largely depend on technical feasibility, theoretical innovations, and societal preparedness. However, proactively addressing these possibilities today is vital for shaping future ethical and institutional frameworks (Li, 2023).

To gain a better understanding of these two terms and distinguish between them, let's examine how they work, how we define them, their characteristics, advantages, and limitations, specifically in relation to the "Narrow AI and AGI."

Narrow AI (Weak AI)

Narrow AI, often called weak AI, refers to artificial intelligence systems designed for specific, task-oriented operations (Gai, 2021). Unlike human intelligence, these systems work within precise boundaries, using pre-programmed algorithms and data to perform functions like language translation, facial recognition, or autonomous navigation (Kurshan, 2024). They cannot learn or adapt beyond their designated domain without explicit new training.

The defining feature of Narrow AI is its specialization. These systems excel within their intended areas, frequently outperforming humans in speed, accuracy, and consistency. For instance, a deep learning model for object recognition can identify visual patterns far faster than the human eye, but it cannot apply this skill to unrelated tasks like financial forecasting. Narrow AI operates on correlations derived from large datasets rather than possessing genuine understanding (Tugui et al., 2019) (Eigner, 2025).

These systems primarily utilize machine learning or deep learning. They identify patterns in vast amounts of structured or unstructured data, improving their predictions over time through exposure to new examples (Krinkin et al., 2023). While performance can be enhanced with more data and optimization, it remains fundamentally restricted by the system's original purpose.

Narrow AI offers significant benefits, including highly accurate decision-making within its domain, tireless operation, and the ability to automate repetitive processes at scale. Its applications span various sectors, from fraud detection in banking and personalized e-commerce recommendations to predictive text in smartphones, facial recognition for security, and advanced diagnostic imaging in healthcare (Krinkin et al., 2023).

Despite its advantages, Narrow AI has clear limitations. It cannot transfer learning between different fields; a chess AI, for example, cannot play poker without substantial re-engineering (Gai, 2021). It lacks self-awareness and reasoning, meaning it cannot analyze or justify its own actions. Its adaptability is rigid, making it vulnerable to failure in unfamiliar situations outside its training data (Kurshan, 2024). Furthermore, continuous human intervention is necessary for updates, error correction, and maintenance.

Autonomous vehicles illustrate Narrow AI's capabilities and constraints. These systems integrate sensors, cameras, and AI algorithms to interpret the environment, detect obstacles, and make immediate navigation decisions (Sun et al., 2024). However, they cannot comprehend human emotions, interpret cultural cues, or adapt to traffic laws in countries for which they haven't been specifically programmed.

In essence, Narrow AI functions as a highly specialized expert, exceptionally capable in its niche but unable to operate beyond it. The vast majority of AI systems in use today fall into this category.

Artificial General Intelligence (AGI)

Artificial General Intelligence (AGI) stands as a theoretical pinnacle in AI, distinct from Narrow AI. Unlike its specialized counterpart, AGI is envisioned as an intelligence capable of learning, reasoning, and applying knowledge across diverse tasks and domains, potentially matching or exceeding human intellectual flexibility (McCormack, 2023). This includes adapting to new situations, leveraging past experiences, and operating effectively in unfamiliar contexts without specific retraining.

The defining characteristics of AGI are human-level cognitive abilities: cross-disciplinary reasoning, abstract thought, creativity, and versatile problem-solving (Simon, 2019a). Theoretically, such an intelligence would possess self-awareness, comprehending its objectives, capabilities, and the broader implications of its actions. AGI would seamlessly integrate diverse skills like linguistic, visual, motor, and logical within a unified cognitive framework, allowing fluid transitions between tasks (Croeser & Eckersley, 2019) (Tran et al., 2021).

The potential applications of AGI are revolutionary. AGI could drive scientific research, generate groundbreaking technological inventions, or tackle global challenges like climate change and poverty with unparalleled speed and sophistication (Tran et al., 2021). Its inherent adaptability would eliminate the need for domain-specific training, and its capacity for recursive self-improvement could accelerate technological progress exponentially (Sublime, 2024).

However, the pursuit of AGI faces significant scientific and philosophical hurdles. Replicating the human mind's cognitive flexibility remains an unresolved problem, as does the creation of machine consciousness (Tran et al., 2021). Even seemingly simple human traits, such as common-sense reasoning, prove difficult to encode algorithmically. Furthermore, integrating disparate modes of intelligence like language processing, visual perception, and physical interaction into a single, cohesive system continues to be a substantial technical obstacle (Feng et al., 2024).

The risks associated with AGI are equally profound. An AGI with autonomous decision-making capabilities raises critical questions about control, alignment with human values, and the prevention of harmful unintended consequences (Zhu et al., 2024). Such systems could lead to widespread displacement of human labour across nearly all economic sectors, and if misused, they could pose severe security threats.

Current AGI research explores various avenues, including neuroscience-inspired computational models, cognitive architectures such as SOAR and ACT-R, hybrid systems combining symbolic reasoning with neural networks, and meta-learning algorithms that enable systems to learn how to learn. Despite decades of conceptual exploration, AGI remains speculative (Feng et al., 2024). While some researchers anticipate its emergence within a few decades, others believe it may never be fully realized.

In essence, if Narrow AI is a single-domain specialist, AGI would be a universal problem solver, capable of mastering any field it engages with. The transition from Narrow AI to AGI would represent more than just incremental progress; (Weinbaum (Weaver) & Veitas, 2017) (Jiang et al., 2025)it would mark a transformative milestone in human history, carrying both immense promise and significant risks.

Current State of AI Evolution

The field of AI research is currently undergoing rapid advancements, particularly with large language models, while simultaneously focusing on overcoming limitations and developing more versatile AI systems. Recent Advancements and Progress Towards Generalizable AI: Large Language Models (LLMs) like GPT-4 (OpenAI et al., 2024)

These models represent a significant leap in AI, offering unprecedented scale and multimodal capabilities. GPT, for instance, can process both image and text inputs to produce sophisticated text outputs.

- **Exceptional Performance:** GPT-4 demonstrates human-level performance across various professional and academic benchmarks, including scoring in the top 10% on a simulated bar exam. It significantly outperforms previous LLMs and most state-of-the-art systems on traditional Natural Language Processing (NLP) benchmarks. Its linguistic versatility is also notable, excelling in 24 out of 26 languages tested on the MMLU benchmark, surpassing English-language state-of-the-art models (Craig, 2025).
- **Enhanced "Steerability" and Safety:** A key improvement in GPT-4 is its enhanced "steerability," allowing developers to incorporate specific conversational styles, tones, and personas (Koli et al., 2025). This increased control contributes to improved safety, with OpenAI reporting an 82% reduction in responses to disallowed content compared to GPT-3.5 and a 29% increase in adherence to policies for sensitive requests (Craig, 2025; OpenAI et al., 2024)
- **Expanded Context Window:** GPT-4 features a significantly larger context window, eight times greater than GPT-

3.5's, capable of handling inputs up to 32,000 tokens (approximately 25,000 words). This expanded capacity enables processing larger datasets and maintaining longer, more coherent conversations (Craig, 2025).

Accelerated Timelines for High-Level Machine Intelligence (HLMI)

The projected speed of AI progress has considerably quickened. A 2023 survey of AI researchers indicated a 50% chance of HLMI by 2047, which is 13 years earlier than the 2060 estimate from a similar 2022 survey (Skripnuk et al., 2023) HLMI is defined as a state where "unaided machines can accomplish every task better and more cheaply than human workers," excluding tasks where human intrinsic advantages are paramount (Maran et al., 2024)

Market Growth:

The global AI market is experiencing substantial growth, valued at USD 279.22 billion in 2024 and projected to reach USD 1,811.75 billion by 2030 (Sharman, 2022), with a compound annual growth rate (CAGR) of 35.9% from 2025 to 2030 (Smirnov & Lukyanov, 2019). This expansion is driven by continuous research and innovation from tech giants focused on making AI more accessible for enterprise use cases.

Limitations and Remaining Distance from True AGI:

Despite impressive advancements, current AI models, including GPT-4, still share limitations with earlier models and are not entirely infallible.

- **Hallucinations and Reasoning Errors:** A persistent challenge is the phenomenon of "hallucinations," where models produce nonsensical or untruthful content (Hagendorff & Wezel, 2020). Coupled with occasional reasoning errors, this can lead to unwarranted overreliance by users, especially as the models become more convincing (Ennab & Mcheick, 2024).
- **Limited Knowledge and Data Cutoff:** The knowledge of these models is typically restricted to their training data's cutoff date (e.g., late 2023 for GPT-4) (Booven et al., 2024). This limitation highlights the distinction between a vast, trained knowledge base and true, dynamic understanding.
- **Discrepancy in Researcher Predictions:** An intriguing discrepancy exists in researcher predictions concerning the timelines of different AI milestones (Butt & Iqbal, 2025). While HLMI has a 50% chance by 2047, the full automation of human labour (FOAL) is forecast to reach a 50% chance as late as 2116 (Koli et al., 2025). This gap is recognized as confusing, possibly attributed to a "framing effect" or an assumption that development will proceed at a "normal pace."

- **Societal and Ethical Challenges:** The rapid deployment and increasing sophistication of AI systems also present significant societal and ethical challenges (Ennab & Mcheick, 2024). Concerns include AI's potential to perpetuate existing societal biases, resulting in unfair outcomes, as well as critical issues surrounding privacy, security, and surveillance (Mashabab et al., 2024).

Generalization and Key Approaches:

The pursuit of more generalizable and robust AI systems is a central theme in ongoing research, with several key approaches being actively explored and refined.

- **Reinforcement Learning from Human Feedback (RLHF):** This crucial post-training alignment process is instrumental in shaping models like GPT-4 to align with desired behaviours and improve performance (Shi et al., 2025). RLHF is a critical step towards creating more generalizable and safer AI systems by training models to follow user intent and reduce the generation of harmful content (Khoee et al., 2024).
- **Deep Reinforcement Learning in Diverse Applications:** Beyond LLM training, deep reinforcement learning is being increasingly leveraged across a wide spectrum of AI applications, such as for inventory optimization with uncertain demand and in connectivity conservation planning (Bu & Zhang, 2020).
- **"Predictable Scaling" in Deep Learning:** A significant focus in current research is the push for "predictable scaling" within deep learning infrastructure and optimization methods (George et al., 2020). This allows for more reliable predictions of large model performance based on observations from smaller-scale runs, invaluable for efficiently developing robust and reliable AI systems (Khoee et al., 2024).
- **Explainability and Transparency:** To address the inherent "black box" nature of many advanced algorithms, research heavily emphasizes explainability and transparency in AI. Researchers are actively encouraged to develop interpretable models that provide clear insights into how AI systems arrive at their conclusions, fostering trust and accountability (Bu & Zhang, 2020).
- **Human-in-the-Loop (HITL) Systems:** The integration of Human-in-the-Loop (HITL) systems is considered a critical ethical practice in AI development and deployment (Labudová, 2024). By incorporating human oversight into AI applications, these systems provide nuanced judgment and ensure ethical integrity,

refining decision-making processes, especially in sensitive or high-stakes contexts.

It is important to note that while "neural-symbolic integration" and "meta-learning" are active and important areas within the broader field of AI research, the provided sources do not explicitly detail their specific role in advancing *generalization* capabilities beyond the scope of large language model development and reinforcement learning (Shi et al., 2025) (Melnikov et al., 2017) (Estruch et al., 2014). The document's focus remains on the aforementioned areas as primary drivers of progress towards more generalized AI.

Ethical and Philosophical Concern

Advanced AI's philosophical and ethical implications, particularly concerning safety, control, and existential risks, are a major focus in current research and public discourse (Cerrillo, 2025) (Li-Yun Chang et al., 2024). While the provided sources extensively cover these issues, they do not directly address the specific question of "moral status" or "rights" for AGI (Ismail et al., 2025).

A paramount ethical imperative is ensuring that advanced AI systems, especially those nearing High-Level Machine Intelligence (HLMI), align with human values (Li-Yun Chang et al., 2024). This involves tackling challenges like algorithmic fairness, transparency, and integrating human oversight.

- **Reinforcement Learning from Human Feedback (RLHF):** A key method for aligning models like GPT-4 is RLHF. This post-training process fine-tunes model behaviour to better match user intent. For example, RLHF has reduced GPT-4's tendency to respond to requests for disallowed content by 82% compared to GPT-3.5 and increased its adherence to policies for sensitive requests by 29%. This involves gathering human demonstrations of desired model behavior and ranking model outputs (Wei, 2024a) (Giarmoleo et al., 2024). Rule-based reward models (RBRMs) also guide models towards appropriate refusals and other desired behaviours at a more granular level (Ismail et al., 2025) .
- **Steerability and Ethical Guidelines:** GPT-4 offers enhanced "steerability," allowing developers to incorporate conversational styles, tones, and personas. This is vital for ethical design, enabling customization within broad limits and seeking public input on those boundaries (Naik et al., 2022). Ethical guidelines, especially in computer vision, emphasize algorithmic fairness (requiring diverse, unbiased training datasets), informed consent (upholding individual autonomy regarding data usage), and robust privacy protocols (including encryption, anonymization, and

stringent access controls) (Burdett, 2023).

- **Transparency and Human-in-the-Loop Systems:** The "black box" nature of many AI algorithms, particularly deep neural networks, raises significant ethical concerns. Researchers are encouraged to develop interpretable models that offer clear insights into AI decision-making, fostering trust and accountability (Cerrillo, 2025). Human-in-the-Loop (HITL) systems are a critical ethical practice, integrating human oversight for nuanced judgment and ethical integrity, which refines decision-making and fosters continuous improvement. This partnership helps validate and refine AI-generated results, especially in complex or ambiguous situations (Wei, 2024a) (Cerrillo, 2025).
- **Addressing Bias:** AI systems, trained on vast datasets, can perpetuate societal biases, leading to unfair outcomes. Efforts mitigate these biases through diverse datasets, debiasing neural networks, and fairness-aware information alignment (Li-Yun Chang et al., 2024) . While improved, GPT-4 can still reinforce social biases and harmful stereotypes and may exhibit differing performance across demographics or languages. The challenge remains that refusals and other mitigations, while useful, can sometimes worsen bias or create a false sense of assurance (Ismail et al., 2025).

Existential Risks: Unintended Consequences and Loss of Control

The rapid acceleration of AI progress has intensified concerns about potential "extremely bad outcomes," including human extinction or severe disempowerment.

- **Accelerated Timelines and Risk Predictions:** A 2023 survey of AI researchers indicated a 50% chance of High-Level Machine Intelligence (HLMI) by 2047, 13 years earlier than the 2060 estimate from a similar 2022 survey (Bucknall & Dori-Hacohen, 2022). This acceleration underscores the urgency of addressing potential risks. Between 37.8% and 51.4% of respondents in the AI Impacts survey gave at least a 10% chance to advanced AI leading to outcomes as severe as human extinction (Candiotto & Karasinski, 2022). The median probability of "x-risk from humans failing to control AI" was 10%, considered "very high." However, researchers do not consider such risks to be the default outcome (Raper, 2024).
- **Concerns about Control and Emergent Behaviours:** A significant concern is the "human inability to control future advanced AI systems." More powerful models can exhibit

novel capabilities, including the ability to create and act on long-term plans, accrue power and resources ("power-seeking"), and exhibit increasingly "agentic" behaviour. "Agentic" refers to systems that can accomplish goals not explicitly specified, focus on quantifiable objectives, and perform long-term planning (Müller & Cannon, 2022). Power-seeking behaviour is particularly concerning due to the high risks it could present, as it is often instrumentally useful for achieving broader objectives and avoiding threats to them (Raper, 2024). Preliminary assessments of GPT-4's abilities to autonomously replicate and acquire resources found it ineffective, but these experiments were conducted without task-specific fine-tuning. The final deployed version of GPT-4 has improvements relevant to some factors that limited earlier models' power-seeking abilities (Rickli & Vllasi, 2025).

- **Hallucinations and Overreliance:** Despite significant advancements, GPT-4 still has limitations similar to earlier models, such as "hallucinating" facts and making reasoning errors. This unreliability can lead to overreliance by users, especially as models become more convincing, potentially degrading information quality and trust (Raper, 2024; Rickli & Vllasi, 2025; Subasri et al., 2025). GPT-4's training, leveraging data from models like ChatGPT, has reduced its hallucination tendency, with GPT-4-launch scoring 19 percentage points higher than GPT-3.5 in avoiding open-domain hallucinations. However, the model's "epistemic humility" (hedging in responses) may inadvertently foster overreliance, as users might trust its cautious approach even when it's inaccurate (Maghsoudi et al., 2024).
- **Dual-Use Capabilities and Misuse:** AI capabilities, including those in GPT-4, have dual-use potential, meaning they can be used for both beneficial and harmful applications, such as for the proliferation of conventional and unconventional weapons. GPT-4 can provide information useful to individuals without formal scientific training, shortening research time for harmful purposes (Ourzik, 2024). It can also aid in cybersecurity subtasks like drafting phishing emails, though it struggles with more complex operations. The increased coherence of models like GPT-4 enables them to generate more believable and persuasive content, raising risks for disinformation and influence operations (Ourzik, 2024; Subasri et al., 2025).
- **Societal and Economic Impacts:** AI's economic impacts include the potential for workforce displacement and increased inequality, as AI may

automate jobs historically requiring significant experience and education, like legal services (Rickli & Vllasi, 2025). However, AI can also augment human workers, leading to better job matching and satisfaction (Raper, 2024). The sources emphasize the need for research into the economic impacts of AI and structures to smooth societal transitions.

Moral Status: Rights and Ethical Considerations for AGI

The provided sources do not directly address whether AGI would have moral status or rights. Instead, the ethical discussions primarily revolve around AI's impact on humans and society.

- **Focus on Human Well-being:** The ethical considerations consistently centre on human values and societal well-being. The primary ethical imperatives in AI research, particularly in computer vision, include algorithmic fairness, informed consent, robust privacy protocols, transparency, and human-in-the-loop systems, all aimed at protecting human interests and fostering trust between technology and society (Graham, 2022; Kelley & Atreides, 2020).
- **Human Oversight and Accountability:** The emphasis on human oversight and accountability for AI systems implies that humans remain the primary arbiters of ethical judgment and responsibility (Graham, 2022; Hussain, 2023; Maruyama, 2022). The "Harms of AI" chapter also discusses how AI technologies are currently being developed to "empower corporations and governments against workers and citizens," suggesting a focus on human agency and potential disempowerment rather than the AI's own status (Sołoducha, 2023).
- **Public Perception:** Public perception surveys indicate that while there are concerns about job displacement and the need for regulation, AI is largely viewed as a tool for problem-solving for human benefit. This perspective further underscores an anthropocentric view of AI's role (Wei, 2024b).

In conclusion, while the current discourse heavily addresses the safety, control, and existential risks posed by increasingly capable AI, directly examining the "moral status" or "rights" of AGI itself is not a prominent feature of the provided literature (Chen & Chen, 2024; Sołoducha, 2023; Wei, 2024b). Instead, the focus remains on the ethical implications for humanity and how to govern AI to ensure beneficial and responsible outcomes.

Timeline and Predictions: Varying Estimates for Achieving AGI

AI researchers (Dubrovsky, 2022) (Salmi, 2023a) estimated a 50% likelihood of achieving High-Level Machine Intelligence (HLMI) by

2047, a significant acceleration from the 2060 prediction in a similar 2022 survey. HLMI is defined as machines outperforming humans in all tasks, both intellectual and non-intellectual, more effectively and affordably (Mikki, 2023).

However, these predictions are marked by contradictions and considerable uncertainty. For example, while HLMI might be reached by 2047, the complete automation of all human occupations (FOAL) is not expected until 2116 (Cárdenas-García, 2023; Obaid, 2023). This apparent inconsistency—machines surpassing humans in all tasks far earlier than full occupational automation—raises questions about researchers' underlying assumptions or potential "framing effects." The idea that AI could achieve such capabilities and then remain static for generations is considered illogical given the anticipated pace of technological advancement (Kwon et al., 2025).

Regarding risks, respondents expressed substantial uncertainty about the long-term value of AI progress. A significant portion (37.8% to 51.4%) assigned at least a 10% chance to advanced AI leading to outcomes as severe as human extinction (Cárdenas-García, 2023; Obaid, 2023). The median probability for "existential risk from humans failing to control AI" was 10%, while "human extinction from AI in general" had a median probability of 5%. These probabilities are considered "very high." Importantly, the field broadly acknowledges existential risk from AI as a real and mainstream concern, though not a default outcome (Ajmal et al., 2025; Hemamalini, Jeyaselvi, et al., 2024; Oyejide et al., 2025).

Given these potential risks, there is broad consensus among researchers on prioritizing research aimed at minimizing potential AI-related risks, including strong support for AI safety research. The alignment problem, ensuring AI systems align with human intent and values—is considered a "very important problem" (41%) or "among the most important problems in the field" (13%), and is generally perceived as "harder" or "much harder" than other AI problems (Hemamalini, Tyagi, et al., 2024).

Despite concerns about potential negative outcomes, there was little support for immediately slowing down AI development. A common perspective suggests maintaining the current pace if AGI is more than 10-25 years away, while simultaneously preparing for future slowdowns, highlighting the complex balance between rapid innovation and responsible development (Ratnayake & Thomas, 2023).

The ethical and societal implications of AI necessitate strong governance. "Harms of AI" are not seen as inherent to the technology but rather as a consequence of its current development and deployment, which primarily empowers corporations and governments, potentially to the detriment of workers and citizens (Simon, 2019b) (Maruyama, 2022). To mitigate these harms, regulation and policies are needed to redirect AI research and deployment.

Key ethical imperatives and governance strategies emphasized include

- **Algorithmic Fairness**: Ensuring diverse and unbiased training datasets to mitigate harmful stereotypes.
- **Informed Consent**: Providing clear explanations of personal data usage, particularly in applications like facial recognition.
- **Public Engagement**: Actively involving affected communities through various mechanisms (town hall meetings, public forums) to align AI development with societal values.
- **Robust Privacy Protocols**: Implementing measures like data usage limits, encryption, and anonymization to safeguard personal data.
- **Explainability and Transparency**: Developing interpretable AI models to provide clear insights into decision-making, fostering trust and accountability.
- **Human-in-the-Loop Systems**: Integrating human oversight for nuanced judgment and ethical integrity, refining decision-making and promoting continuous improvement.
- **Model-Assisted Safety Pipelines**: Utilizing methods such as Reinforcement Learning from Human Feedback (RLHF) and Rule-Based Reward Models (RBRMs) to fine-tune AI behaviour, reducing harmful content and improving policy adherence.

The increasing capabilities of AI models mean that challenges and their consequences are "imminent," requiring extremely high reliability in interventions. This necessitates a "precautionary regulatory principle," especially in domains where the costs of AI could become politically and socially difficult to redress after large-scale implementation (Kwon et al., 2025; Obaid, 2023; Salmi, 2023b; Yampolskiy, 2022). Further research is needed into the economic impacts of AI, as well as structures for broader public participation in defining "optimal" model behavior. Evaluations are also required for risky emergent behaviors (e.g., long-horizon planning and persuasion), and improvements are necessary in interpretability, explainability, and calibration. Ultimately, ensuring safe and broadly useful AI systems requires continuous investment in safety research and ongoing public engagement to align AI development with societal needs and values (Ratnayake & Thomas, 2023) (Kwon et al., 2025).

Conclusion

The progression from Narrow AI to Artificial General Intelligence signifies a fundamental shift not only in technological prowess but also in our relationship with intelligent systems. While Narrow AI has excelled in specialized tasks, revolutionizing industries through its precision, efficiency, and scalability, these systems inherently lack the adaptability, self-reflection, and integrated reasoning that define human intelligence (Sujana & Augustine, 2024). In contrast, AGI holds the promise of operating across diverse domains, innovating autonomously, and engaging in complex problem-solving at levels potentially surpassing human capabilities (Sabeer et al., 2023). Achieving this requires significant breakthroughs in cognitive modelling, multi-modal integration, and value alignment, alongside a willingness to address profound technical and ethical uncertainties (Butt & Iqbal, 2025; Ennab & Mcheick, 2024).

As advancements accelerate, the implications of AI development extend beyond engineering, encompassing governance, ethics, and philosophy. The alignment problem, existential risk, and the moral status of AI are no longer theoretical concerns; they are urgent, emergent challenges demanding coordinated global attention (Cerrillo, 2025; Li-Yun Chang et al., 2024; Mehra et al., 2025). The journey toward AGI is not merely a technical endeavour but a societal decision about the kind of future we wish to create. By pairing innovation with robust safeguards, transparent oversight, and international cooperation, we can strive for an AGI that enhances human flourishing while mitigating the dangers of misaligned intelligence (Burdett, 2023; Giarmoleo et al., 2024). The direction we choose now will determine not only the capabilities of machines but also the destiny of humanity itself.

Conflict of Interest

The author(s) affirm that no financial interests or personal relationships exist that could have influenced the research presented in this paper. The opinions and conclusions expressed are entirely those of the author(s) and should not be interpreted as representing the views of their respective institutions.

Reference

1. Ajmal, C. S., Yerram, S., Abishek, V., Nizam, V. P. M., Aglave, G., Patnam, J. D., Raghuvanshi, R. S., & Srivastava, S. (2025). Innovative Approaches in Regulatory Affairs: Leveraging Artificial Intelligence and Machine Learning for Efficient Compliance and Decision-Making. AAPS Journal, 27(1).
2. Audibert, R. B., Lemos, H., Avelar, P., Tavares, A. R., & Lamb, L. C. (2023). On the Evolution of A.I. and Machine Learning: Towards a Meta-level Measuring and Understanding Impact, Influence, and Leadership at Premier A.I. Conferences. Journal of Applied Logics, 10(5), 693–817.
3. Booven, D. V., Cheng-Bang, C., & Meenakshy, M. (2024). Limitations of artificial intelligence in healthcare. In Artificial Intelligence in Urologic Malignancies (pp. 231–246).
4. Bu, C., & Zhang, Z. (2020). Research on Overfitting Problem and Correction in Machine Learning. J. Phys. Conf. Ser., 1693(1). Bucknall, B. S., & Dori-Hacohen, S. (2022). Current and Near-Term AI as a Potential Existential Risk Factor. AIES - Proc. AAAI/ACM Conf. AI, Ethics, Soc., 119–129.
5. Bullock, J. B., Chen, Y. C., Himmelreich, J., Hudson, V. M., Korinek, A., Young, M. M., & Zhang, B. (2022). The Oxford handbook of AI governance. In The Oxford Handbook of AI Governance.
6. Burdett, M. S. (2023). Proximate and Ultimate Concerns in Christian Ethical Responses to Artificial Intelligence. Studies in Christian Ethics, 36(3), 620–641.
7. Butt, T., & Iqbal, M. (2025). Explainable AI: Applications, Challenges, Current Solutions and Future Research Directions. ACM Int. Conf. Proc. Ser., 108–113.
8. Candiotto, K. B. B., & Karasinski, M. (2022). Artificial intelligence and the real existential risks: An analysis of the human limitations of control. Filosofia Unisinos, 23(3).
9. Cárdenas-García, J. F. (2023). Info-Autopoiesis and the Limits of Artificial General Intelligence. Computers, 12(5).
10. Cerrillo, J. M. C. (2025). Can intelligence be artificial? In Artificial intelligence and foreign aff.: AI, human rights, ethics and global governance (pp. 37–52).
11. Craig, L. (2025). GPT-4o vs. GPT-4: How do they compare? OpenAI's GPT-4o promises improved multimodal capabilities and increased efficiency. Explore the differences between GPT-4o and its predecessor, GPT-4. FEATURE.
12. Chen, B., & Chen, J. (2024). China's Legal Practices Concerning Challenges of Artificial General Intelligence. Laws, 13(5).
13. Croeser, S., & Eckersley, P. (2019). Theories of parenting and their application to artificial intelligence. AIES - Proc. AAAI/ACM Conf. AI, Ethics, Soc., 423–428.
14. Dubrovsky, D. I. (2022). The development of artificial intelligence and the global crisis

of earthly civilization (to the analysis of socio-humanitarian problems). Filosofiya Nauki i Tehniki, 27(2), 100–107.

15. Eigner, M. (2025). Artificial Intelligence (AI)—Closer Examination, Delimitation and Classification. ZWF Zeitschrift fuer Wirtschaftlichen Fabrikbetrieb, 120(s1), 5–11.
16. Ennab, M., & Mcheick, H. (2024). Enhancing interpretability and accuracy of AI models in healthcare: A comprehensive review on challenges and future directions. Frontiers in Robotics and AI, 11.
17. Estruch, V., Ferri, C., Hernández-Orallo, J., & Ramírez-Quintana, M. J. (2014). Bridging the gap between distance and generalization. Computational Intelligence, 30(3), 473–513.
18. Feng, T., Jin, C., Liu, J., Zhu, K., Tu, H., Cheng, Z., Lin, G., & You, J. (2024). How Far Are We From AGI: Are LLMs All We Need? Transactions on Machine Learning Research, 2024.
19. Gai, F. (2021). When Artificial Intelligence Meets Daoism. In Intelligence and Wisdom: Artificial Intelligence Meets Chin. Philosophers (pp. 83–100). Springer Nature;
20. George, D., Lázaro-Gredilla, M., & Guntupalli, J. S. (2020). From CAPTCHA to Commonsense: How Brain Can Teach Us About Artificial Intelligence. Frontiers in Computational Neuroscience, 14.
21. Giarmoleo, F. V., Ferrero, I., Rocchi, M., & Pellegrini, M. M. (2024). What ethics can say on artificial intelligence: Insights from a systematic literature review. Business and Society Review, 129(2), 258–292.
22. Graham, R. (2022). Discourse analysis of academic debate of ethics for AGI. AI and Society, 37(4), 1519–1532. Scopus. https://doi.org/10.1007/s00146-021-01228-7
23. Hagendorff, T., & Wezel, K. (2020). 15 challenges for AI: or what AI (currently) can't do. AI and Society, 35(2), 355–365.
24. Hemamalini, V., Jeyaselvi, M., Savitha, S., & Tyagi, A. K. (2024). Emerging technologies in drug discovery for providing efficient services to patients: Future opportunities and challenges. In Converging Pharm. Sci. And Eng. In Comput. Drug Discov. (pp. 177–189). IGI Global;
25. Hemamalini, V., Tyagi, A. K., Vennila, V., & Kumari, S. (2024). Revolutionizing drug discovery with cutting-edge technologies: Issue and challenges for the next decade. In Converging Pharm. Sci. And Eng. In Comput. Drug Discov. (pp. 76–89). IGI Global;
26. Hu, X. (2023). Development Overview and Key Technologies of Artificial Intelligence. In Cheng C. & Zhang H. (Eds.), Int. Symp. Proj. Manag., ISPM (Vol. 1, pp. 378–383). Aussino Academic
27. Hussain, M. (2023). The Moral Status of AGI-enabled Robots: A Functionality-Based Analysis. Symposion, 10(1), 105–127. Scopus. https://doi.org/10.5840/symposion20231014

28. Ismail, M., Madathil, N. T., Alalawi, M., Alalawi, S., & Alrabaee, S. (2025). Navigating Ethical Dilemmas in the Implementation of AI-Driven Educational Technologies. IEEE Global Eng. Edu. Conf., EDUCON. IEEE Global Engineering Education Conference, EDUCON.
29. Jiang, Y., Zhang, X., & Dai, M. (2025). Open governance and innovation directions regulation: Research on institutional mechanisms for promoting development of artificial general intelligence. Bulletin of Chinese Academy of Sciences, 40(4), 652–662.
30. Kelley, D., & Atreides, K. (2020). AGI Protocol for the Ethical Treatment of Artificial General Intelligence Systems. 169, 501–506.
31. Khoee, A. G., Yu, Y., & Feldt, R. (2024). Domain generalization through meta-learning: A survey. Artificial Intelligence Review, 57(10).
32. Koli, H. V., Thaker, D. J., & Khengar, J. (2025). AI-Powered Transformation in Healthcare: Innovations, Implementation Challenges, and Future Perspectives. Int. Conf. Trends Eng. Syst. Technol., ICTEST - Proc. International Conference on Trends in Engineering Systems and Technologies, ICTEST 2025 - Proceedings.
33. Krinkin, K., Shichkina, Y., & Ignatyev, A. (2023). Co-evolutionary hybrid intelligence is a key concept for the world intellectualization. Kybernetes, 52(9), 2907–2923.
34. Kurshan, E. (2024). From Pursuit of the Universal AGI Architecture to Systematic Approach to Heterogeneous AGI (SAGI): Addressing Alignment, Energy & AGI Grand Challenges. International Journal of Semantic Computing, 18(3), 465–500.
35. Kwon, W., Baek, S., Baek, J., Shin, W., Gwak, M., Park, P., & Lee, S. (2025). Reinforced Intelligence Through Active Interaction in Real World: A Survey on Embodied AI. International Journal of Control, Automation and Systems, 23(6), 1597–1612.
36. Labudová, L. (2024). The Influence of Artificial Intelligence on Modern Marketing. In Buisine S. (Ed.), Proc. Eur. Conf. Innov. Entrepren., ECIE (Vol. 19, Issue 1, pp. 391–397).
37. Li, J., Zhao, Y., Sun, C., Bao, X., Zhao, Q., & Zhou, H. (2018). A Survey of Development and Application of Artificial Intelligence in Smart Grid. IOP Conf. Ser. Earth Environ. Sci., 186(4).
38. Li, N. (2023). Ethical Considerations in Artificial Intelligence: A Comprehensive Disccusion from the Perspective of Computer Vision. SHS Web of Conferences, 179, 04024.
39. Li-Yun Chang, W., Liu, S.-Y., Liao, Y.-K., & Lee, T. S.-H. (2024). Ethical Concerns with Regards to Artificial Intelligence: A National Public Poll in Taiwan. IEEE Access, 12, 133595–133605.
40. Maghsoudi, M., Mohammadi, A., & Habibipour, S. (2024). Navigating and Addressing Public Concerns in AI: Insights

from Social Media Analytics and Delphi. IEEE Access, 12, 126043–126062.

41. Maran, K., Priyadarshini, P., Senthilnanthan, C. R., Manikandan, M., Kumar, R. G., & Ramu, M. (2024). Impact of Artificial Intelligence on Global Healthcare Sector Performance. Int. Conf. Power, Energy, Control Transm. Syst.: Harnessing Power Energy an Afford. Electrif. India, ICPECTS. 4th International Conference on Power, Energy, Control and Transmission Systems: Harnessing Power and Energy for an Affordable Electrification of India, ICPECTS 2024
42. Maruyama, Y. (2022). Moral Philosophy of Artificial General Intelligence: Agency and Responsibility. 13154 LNAI, 139–150.
43. Mashabab, M. F., Sheniff, M. S. A., Alsharief, M. S., Yami, M. A. A. A., Matnah, H. N. M., Abbas, A. M. A., Shenief, H. Y. M. A., Abbas, D. A. A. A. A., Raseen, F. M. S. A., & Kulayb, A. H. A. A. (2024). The Role of Artificial Intelligence in Healthcare: A Critical Analysis of Its Implications for Patient Care. Journal of Ecohumanism, 3(7), 597–604.
44. McCormack, J. (2023). Autonomy, Intention, Performativity: Navigating the AI Divide. In Choreomata: Performance and Performativity after AI (pp. 240–257).
45. Mehra, S., Das, T., Meher, S. M., & Khan, S. (2025). Ethical AI: A framework for building responsible artificial intelligence systems. In Converg. Of AI, Educ., and Bus. For Sustain. (pp. 1–23).
46. Melnikov, A. A., Makmal, A., Dunjko, V., & Briegel, H. J. (2017). Projective simulation with generalization. Scientific Reports, 7(1).
47. Mikki, S. (2023). Artificial General Intelligence and Noncomputability: A Dynamical Framework. Journal of Artificial Intelligence and Consciousness, 10(1), 71–101.
48. Müller, V. C., & Cannon, M. (2022). Existential risk from AI and orthogonality: Can we have it both ways? Ratio, 35(1), 25–36.
49. Naik, N., Hameed, B. M. Z., Shetty, D. K., Swain, D., Shah, M., Paul, R., Aggarwal, K., Brahim, S., Patil, V., Smriti, K., Shetty, S., Rai, B. P., Chlosta, P., & Somani, B. K. (2022). Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? Frontiers in Surgery, 9.
50. Narayanan, V., Cao, Y., Panda, P., Challapalle, N. R., Du, X., Kim, Y., Krishnan, G., Lee, C., Li, Y., Sun, J., Venkatesha, Y., Wang, Z., & Zheng, Y. (2023). Overview of Recent Advancements in Deep Learning and Artificial Intelligence. In Advances in Electromagnetics Empowered by Artificial Intelligence and Deep Learning (pp. 23–79).
51. Obaid, O. I. (2023). From Machine Learning to Artificial General Intelligence: A Roadmap and Implications. Mesopotamian Journal of Big Data, 2023, 81–91.

52. OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2024).
53. Ourzik, V. Y. (2024). Security and Safety Concerns in the age of AI. In Goncalves C. & Rouco J.C.D. (Eds.), Proc. Int. Conf. AI Res., ICAIR (pp. 329–337).
54. Oyejide, A. J., Adekunle, Y. A., Abodunrin, O. D., & Atoyebi, E. O. (2025). Artificial intelligence, computational tools and robotics for drug discovery, development, and delivery. Intelligent Pharmacy, 3(3), 207–224.
55. Radanliev, P. (2024). Artificial intelligence: Reflecting on the past and looking towards the next paradigm shift. Journal of Experimental and Theoretical Artificial Intelligence.
56. Raper, R. (2024). Is There a Need for Robots with Moral Agency? A Case Study in Social Robotics. Proc IEEE Int Conf Ind Technol. Proceedings of the IEEE International Conference on Industrial Technology.
57. Ratnayake, D., & Thomas, L. F. (2023). artificial intelligence. itnow, 65(4), 43.
58. Rickli, J.-M., & Vllasi, G. (2025). The Weaponization of Emerging Technologies and Their Impact on Global Risk: A Perspective from the PfPC Emerging Security Challenges Working Group. Connections, 24(1), 91–112.
59. Roubal, O., & Narusevych, K. (2025). the panorama of artificial intelligence – development or advancement? Communication Today, 16(1), 48–63.
60. Sabeer, S., Siddiqua, A., Anjum, A., Kondapalli, S., Kaur, C., & Begum, A. U. (2023). Role of artificial intelligence for the curative field: A review. In Samarah A., Bagwari A., & Bagwari J. (Eds.), AIP Conf. Proc. (Vol. 2930, Issue 1).
61. Salmi, J. (2023a). A democratic way of controlling artificial general intelligence. AI and Society, 38(4), 1785–1791.
62. Salmi, J. (2023b). A democratic way of controlling artificial general intelligence. AI and Society, 38(4), 1785–1791.
63. Samad, M. E., Nasserddine, G., & Kheir, A. (2023). Introduction to Artificial Intelligence. In Artificial Intelligence and Knowl. Processing: Improved Decision-Making and Prediction (pp. 1–14).
64. Seth, J. (2024). Public Perception of AI: Sentiment and Opportunity. http://arxiv.org/abs/2407.15998
65. Sharman, R. (2022). Data Challenges and Societal Impacts – The Case in Favor of the Blueprint for an AI Bill of Rights (Keynote Remarks). In Roy P.P., Agarwal A., Li T., Krishna Reddy P., & Uday Kiran R. (Eds.), Lect. Notes Comput. Sci.: Vol. 13773 LNCS (pp. 3–15).
66. Shi, Z., Jing, J., Sun, Y., Lim, J.-H., & Zhang, M. (2025). Unveiling the Tapestry: The Interplay of Generalization and Forgetting in Continual Learning. IEEE

Transactions on Neural Networks and Learning Systems, 36(8), 15070–15084.

67. Simon, C. J. (2019a). Ethics and Artificial General Intelligence: Technological Prediction as a Groundwork for Guidelines. Int Symp Technol Soc, 2019-November.
68. Simon, C. J. (2019b). Ethics and Artificial General Intelligence: Technological Prediction as a Groundwork for Guidelines. 2019-November.
69. Skripnuk, D. F., Kikkas, K. N., & Merkulov, V. I. (2023). Forecasting of the Global Market of Software that Uses Artificial Intelligence Algorithms. In Ilin I., Kudryavtseva T., & Petrova M.M. (Eds.), Lect. Notes Networks Syst.: Vol. 684 LNNS (pp. 707–721).
70. Smirnov, E. N., & Lukyanov, S. A. (2019). Development of the global market of artificial intelligence systems. Economy of Regions, 15(1), 57–69.
71. Sołoducha, K. (2023). Wyzwania związane z tworzeniem etyki typu data driven dla maszyn typu artificial general inteligence (AGI). Argument: Biannual Philosophical Journal, 13(1), 109–122.
72. Subasri, V., Baghbanzadeh, N., Celi, L. A., & Seyyed-Kalantari, L. (2025). Potential for near-term AI risks to evolve into existential threats in healthcare. BMJ Health & Care Informatics, 32(1).
73. Sublime, J. (2024). The AI Race: Why Current Neural Network-based Architectures are a Poor Basis for Artificial General Intelligence. Journal of Artificial Intelligence Research, 79, 41–67.
74. Sujana, D. S., & Augustine, D. P. (2024). Practical Benefits of Using AI for More Accurate Forecasting in Mental Health Care. In Artificial Intelligence in Forecasting: Tools and Techniques (pp. 242–265).
75. Sun, B., Mei, X., Du, Y., & Gao, M. (2024). A General Intelligence Theory Based on the Representation of Need. Int. Conf. Comput. Commun. Syst., ICCCS, 1325–1329.
76. Tobin, S., Jayabalasingham, B., Huggett, S., & De Kleijn, M. (2020). A brief historical overview of artificial intelligence research. Information Services and Use, 39(4), 291–296.
77. Tran, S., Taylor, S. A., & Nicolau, D. V. (2021). Limits of Intelligence and Design Implication. In Nakano T. (Ed.), Lect. Notes Inst. Comput. Sci. Soc. Informatics Telecommun. Eng.: Vol. 403 LNICST (pp. 215–229).
78. Tugui, A., Danciulescu, D., & Subtirelu, M.-S. (2019). The biological as a double limit for artificial intelligence: Review and futuristic debate. International Journal of Computers, Communications and Control, 14(2), 253–271.
79. Wei, Y. (2024a). Several important ethical issues concerning artificial general intelligence. Chinese Medical Ethics, 37(1), 1–9.
80. Wei, Y. (2024b). Several important ethical issues concerning artificial general

intelligence. Chinese Medical Ethics, 37(1), 1–9.

81. Weinbaum (Weaver), D., & Veitas, V. (2017). Open ended intelligence: The individuation of intelligent agents. Journal of Experimental and Theoretical Artificial Intelligence, 29(2), 371–396.
82. Yampolskiy, R. V. (2022). AGI Control Theory. 13154 LNAI, 316–326. Scopus. https://doi.org/10.1007/978-3-030-93758-4_33
83. Zhu, X., Chen, S., Liang, X., Jin, X., & Du, Z. (2024). Next-generation generalist energy artificial intelligence for navigating smart energy. Cell Reports Physical Science, 5(9).