scientific reports

Check for updates

OPEN

Navigating artificial general intelligence development: societal, technological, ethical, and brain-inspired pathways

Raghu Raman[1]✉, Robin Kowalski[2], Krishnashree Achuthan[3], Akshay Iyer[4] & Prema Nedungadi[5]

This study examines the imperative to align artificial general intelligence (AGI) development with societal, technological, ethical, and brain-inspired pathways to ensure its responsible integration into human systems. Using the PRISMA framework and BERTopic modeling, it identifies five key pathways shaping AGI's trajectory: (1) societal integration, addressing AGI's broader societal impacts, public adoption, and policy considerations; (2) technological advancement, exploring real-world applications, implementation challenges, and scalability; (3) explainability, enhancing transparency, trust, and interpretability in AGI decision-making; (4) cognitive and ethical considerations, linking AGI's evolving architectures to ethical frameworks, accountability, and societal consequences; and (5) brain-inspired systems, leveraging human neural models to improve AGI's learning efficiency, adaptability, and reasoning capabilities. This study makes a unique contribution by systematically uncovering underexplored AGI themes, proposing a conceptual framework that connects AI advancements to practical applications, and addressing the multifaceted technical, ethical, and societal challenges of AGI development. The findings call for interdisciplinary collaboration to bridge critical gaps in transparency, governance, and societal alignment while proposing strategies for equitable access, workforce adaptation, and sustainable integration. Additionally, the study highlights emerging research frontiers, such as AGI-consciousness interfaces and collective intelligence systems, offering new pathways to integrate AGI into human-centered applications. By synthesizing insights across disciplines, this study provides a comprehensive roadmap for guiding AGI development in ways that balance technological innovation with ethical and societal responsibilities, advancing societal progress and well-being.

Keywords Artificial general intelligence, Strong AI, Weak AI, Ethical AI, Responsible AI, Human-like AI, Superintelligence, Ethics, Topic modeling, Brain inspired

Artificial intelligence (AI) marks the foundation of modern technological advancements, focusing on narrow, task-specific applications. Artificial general intelligence (AGI) aims to replicate human cognitive capabilities across domains, setting the stage for transformative societal and economic impacts, particularly in areas involving human decision-making and adaptive reasoning[1–3]. The ultimate leap to artificial superintelligence (ASI) involves machines that surpass human intelligence, presenting unparalleled opportunities for innovation alongside significant ethical and existential challenges[4–7]. AGI is predicted to significantly alter the trajectory of human civilization, potentially leading to posthuman conditions while reshaping human–computer interactions and cognitive frameworks[8–10]. AGI promises a shift from task-specific algorithms to systems that mimic human cognitive abilities, offering unprecedented capabilities in learning, reasoning, and decision-making, which are central to fields like cognitive psychology and behavioral research[11]. AGI could revolutionize areas such as biomedical research, nanotechnology, energy research, and cognitive enhancement, leading to an

[1]Amrita School of Business, Amrita Vishwa Vidyapeetham, Amritapuri, Amritapuri, Kerala 690525, India. [2]College of Behavioral, Social and Health Sciences, Clemson University, Clemson, SC 29634, USA. [3]Center for Cybersecurity Systems and Networks, Amrita Vishwa Vidyapeetham, Amritapuri, Amritapuri, Kerala 690525, India. [4]Department of Microbiology and Immunology, University of Miami, Miller School of Medicine, Miami, FL 33136, USA. [5]Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, Amritapuri, Kerala 690525, India. ✉email: raghu@amrita.edu

"intelligence explosion," where AGIs could program other AGIs, resulting in rapid and radical technological advancements[12–14]. Moreover, the cognitive dimensions of AGI are critical in designing systems that align with human behavior, fostering collaboration and trust in decision-making contexts such as healthcare and education.

While AGI, driven by generative AI and domain-specific technologies, promises efficiency, safety, and profitability, its reliance on Generative Adversarial Networks (GANs) remains debated. GANs enhance data synthesis, unsupervised learning, and decision-making, yet critics highlight their lack of intrinsic reasoning, poor generalization, and bias susceptibility. Alternative approaches, including neurosymbolic AI, hybrid cognitive architectures, and evolutionary computing, prioritize explainability, adaptability, and reasoning, challenging GAN-driven models[15]. This debate underscores the need for a pluralistic AGI framework, evaluating diverse methodologies rather than over-relying on a single paradigm.

Some argue that symbolic reasoning, neuromorphic computing, or hybrid AI architectures may offer more viable pathways to AGI, underscoring the need for further interdisciplinary exploration[16–20]. As AGI systems increasingly influence human decision-making in critical domains such as healthcare, finance, and governance, they introduce significant risks and ethical challenges, including bias, accountability, and unintended consequences. Ensuring alignment with human cognitive and behavioral norms requires careful oversight, robust regulatory frameworks, and proactive measures to mitigate potential harm while maximizing societal benefits[13,21,22]. This systematic review explores the multifaceted dimensions of AGI, examining its conceptual foundations, technological advancements, ethical considerations, and potential societal impacts, particularly in contexts where human behavior and machine intelligence converge.

AGI's theoretical potential aligns with advanced human reasoning, as imagined in films such as *Her and Ex Machina*, where AI surpasses human comprehension. While these portrayals remain speculative, companies like DeepMind and OpenAI have made notable advancements with systems such as AlphaGo and GPT models. These systems, though not AGI, represent significant milestones in AI development. For instance, AlphaGo's victory over the world's best Go player[23] marked a groundbreaking achievement in AI reasoning, mastering intuition and strategic thinking previously considered uniquely human. Despite these advancements, current AI remains limited to task-specific applications, underscoring the considerable gap between existing technologies and the generalized, flexible capabilities envisioned for AGI[24]. This gap highlights the ongoing challenges in bridging advanced AI with the cognitive adaptability and domain-spanning reasoning characteristic of human intelligence.

A proposed paradigm shift in AGI research involves viewing intelligence not only as a static competence of individual agents but also as a formative process of self-organization. This perspective aligns with the natural progression of complex systems and could help in building AGI systems that are more robust, adaptable, and capable of learning in dynamic environments[25]. AGI requires approaches that transcend current AI limitations and account for both the technical and philosophical dimensions. One critical factor in AGI development is the role of deep learning and big data**.** These modern techniques have driven unprecedented advancements in AI, enabling systems to excel in tasks such as natural language processing and computer vision. However, they remain insufficient for achieving true AGI. While deep learning systems rely on vast datasets to extract patterns and make predictions, they lack the ability to generalize knowledge across domains or reason abstractly in novel situations[1,26]. This gap highlights the need for more sophisticated learning paradigms capable of mimicking the flexibility and adaptability of human cognition.

Recent advancements in AI, particularly in machine learning and deep neural networks, have brought us closer to achieving AGI. The review by Wickramasinghe et al.[27] explores continual learning (CL) as a vital step toward AGI, highlighting the need for brain-inspired data representations and learning algorithms to overcome catastrophic forgetting and enable adaptive, knowledge-driven systems. Here, "brain-inspired" means designing AI to learn and remember like the human brain by mimicking how we process, store, and recall information to avoid forgetting. Fei et al.[28] highlight the progress toward multimodal foundation models as a step to achieving true AGI, suggesting that these models might soon enable machines to mimic core cognitive activities historically unique to humans. The AGI protocol is designed to provide a basis for working with AGI systems, especially those that have the possibility of having emotional, subjective experiences from a theoretical standpoint[29].

However, these technologies still have significant limitations that need to be addressed[26,30]. One of the major challenges is the control of AGI, i.e., ensuring that AGI systems operate in ways consistent with human values and priorities[9,10,31–33]. Generative AI, exemplified by LLMs, has intensified debates about the proximity of achieving AGI, with perspectives ranging from optimism about its potential for equitable resource distribution and reduced human labor to fears of catastrophic societal dominance by machines[34,35].

In addition to technical challenges, AGI raises ethical issues such as subjectivity, safety, ethical responsibility, controllability, and socialization as an artificial autonomous moral system[22,36]. The risks associated with AGIs include existential risks, inadequate management, and AGIs with poor ethics, morals, and values[37,38]. Current regulatory frameworks may be inadequate for addressing these risks[21,39]. Establishing ethical guidelines for the development and deployment of AGI is crucial. This includes considering the values embedded in AI systems and their potential societal impacts[22,39]. Lenharo[40], for example, highlights the growing debate on AI consciousness, urging the development of welfare policies to address potential ethical challenges. Researchers advocate for frameworks to assess AI consciousness and caution against the neglect or misallocation of resources. While AI consciousness remains uncertain, proactive planning is crucial for balancing safety, ethics, and societal impacts. Shankar[41] highlights the dual role of AI in driving efficiency and innovation while raising ethical, privacy, and job-related concerns. He proposes a framework to assess AI's impact, explores real-world applications, and emphasizes the need for responsible AI principles to balance benefits with risks as advancements in AGI and ASI emerge. Wu[42] examines the transformative impact of AGI on information professions, emphasizing the need for updated skills, curricula, and ethical frameworks to adapt to an AGI-driven landscape while maintaining human-centric values. Sukhobokov et al.[43] propose a reference cognitive architecture for AGI, integrating

diverse knowledge representation methods and modules such as consciousness, ethics, and social interaction to address gaps in existing architectures and advance AGI capabilities. Salmon et al.[44] underscore the potential existential risks posed by AGI and highlight the critical yet underutilized role of human factors and ergonomics (HFE) in ensuring safe, ethical, and usable AGI design, advocating for collaboration, lifecycle integration, and systems HFE approaches to address these challenges. This mapping highlights the evolving role of AI across industries, providing a foundation for identifying industry-specific research opportunities.

Existing research on AI focuses on narrow applications, leaving gaps in understanding AGI's adaptability, reasoning, and alignment with human cognition. While studies explore deep learning and neuromorphic computing, limited attention is given to cognitive science perspectives, ethical concerns, and societal impacts. This study addresses these gaps by systematically identifying key AGI research pathways bridging technological advancements with cognitive, ethical, and societal dimensions to guide responsible development aligned with human behavior and values. This study addresses the critical need for a deeper understanding of AGI as it transitions from theoretical exploration to practical development. The rapid evolution of AI technologies necessitates the creation of frameworks that align AGI with human values, ensure equitable access, and mitigate potential risks such as job displacement, privacy violations, and ethical dilemmas. To this end, we propose the following research questions.

1. What pathways are needed to enable scalable, adaptable, and explainable AGI across diverse environments?
2. How can AGI systems be developed to align with ethical principles, societal needs, and equitable access?
3. What pathways can ensure effective collaboration, trust, and transparency between humans and AGI systems?
4. How can AGI contribute to advancements through interdisciplinary integration?

These research questions address critical gaps in AGI development, ensuring its adaptability, ethical alignment, and societal integration. While current AI systems struggle with scalability and explainability, understanding the advancements needed for AGI to operate across diverse environments is essential. Ethical concerns, including fairness and equitable access, highlight the need for frameworks that align AGI with societal values. Trust and transparency remain major challenges, making it crucial to establish collaborative frameworks that foster human-AI interaction. Finally, interdisciplinary integration is key to unlocking AGI's full potential, bridging technological innovation with real-world applications.

This paper makes several significant contributions to the AGI-related literature. First, it identifies and categorizes underexplored themes via machine learning-based BERTopic modeling, offering a novel perspective on emerging challenges and opportunities within AGI research. Second, it establishes a conceptual framework that distinguishes key AI concepts, positions AGI as a pivotal milestone, and links AI types to industry applications for targeted innovation while addressing ethical implications. Third, the paper addresses the multifaceted challenges of achieving AGI, spanning the technical, theoretical, ethical, and societal domains, and proposes strategies to develop adaptive, efficient systems aligned with human values and global needs. Fourth, it analyzes the societal, ethical, economic, legal, and psychological impacts of AGI, emphasizing the need for interdisciplinary frameworks to ensure its responsible development and integration into society. Finally, it identifies interconnected themes in AGI research, such as ethical governance, collective intelligence, brain-inspired designs, consciousness, and generative AI in education, providing a comprehensive roadmap for aligning AGI development with societal needs, ethical considerations, and transformative advancements.

Conceptual background

To effectively map research trends in AGI, establishing a robust conceptual foundation is essential. This section clarifies key AI concepts, their relationships, and their relevance to technological advancements and industry applications while addressing ethical considerations and emerging future directions. The journey toward AGI is marked by key distinctions among closely related AI concepts. Understanding these terms provides clarity on their scope, capabilities, and implications (Table 1).

- **Weak AI (also known as narrow AI)**: This encompasses AI systems designed to excel in specific, well-defined tasks without possessing general cognitive abilities or reasoning capabilities. Often, task-specific, weak AI is highly efficient within its scope but cannot adapt or apply knowledge beyond its trained functions. Examples include voice assistants such as Siri, language translation tools, recommendation algorithms, and specialized systems such as AlphaGo[45–47].
- **Human-level AI:** This concept describes AI that matches human cognitive abilities, such as natural reasoning, emotional understanding, and complex decision-making. For example, human-level AI could mimic human intuition in fields such as diplomacy or creative storytelling, exhibiting behavioral parity with humans[48–50].
- **Human-like AI:** This concept focuses on mimicking human behaviors, such as speech, facial expressions, and emotional responses. Unlike AGI, human-like AI prioritizes interaction and relatability over cognitive versatility. Examples include conversational systems such as ChatGPT or humanoid robots such as Sophia, which simulate human emotions[51,52].
- **Artificial general intelligence (AGI):** AGI refers to AI systems that can understand, learn, and adapt to perform any intellectual task a human can perform. Unlike narrow AI, AGI generalizes learning across multiple domains and applies reasoning to new, unfamiliar problems. A true AGI system could diagnose diseases, compose symphonies, and design complex engineering systems—all without task-specific programming[7,9,53,54].
- **Strong AI:** Strong AI, often synonymous with AGI, suggests an AI system that possesses actual understanding, consciousness, and self-awareness. While AGI focuses on functional performance, strong AI questions whether machines can "think" in a human-like way, raising philosophical debates about sentience[1,55–57].

Type of AI	Definition	Key characteristics	Ethical considerations	Examples
Weak AI	AI systems are designed to perform specific tasks efficiently without general cognitive abilities	Task-specific and goal-oriented; Limited adaptability beyond trained functions	Biases in outputs, limited explainability, and reliance on potentially flawed training datasets	Image recognition systems, recommendation engines, and virtual assistants like Alexa or Google Assistant
Human-level AI	AI systems that can match human capabilities in intellectual and cognitive tasks	Comparable performance to humans; Reasoning across diverse tasks	Biases in AI decisions, the replacement of human jobs, and the potential erosion of privacy	Discussed in Turing Test scenarios and human-level game-playing AIs like AlphaGo
Human-like AI	AI systems are designed to mimic human cognitive behaviors, reasoning, and problem-solving skills	Human-like responses and behaviors; Focus on imitation	User manipulation, transparency in AI behavior, and maintaining ethical human-AI interactions	ChatGPT, Siri, Sophia the Robot, and conversational AIs
Artificial General Intelligence (AGI)	AI systems are capable of performing any intellectual task that a human can, with adaptability across domains	Generalized learning and reasoning abilities; Task-agnostic	Misuse, control, and unintended consequences impact human society, requiring frameworks for governance and safety	Hypothetical systems like OpenCog and concepts explored in DeepMind's research on AGI
Strong AI	AI that exhibits true understanding, reasoning, and consciousness similar to human beings	Ability to think, understand, and self-reflect, Not just task execution	Machine rights, moral agency, and accountability in decision-making processes	No real-world examples yet; explored in philosophical AI debates (e.g., John Searle's "Chinese Room Argument")
Artificial Superintelligence	Hypothetical AI that surpasses human intelligence in all domains, including decision-making and creativity	Exceeds human cognitive abilities; Capable of rapid self-improvement	Loss of human control, unequal power distribution, and existential threats to humanity	Nick Bostrom's scenarios in *Superintelligence*; futuristic portrayals in movies like *"Her"* or *"Ex Machina."*

Table 1. Comparison of different types of AI.

Type of AI	Industries and domains	Example applications
Human-level AI	- Customer Service: AI agents replicating human communication.- Gaming: Human-like opponents for adaptive gameplay.- Education: Tutors are as effective as human teachers.- Manufacturing: Robots capable of cognitive task decision-making.- Entertainment: Film, animation, or content creation	- AI-powered tutors providing personalized support.- AI-driven gaming characters that adapt dynamically to users
Human-like AI	- Retail: Virtual customer assistants providing human-like interactions.- Healthcare: Mental health support via AI companions.- Hospitality: Human-like robots assisting in customer care.- Education: Socially interactive tutors for children.- Marketing: Personalized AI communications and campaigns	- AI robots assisting elderly patients.- Virtual shopping assistants enhancing customer experience
Artificial General Intelligence (AGI)	- Healthcare: Diagnosis, personalized treatments, advanced decision-making.- Education: Adaptive learning systems.- Finance: Dynamic risk modeling, market predictions.- Defense: Strategic simulations.- R&D: Scientific discovery and innovation across disciplines	- AI-driven doctors capable of autonomous diagnosis.- AI-powered research assistants discovering new materials or drugs
Strong AI	- Philosophy/Ethics: Understanding consciousness and cognition.- Healthcare: Cognitive reasoning for personalized care.- Legal Sector: Independent decision-making in complex cases.- Creative Industries: True artistic or creative generation.- Robotics: Fully autonomous robots	- AI systems capable of moral reasoning or empathy.- AI judges handling legal cases with fairness and ethics
Artificial Superintelligence	- Defense and Security: Strategic autonomous decision-making beyond human capabilities.- Finance: Global economic optimization.- Energy: Highly efficient resource allocation systems.- Scientific Research: Quantum mechanics, climate modeling, or deep space exploration.- Governance: AI-driven policymaking and problem-solving	- Climate-change modeling far beyond human accuracy.- Optimized, AI-led governance systems

Table 2. Relevance of AI types to industries and domains.

- **Artificial superintelligence (ASI):** ASI refers to hypothetical AI systems that surpass human intelligence in every measurable aspect, including creativity, strategic thinking, and problem solving. While AGI matches human abilities, ASI achieves capabilities beyond human comprehension, raising significant ethical and existential questions[4,6,7,58].

By defining these concepts, this study positions AGI as a critical milestone in AI research while distinguishing it from speculative ideas such as superintelligence technologies. AGI systems are expected to be capable of operating in unknown environments, reusing knowledge gained in different problem domains, and autonomously learning and understanding problem domains, which sets them apart from weak AI (narrow AI). Despite significant advancements in AI, most current systems are still classified as narrow AIs, which excel in specific domains but lack general intelligence. Unlike narrow AI, AGI aims to distill principles of intelligence that operate independently of specific problem domains.

Each AI type aligns with industries and domains based on its capabilities. As shown in Table 2, human-like AI is already being adopted in customer service, healthcare, and education, whereas AGI and strong AI are poised for future breakthroughs in domains requiring human-level cognition, such as medicine, education, and manufacturing. Artificial superintelligence systems remain hypothetical but hold the potential for transformative advancements in defense, global governance, and scientific discovery.

Related work

The potentialities and promise of AGI have been the subject of much research debate. According to Stewart[59], while AGI may surpass humans in capabilities such as reasoning and problem solving, its lack of a central nervous system, and thus its ability to experience emotions and their consequences, fundamentally limits its ability to develop ethical systems and assume universal leadership. At the same time, Vaidya[60] critically examines

the argument that machines cannot have emotions due to a lack of feelings, proposing instead that AGI systems could possess emotions through their capacity for judgment, as demonstrated in emotions such as anger. Triguero et al.[61] explore general-purpose artificial intelligence systems (GPAISs), bridging the gap between specialized AI applications and the aspirational goal of artificial general intelligence (AGI) by proposing a definition and taxonomy that classifies GPAISs based on autonomy, adaptability.

The AGI-related review studies in Table 3 present a multifaceted exploration of AGI across diverse domains, underscoring its transformative potential while highlighting significant research gaps. From enhancing innovation synergy in Industry 5.0 through AGI-driven processes[62] to advancing intelligent crisis management systems that integrate AGI with blockchain technologies[63], these studies emphasize AGI's capacity to address complex, multilayered challenges. The upstream geoenergy industry benefits from the application of AGI in optimizing operational efficiency and domain-specific knowledge[18], whereas information professions necessitate AGI literacy and ethical reasoning to adapt to a rapidly evolving landscape[42].

Despite these advancements, recurring research gaps are evident. Several studies identify a lack of comprehensive frameworks to address AGI ethical and security concerns, such as AGI safety in medical IoT systems[64] and the risks associated with AGI autonomy and goal misalignment[37]. Visual domains also present challenges, where the intricacies of prompt engineering for large vision models remain underexplored[65,66]. Similarly, while AGI principles hold promise in bridging behavioral and design sciences for Society 5.0[67], critical gaps in transparency and interpretability persist. Smart city technologies incorporating AGI[69,70] highlight AGI's role in fostering disruptive innovation but also expose the need for robust frameworks to navigate implementation barriers. Moreover, the development of AGI frameworks for autonomous systems[71] and safety literature[72] calls for more comprehensive risk mitigation strategies and interdisciplinary collaboration.

There is debate over how quickly AGI might emerge and how dangerous it could be. Some speculate that its development will be gradual and only moderately dangerous, whereas others fear sudden and uncontrollable emergence[26,22]. Interdisciplinary efforts underline the synthesis of neuroscience, cognitive science, and advanced computational models to replicate human reasoning processes[73]. This endeavor is not merely technical but philosophical, questioning the very nature of intelligence and consciousness[74]. Some researchers have proposed developing AGI by mimicking the human brain's structure and functions, incorporating elements such as sensory processing and memory storage[75,76]. Questions surrounding sentience, consciousness, and moral responsibility become increasingly relevant as AGI systems approach human-like reasoning and decision-making abilities. Achieving true AGI requires breakthroughs not only in hardware and algorithms but also in understanding human intelligence itself.

Efforts to address current limitations often focus on cognitive architectures, which aim to model the underlying processes of human intelligence[43]. Cognitive architectures, such as Soar and the ACT-R, provide frameworks for integrating perception, reasoning, and learning into unified systems. These models emphasize structured, hierarchical approaches to information processing, drawing inspiration from neuroscience and cognitive psychology. By simulating cognitive functions, these architectures offer a path toward creating AGI systems that can solve problems, learn iteratively, and adapt to new challenges—critical milestones in the journey toward general intelligence.

Author(s)	Research focus	Research gaps
Bécue et al.[62]	Explores AGI-driven innovation through the alignment of AI maturity, manufacturing strategies, and innovation capacity, emphasizing AGI's role in decision-making and complex problem-solving in Industry 5.0	Alignment of AI maturity with innovation metrics
Yue and Shyu[63]	Leverages AGI in the creation of intelligence fusion networks for proactive crisis management, incorporating principles of multisourced data integration, situational awareness, and decision-making	Scalability of AGI-driven intelligence fusion networks
Li et al.[18]	Utilizes AGI principles such as multimodal learning, domain-specific knowledge extraction, and operational optimization for addressing complex challenges in the geoenergy sector	Domain-specific knowledge for AGI application
Wu[42]	Examines AGI's impact on redefining professional roles and skills, focusing on AGI-human collaboration, ethical reasoning, and adaptability in an AGI-driven information environment	Skillset adaptation for AGI-driven environments
Chiroma et al.[64]	Focuses on AGI-related applications in IoMT, including explainable AI for healthcare decision-making, and predictive analytics for real-time health monitoring systems	Addressing security and privacy in IoMT
McLean et al.[37]	Analyzes AGI-related risks such as goal misalignment, autonomous decision-making, and the existential threats posed by AGI, proposing governance frameworks to mitigate such risks	Lack of standard AGI terminology and definitions
Wang et al.[65,66]	Reviews AGI-driven large vision models and visual prompt engineering, emphasizing AGI's capability to adapt prompts for generalizable and context-aware visual tasks	Designing efficient visual prompts for AGI systems
Daase and Turowski[67]	Proposes AGI-compatible methodologies for explainable AI, connecting behavioral and design sciences to develop general-purpose AGI systems for Society 5.0	Unified methodologies for explainable AI
Yang et al.[68]	Identifies AGI challenges in diagram analysis, focusing on AGI's ability to understand shape, topology, and content-based image retrieval for technical applications	Advances in diagram-specific retrieval techniques
Krishnan et al.[69]	Investigates AGI integration with disruptive technologies like IoT and autonomous systems, emphasizing its role in adaptive, scalable, and human-centered smart city frameworks	Implementation challenges for nested technologies
Krishnan et al.[70]	Explores AGI principles such as adaptability and multiagent interaction for optimizing disruptive technologies in smart cities, using empirical models to assess AGI's impact	Framework validation for disruptive AGI technologies
Long and Cotner[71]	Proposes a conceptual framework for AGI development, focusing on generalization, autonomy, and system-level integration for multidomain applications	Scalability of autonomous AGI systems
Everitt et al.[72]	Provides a comprehensive review of AGI safety challenges, addressing goal alignment, system control, and ethical AI deployment strategies	Gaps in comprehensive AGI safety strategies

Table 3. AGI-related review studies.

In summary, achieving AGI requires a paradigm shift that reimagines intelligence as an evolving, self-organizing process. While deep learning and big data have set the stage for AI advancements, their limitations highlight the need for cognitive architectures and novel frameworks that mirror human adaptability. Moreover, ethical and philosophical challenges must remain at the forefront to ensure that AGI development aligns with societal well-being and safety[25,26].

Methods

PRISMA protocol

The research design (Fig. 1) uses the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework[77] to ensure a systematic and rigorous approach to identifying and including studies relevant to AGI-related research. This framework was applied in a structured process that involved defining inclusion and exclusion criteria, conducting a comprehensive database search, screening retrieved studies at multiple levels (title and abstract), and synthesizing findings to ensure transparency and reproducibility. By following PRISMA guidelines, the study minimizes selection bias, enhances methodological rigor, and provides a structured approach for assessing the relevance and quality of AGI research contributions. To ensure a comprehensive and accurate analysis, this study utilized the Scopus database for literature retrieval. Scopus is widely recognized as a leading bibliographic database, offering significantly broader coverage than Web of Science, with approximately 60% more indexed publications[78,79]. The search in the Scopus database, covering the years 2003 to 2024, was conducted on October 30, 2024. The search terms were carefully designed to capture the full spectrum of AGI-related concepts, ranging from "Artificial General Intelligence" (AGI) to “Strong AI” and “Artificial Superintelligence” (ASI), reflecting the evolutionary trajectory of these technologies. The final search terms used in the titles and abstracts of publications were as follows: ("Artificial General Intelligence" OR “artificial GI” OR “strong AI” OR "strong artificial intelligence" OR "human level AI" OR "human level artificial intelligence" OR "artificial superintellig*" OR "artificial general intelligence system*" OR "human like AI" OR “superintelligent systems” OR “machine superintelligence”). A total of 1296 records were identified, with inclusion criteria limiting results to articles, reviews, conference papers, books, and book chapters in English. Following screening and eligibility assessment, all 1296 records met the predefined criteria, progressing to the final selection stage. This process underscores the methodological rigor in curating a focused dataset for examining AGI research trends and developments.

The systematic identification and selection of relevant studies through the PRISMA framework provided a robust foundation for applying topic modeling. Topic modeling was chosen over methods such as cocitation analysis, bibliographic coupling, and keyword co-occurrence because it goes beyond structural relationships in the literature, focusing instead on the semantic content of documents[80]. While cocitation and bibliographic coupling reveal historical connections and shared references between studies and keyword co-occurrence frequently highlights paired terms, topic modeling uncovers latent themes within the text, enabling a more context-rich exploration of AGI-related research trends and thematic evolution[81]. This approach provides deeper insights into emerging concepts and their semantic associations, which are essential for understanding a rapidly evolving field such as AGI.

BERTopic modeling

Various topic modeling methods are available, such as nonnegative matrix factorization (NMF), latent Dirichlet allocation (LDA), probabilistic latent semantic analysis (PLSA), and To2Vec. Despite their widespread use, these approaches often fail to account for semantic relationships between terms and face challenges when dealing with short-text formats[82]. Additionally, traditional models like LDA and PLSA rely on bag-of-words (BoW) frameworks, which treat words independently and disregard word order and contextual meaning, limiting their ability to capture nuanced language patterns, particularly in AGI research.

BERT, an advanced language model developed by Google, stands for bidirectional encoder representations from transformers[83]. In contrast to conventional topic modeling techniques that rely on BoW-based approaches, which focus solely on word frequency, BERTopic employs transformer-based embeddings[84] to encode documents into a lower-dimensional space while preserving semantic nuances. This allows BERTopic to dynamically cluster similar concepts and better capture evolving interdisciplinary research topics, such as AGI. Furthermore,

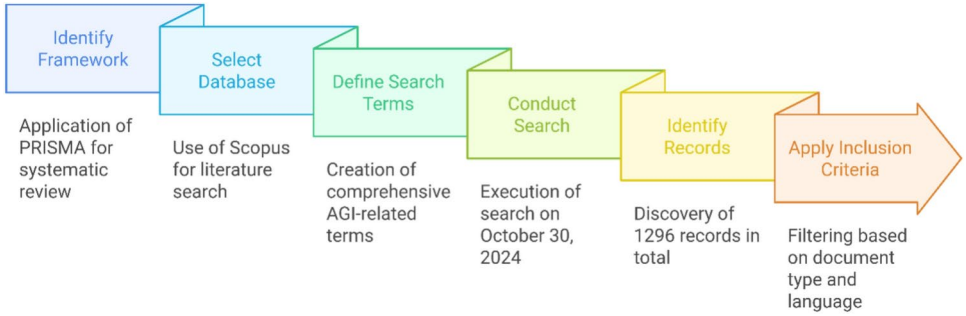

Fig. 1. Research design.

BERTopic integrates contextualized word embeddings, enhancing the interpretability of topics and reducing the need for extensive manual preprocessing[85,86].

Recent studies have demonstrated the superior performance of BERT-based topic modeling techniques over other topic modeling methods. Research comparing BERT-integrated neural topic models with conventional stochastic models found that BERT-based approaches significantly improve domain-specific text understanding and continuous learning Um and Kim[87]. BERTopic outperforms LDA and Top2Vec in topic clustering, achieving at least 34.2% better performance in Chinese and English datasets, with superior topic separation, clearer semantics, and a deeper understanding of text structure and content[88]. Given the need for a model that captures evolving trends and complex linguistic structures in AGI research, BERTopic offers a more adaptive and semantically rich alternative to conventional BoW-based topic modeling techniques.

The process of topic modeling begins with transforming input text into numerical embeddings through vectorization (Fig. 2). Dimensionality reduction is then performed via Unified Manifold Approximation and Projection (UMAP), which enhances the clarity of topic clusters by grouping similar data points[89]. Clusters are identified via hierarchical density-based spatial clustering of applications with noise (HDBSCAN), which detects dense regions of data while filtering out unrelated points. Key terms within each cluster are extracted via class-based term frequency-inverse document frequency (c-TF-IDF), which emphasizes terms based on their occurrence patterns across documents[90]. In this study, the "all-MiniLM-L6-v2" text representation model was utilized, as it is well suited for clustering and semantic search tasks[91]. Topics are assigned to documents based on these representative terms, with associated probabilities indicating each document's relevance to specific topics[92].

To refine topic modeling, three key hyperparameters were carefully tuned: the n-gram range, the number of topics, and the minimum topic size, ensuring optimal semantic richness, interpretability, and coherence of extracted topics.

The n-gram range was set to (1,2) to capture both single words and two-word phrases, striking a balance between granularity and contextual representation. Unigrams alone may fail to capture domain-specific terms, while longer n-grams can lead to data sparsity and model overfitting. This approach allowed the model to identify meaningful phrases without excessive noise, improving interpretability[84]. To determine the optimal number of topics, an iterative evaluation process was conducted, testing topic counts from 4 to 20. Each iteration was assessed using coherence scores (Cv, UMass, C_npmi), intertopic distance metrics, and manual inspection to ensure topics were distinct yet cohesive. The lower limit of four topics prevented overly broad themes, while the upper limit of sixteen was selected to avoid redundant, fragmented topics with minimal substantive difference. This tuning process was guided by coherence score stabilization, ensuring that increasing the number of topics did not lead to a drop in interpretability. A minimum topic size of 20 was applied to ensure that each topic had sufficient representation while avoiding excessively small, noisy topics that lacked generalizability. Additionally, the number of top words per topic was limited to 20 to emphasize the most relevant terms while maintaining topic coherence. Stopword removal was systematically applied to filter out generic, non-informative words such as "use," "add," and "related," which could otherwise dilute topic significance.

Dimensionality reduction was handled using Uniform Manifold Approximation and Projection (UMAP) with default settings, ensuring effective clustering in a reduced latent space while preserving semantic relationships. The "calculate probabilities" option was enabled to provide document-topic associations, improving soft clustering accuracy. Cosine similarity was used to measure the angular distance between topic embeddings, ensuring that closely related topics remained distinct while maintaining meaningful overlap. To further refine the topic modeling process, an extensive validation phase was conducted using manual qualitative assessments, coherence score comparisons, and intertopic distance distributions. A minimum intertopic distance threshold of 0.05 was applied, ensuring clear separation between topics while preventing excessive overlap. A random state of 100 was set for reproducibility, and n_neighbors was configured at 15 to balance local topic relationships with broader thematic patterns[93]. Through this multi-stage hyperparameter tuning process, the model identified five

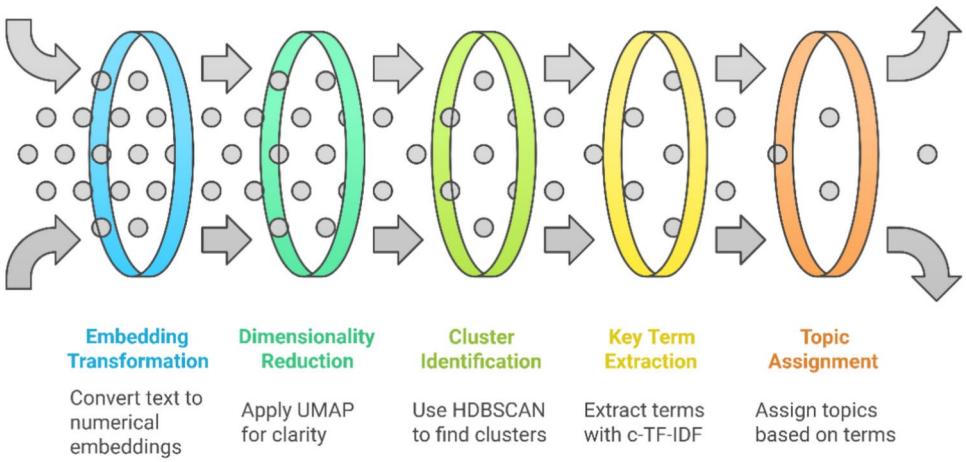

Fig. 2. BERTopic modeling steps.

Pathways	Theories/Frameworks mentioned	AGI concepts used	Keywords
Societal pathways of human-like AI	Computers as Social Actors (CASA), PESTEL Framework	Empathy, Interaction Quality, Anthropomorphic Design	"human level", "human-like AI", "AI safety", "decision making", "machine learning", "Turing test", "progression", "social change"
Technological pathways toward AGI	Explainable AI (XAI), Reward Maximization Hypothesis, Logical AI	Transparency, Reinforcement Learning, Representation Problems	"evolution", "advancements", "real-world effects", "artificial general intelligence", "big data", "human-level", "ethical considerations", "policy", "governance"
Pathways to explainability in AGI	Reinforcement Learning Framework, LDA Topic Modeling, AI Experimentation Platforms	Problem Solving, Task-Specific AI	"reinforcement learning", "machine learning", "artificial general intelligence", "natural language processing", "next-generation", "innovation", "trends"
Cognitive pathways and ethics in AGI	LIDA Model, Global Workspace Theory, AI-Completeness Theory	Moral Decision Making, Turing Test, Neurocomputers	"cognitive architectures", "machine consciousness", "common sense", "strong AI", "ethical considerations", "governance", "future"
Brain inspired pathways toward AGI	Hybrid Tianjic Architecture, Neuroevolution, Statistical Sample Complexity Theory	Real-time Processing, Network Optimization, High-Dimensional Processing	"neuromorphic computing", "working memory", "brain-inspired computing", "cutting-edge", "breakthrough", "projections", "next-generation"

Table 4. Overview of key pathways derived from BERTopic modeling.

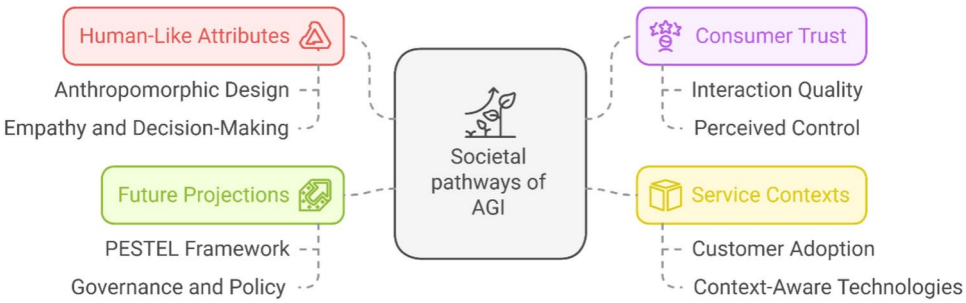

Fig. 3. Societal pathways of AGI.

distinct, well-separated topics that provided a coherent, interpretable representation of AGI-related research, ensuring both semantic accuracy and thematic depth.

To validate the accuracy of the machine learning outcomes, a manual review of the five identified topics and their representative publications was performed. Three domain experts assessed the coherence and relevance of the topics by qualitatively evaluating the associated keywords and publications, following approaches used in similar BERTopic studies[65,66,94]. Additionally, probability values and citation counts were factored into the selection process to identify the top three representative articles for each topic, ensuring that the unsupervised modeling provided meaningful and actionable insights.

Results

Table 4 offers an overview of the key pathways derived from BERTopic modeling, linking the technological evolution, societal impacts, and future trajectories of AGI. It highlights theories and AGI concepts, setting the stage for detailed exploration in the sections that follow.

Societal pathways of AGI

This theme highlights both the developmental trajectories and the broader societal impacts of AGI, particularly its human-like attributes (Fig. 3). Keywords such as "human level", "decision making", and "Turing test" reflect the dual focus on the theoretical evolution of AGI and its real-world consequences, connecting AGI's advancements to its transformative social, ethical, and economic effects.

The role of human-like attributes in the development and acceptance of AGI is critical to its technological progression. Pelau et al.[95] applied CASA theory to explore psychological anthropomorphic traits, perceived empathy, and interaction quality in service AI. While technological advancements have enabled AGI devices to perform tasks efficiently, consumer trust relies on interaction quality rather than mere human-like design. The incorporation of innovation and advancement trajectories into AGI devices enhances their social acceptance, particularly in service-oriented industries. The findings highlight the broader impact of AGI in maintaining human-centered service experiences and fostering ethical considerations around AI-human interactions.

Yang et al.[96] examined the advancement of AGI designs in various service contexts and their effects on customer adoption. Anthropomorphic attributes in AGI are innovative, yet their acceptance varies based on perceived control within social scenarios. When humans perceive high control, human-like AI is preferred, aligning with the goal of the AGI to mimic empathy and decision-making processes. Conversely, in low-control situations, highly anthropomorphic AGI induces apprehension, underscoring the nuanced effect of AGI on customer trust. This study's insights into design progression highlight the importance of context-aware AGI technologies for sustainable development.

Focusing on future projections and innovation, Kaplan and Haenlein[97] frame AGI within six key debates and dilemmas via the PESTEL framework. They identify the societal consequences and geopolitical trends of the AGI, emphasizing superintelligence, robotics, and the need for governance. As AGI capabilities evolve, governments and organizations must anticipate social change implications, including employment challenges, policy enforcement, and ethical considerations. By forecasting future development directions, this work bridges the technological evolution of AGI with its transformative effects on global structures.

Technological pathways toward AGI

Technological pathways and real-world implications of AGI highlight the developmental focus on the evolution, advancement, and integration of AGI while connecting these trajectories to real-world applications and challenges (Fig. 4). Keywords such as "reinforcement learning", "human level", "deep learning", and "neural networks" focus on the progress of AGI and its role in mimicking human cognitive ability. The theme captures both the technological pathways and the broader impact of AGI across sectors.

Sallab et al.[98] introduced a deep reinforcement learning (DRL) framework for autonomous driving, marking progress in AGI systems that interact with dynamic environments. Unlike traditional supervised learning approaches, DRL emphasizes evolution through trial-and-error, enabling machines to adapt to complex, real-time situations involving road interactions and uncertainties. Integrating neural networks and attention models reduces computational complexity, a significant step in AGI development. The findings highlight the role of AGI in automating critical decision-making processes while addressing challenges in partially observable environments. This trajectory emphasizes future advancements in AGI technologies and their impact on automotive systems, raising questions about policy and governance in AI-driven mobility.

The rapid progression of large language models (LLMs), such as ChatGPT, highlights the potential of AGI to impact jobs, creativity, and human–machine interactions. By analyzing early user responses via topic modeling, Taecharungroj[99] categorized ChatGPT's functionalities into domains such as creative writing, code generation, and question answering. These findings underscore the emerging applications of the AGI in diverse industries while simultaneously raising ethical considerations regarding job displacement and technology governance. This research exemplifies how technological advancements influence societal structures and economic implications as AGI evolves, encouraging deeper exploration into the responsible integration of human-like AI systems into daily life.

Hohenecker and Lukasiewicz[100] introduced Project Malmo, an AI experimentation platform built within Minecraft, which was designed to advance AGI systems capable of solving diverse tasks in complex environments. By enabling flexible agents to engage in problem-solving, collaboration, and learning, the platform mimics real-world challenges, fostering the progression of AGI toward human-like abilities. This work highlights innovation pathways for AGI development by testing learning algorithms and decision-making strategies across simulated environments. Such platforms not only accelerate technological advancements but also address broader consequences of AGI deployment, including adaptability, resource optimization, and ethical considerations.

Pathways to explainability in AGI

The proposed theme reflects the convergence of the "evolution", "progression", and "advancements" of the AGI while emphasizing its "real-world effects" (Fig. 5). Keywords such as "artificial general intelligence", "decision making", "big data", and "AI systems", align with both the developmental trajectories of AGI technologies and their broader implications. The theme connects the theoretical pathways of AGI to its applications while highlighting societal, economic, and ethical considerations.

Future advancements in explainable AI (XAI) should emphasize real-time interpretability and causal explanations, allowing stakeholders to understand AGI's decision-making processes in critical domains. These approaches will bridge the transparency gap and enhance user trust. Došilović et al.[101] identified advancements in explainable AI (XAI), addressing the transparency gap that often limits the practical deployment of AGI. By connecting explainability with deep learning, decision-making, and machine learning, their study contributes to the broader developmental trajectory of AGI while proposing pathways for future research. Transparency in

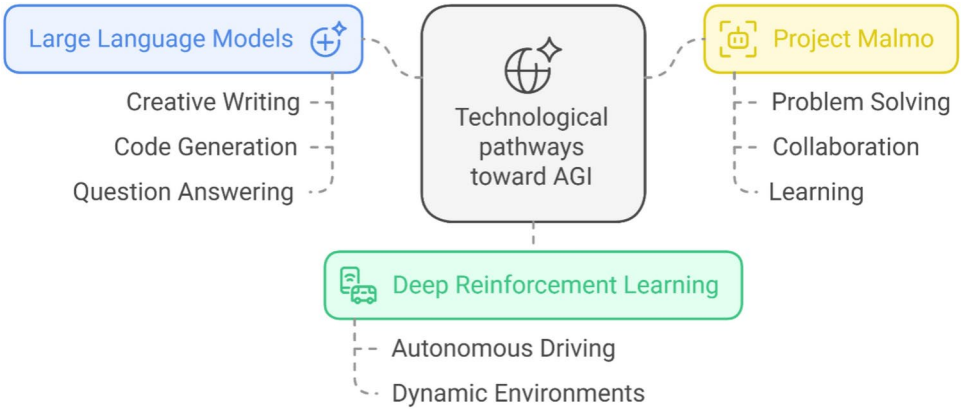

Fig. 4. Technological pathways toward AGI.

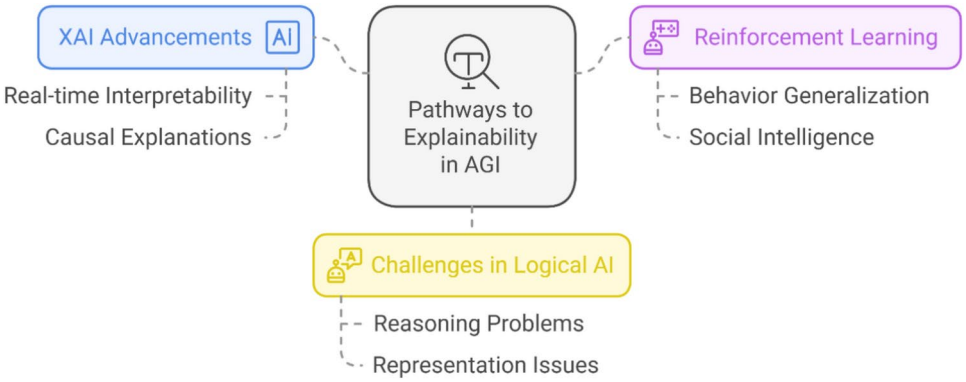

Fig. 5. Pathways to explainability in AGI.

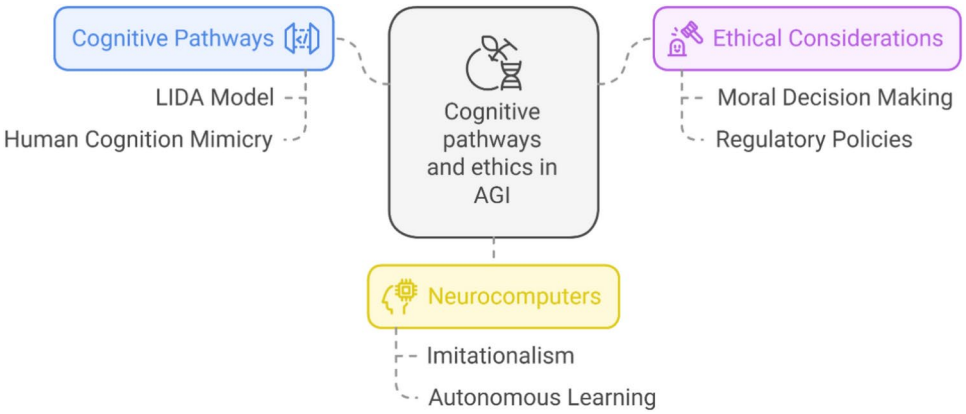

Fig. 6. Cognitive pathways and ethics in AGI.

AI fosters societal trust and mitigates ethical considerations related to algorithmic decisions, underscoring the real-world effects of AGI on critical systems where accountability is essential.

Advancing the discussion on generalization and learning systems in AGI, Silver et al.[102] hypothesize that maximizing reward can drive behavior similar to natural and artificial intelligence. This work emphasizes the progression of reinforcement learning as a potential solution to achieve artificial general intelligence, aligning with the development of AGI through trial-and-error methods. By connecting AGI's ability to generalize behaviors across multiple domains, the study sheds light on how future systems might address social intelligence and decision-making. Such frameworks highlight economic implications and policy considerations as AGI evolves toward human-like adaptability.

The complexity of achieving human-level AI is tied to overcoming brittleness in existing systems. McCarthy[103] identifies reasoning and representation problems as core challenges within logical AI, requiring breakthroughs in learning systems to account for common-sense scenarios. Addressing these limitations marks a significant evolution in AGI, where the ability to adapt to unpredictable conditions defines its potential. The discussion connects the pathways of AGI to its broader consequences for societal automation, emphasizing the need for scalable frameworks to achieve practical, human-level general intelligence.

Cognitive pathways and ethics in AGI

The theme of cognitive pathways and ethical considerations in AGI connects the development, evolution, and cognitive architectures of AGI with ethical and societal consequences (Fig. 6). Keywords such as "cognitive architectures", "machine consciousness", "common sense", "human cognition", and "moral decision making" illustrate AGI's capacity to mimic human reasoning while raising significant ethical questions. The theme balances AGI's developmental pathways with its social impact" and governance concerns.

Wallach et al.[104] explore moral decision-making within artificial general intelligence via the Learning Intelligent Distribution Agent (LIDA) cognitive model, which is rooted in global workspace theory. The LIDA model[105] offers a framework for understanding cognition, perception, and decision-making in AGI by simulating human-like attention, learning, and memory mechanisms. Integrating LIDA into AGI research can provide insights into ethical reasoning and adaptive decision-making, ensuring AGI systems align with human cognitive processes while addressing concerns of transparency, autonomy, and accountability. The research demonstrates how human cognition mechanisms—such as affective and rational processes—can be integrated

into AGI systems to address ethical problems. By mimicking human deliberation processes, the study highlights AGI's advancements toward making context-aware decisions, particularly in morally complex environments. This framework addresses ethical considerations and potential governance implications as autonomous agents become increasingly capable. The evolution of the AGI toward moral reasoning underscores the need for regulatory policies to ensure alignment with societal values.

AI completeness is formalized as a classification framework for evaluating AGI problems, with the Turing test identified as an AI-complete benchmark that signifies the progression of artificial systems toward general intelligence[106]. By reducing problems in AI-hard tasks, this research advances methodologies for understanding the capabilities and limitations of AGI. The classification system forecasts the AGI's future projections by identifying tasks requiring human-like cognition, including common sense reasoning. These advancements have economic implications as AGI systems increasingly interface with labor-intensive sectors requiring flexible, high-level problem-solving.

Brain-inspired neurocomputers are positioned as a next-generation approach to achieving AGI, emphasizing the imitationalism method, which focuses on developing neuromorphic systems that mimic brain structures rather than explicitly replicating human-level intelligence[107]. Neurocomputers incorporate cognitive architectures for autonomous learning and environmental interaction, accelerating the advancement of AGI across physical and computational domains. This development raises broader consequences for technology adoption in autonomous systems, underscoring the need for ethical and policy frameworks to govern AGI's integration into critical sectors such as robotics and healthcare.

Brain-inspired pathways toward AGI

The theme of bridging brain-inspired pathways toward AGI addresses the development of AGI at the intersection of neural networks, brain-inspired computing, and cognitive architectures (Fig. 7). Keywords such as "neuromorphic computing", "neural networks", and "working memory" highlight how AGI research combines inspiration from the human brain and advances in neural architectures. This theme connects technological progress with the pathways of AGI while acknowledging its societal implications.

Pei et al.[108] introduced the Tianjic chip, a groundbreaking hybrid platform that integrates both brain-inspired spiking neural networks (SNNs) and traditional artificial neural networks (ANNs), demonstrating its potential to bridge neuroscience and AI. This integration enables more efficient processing, real-time adaptability, and energy-efficient learning, making it a significant step toward AGI. By reconciling these traditionally distinct paradigms, the chip achieves real-time multitasking capabilities such as voice control, object detection, and obstacle avoidance. Such technological advancements mark a significant step toward developing next-generation hardware platforms essential for artificial general intelligence. This integration also reflects future trends in AGI systems, emphasizing adaptability and innovation. The chip's ability to process diverse tasks highlights its potential impact on robotics and autonomous systems, underscoring the need for policy frameworks addressing the ethical and societal implications of AGI deployment.

Neuroevolution leverages evolutionary algorithms to optimize neural network architectures, offering an alternative to the gradient-based approaches of Stanley et al.[109]. This method enhances advancements in AGI by enabling meta-learning processes, network customization, and optimization of building blocks such as activation functions and learning algorithms. This study explores evolution as a mechanism for AGI progression, drawing parallels to biological development. Neuroevolution supports trends such as parallel computation and innovation in deep reinforcement learning, pushing AGI closer to flexible cognitive systems. The impact of such developments spans economic applications, requiring governance to address scalability and accessibility while mitigating challenges such as system unpredictability.

Deep learning networks demonstrate unreasonable effectiveness in tasks ranging from language translation to image recognition. However, Sejnowski[110] addresses the gap in understanding why these networks perform so well, linking their success to high-dimensional geometry and advancements in optimization techniques. The foundational role of deep learning in AGI systems enables progression toward cognitive functions such as planning and general intelligence. The authors predict that breakthroughs will emerge by exploring brain regions

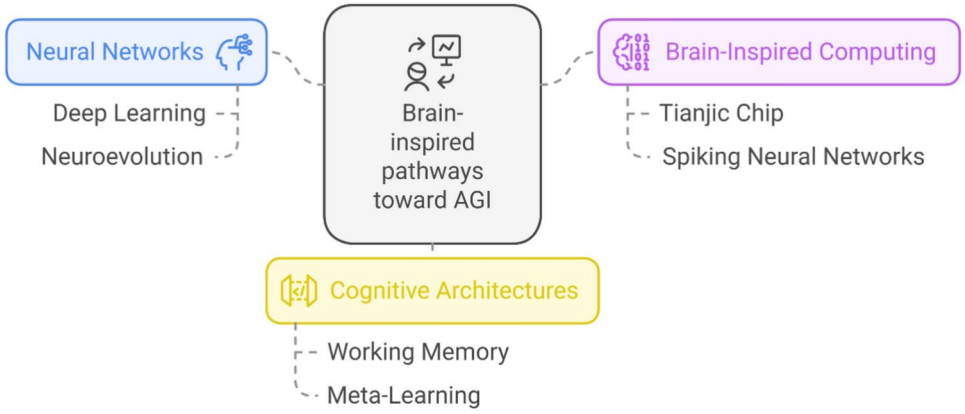

Fig. 7. Brain-inspired pathways toward AGI.

beyond the cerebral cortex, such as those responsible for survival behaviors. This next-generation development raises ethical considerations regarding dependency on black-box models and their effect on decision-making processes in critical fields.

Discussion

The current study not only clarified the distinctions among closely related AI constructs but also, via BERTopic modeling, provides a novel perspective on the challenges and possibilities associated with AGI research. The five pathways that emerged highlight the technological advancements that allow AGI technologies to more closely mirror human interactions and cognitions, the potentialities associated with doing so, as well as the inherent social, psychological, and economic costs. The study also integrates primary, rising, and novel themes, providing a roadmap for future directions in AGI research. Notably, the five core pathways reflect both the opportunities and the potential costs associated with AGI (Fig. 8).

The societal pathways of AGI highlight the complex intersection between technological evolution and societal ramifications, particularly as AGI systems become more human-like in interactions. As seen in previous studies[95,97], AGI is increasingly shaped by how well it engages empathetically and contextually with humans. While empathic AI can enhance trust and acceptance, over-reliance on AGI for emotionally driven decisions, ethical concerns regarding AI persuasion techniques, and unintended psychological effects of anthropomorphic AI interactions pose significant risks[96,111]. This duality underscores the need for a critical evaluation of AGI's design and deployment. While enhanced sensory technologies—such as vision, olfactory, and affective computing—could improve AGI's contextual understanding, excessive realism in human-like AI could lead to trust erosion, manipulation risks, and user discomfort. Additionally, biases in emotion recognition models and ethical concerns regarding AI's role in human decision-making require structured governance frameworks. Boltuc[112] suggests that integrating social scientists into AGI training and development can help ensure that these systems align with human values rather than merely replicating human behavior. Thus, while AGI's societal pathways offer transformative potential, they also introduce risks that require interdisciplinary oversight, transparent AI-human interactions, and regulatory safeguards to prevent misuse and ensure ethical alignment.

"Technological pathways toward AGI" connect AGI's development, practical challenges, and societal implications. AGI systems, driven by deep reinforcement learning and multimodal frameworks, address real-world tasks, advancing human-like reasoning for autonomous systems, games, and cognitive processes. However, rapid adoption, as seen with ChatGPT, raises societal challenges such as employment shifts and policy gaps[99,113]. Platforms such as Project Malmo highlight innovation and governance needs[100]. Future AGI systems aim to integrate social, emotional, attentional, and ethical intelligence to enhance problem-solving[114,115], but risks related to algorithmic bias, security vulnerabilities, and the unintended consequences of autonomous decision-making must be addressed. Additionally, the opacity of deep learning models raises concerns about accountability, particularly in high-stakes applications such as healthcare and finance. Overcoming challenges in interpretability, adaptability, and scalability is necessary. Therefore, AGI development must be accompanied by proactive regulatory measures, transparency standards[116], and interdisciplinary oversight to ensure its safe and equitable deployment. Advancing AGI will demand interdisciplinary collaboration and ethical frameworks to align technology with societal well-being.

"Pathways to explainability in AGI" highlights the critical need for transparent, trustworthy, and ethically aligned AGI systems. Techniques such as LIME and SHAP enhance transparency, which is essential for trust in sensitive areas such as healthcare[101]. Beyond technical clarity, models inspired by interspecies communication emphasize empathy and trust in human-AI collaboration[117]. Reinforcement learning, as a unifying principle, enables AGI to generalize across domains, linking technological progress to societal benefits[102]. However, explainability challenges persist, particularly in deep learning-based AGI, where complex decision-making processes remain opaque, limiting accountability in high-stakes applications such as law and finance. Additionally, reliance on post-hoc interpretability methods may not fully capture model behavior, leading to potential risks

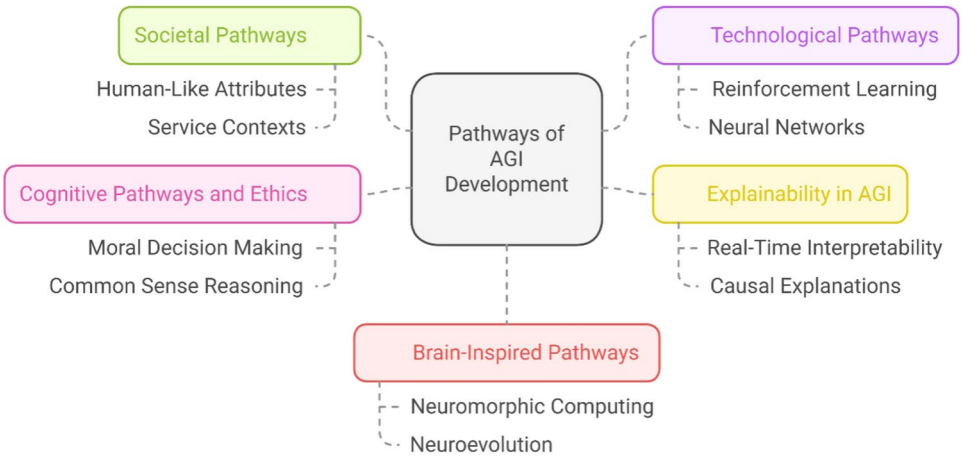

Fig. 8. Pathways of AGI development.

in bias mitigation and fairness. Without robust governance, AGI risks deepening inequalities and workforce disruptions[118,119]. Therefore, integrating explainability into AGI must go beyond technical transparency to include proactive regulatory oversight, interdisciplinary evaluation, and mechanisms for public accountability. These insights collectively underline the necessity of integrating explainability, ethical frameworks, and inclusive policies to ensure AGI's societal alignment and responsible evolution.

"Cognitive pathways and ethics in AGI" examines the interplay between the cognitive development of AGI and its societal impact. Frameworks such as LIDA[104,105] and brain-inspired neurocomputers[107] illustrate AGI's progression toward emulating human reasoning, enabling it to address complex tasks such as common-sense reasoning and moral decision-making. The classification of the Turing test as an AI-complete benchmark highlights the ongoing challenge of bridging narrow AI with general intelligence[106]. However, as AGI systems evolve, concerns around their ability to develop independent cognitive models and self-improve raise risks related to unpredictable decision-making, ethical misalignment, and loss of human oversight. Additionally, value alignment remains a persistent challenge, as AGI must navigate conflicting ethical frameworks across diverse global contexts. These advancements raise ethical concerns about decision-making, privacy, and value alignment[120], alongside governance challenges requiring transparency and accountability in sensitive applications. The economic implications, including workforce automation and task displacement, further underscore the need for policies that balance technological innovation with societal impact. To mitigate these risks, interdisciplinary collaboration between cognitive scientists, ethicists, and policymakers is essential in shaping regulatory safeguards, ensuring AGI development aligns with ethical principles, human oversight, and long-term sustainability goals. Neurocomputers and other AGI technologies forecast applications in robotics, healthcare, and decision-making, emphasizing the importance of interdisciplinary collaboration and regulatory measures to ensure that AGI development aligns with human values and sustainability goals[121].

"Brain inspired pathways toward AGI" highlights the integration of neuroscience-inspired and artificial neural network approaches to advance AGI research. Emerging neuromorphic hardware, such as brain-like chips and hybrid platforms, offer energy-efficient, real-time cognitive capabilities, which are essential for AGI's deployment in dynamic environments. This advancement expands the scope of brain-inspired computing to new applications, including robotics and adaptive learning systems. Hybrid platforms such as the Tianjic chip[108] and neuro-evolution methods using evolutionary algorithms[109] demonstrate innovations in optimizing neural network structures. Layered processing in brain-inspired AGI models supports abstraction, context awareness, and dynamic task prioritization, which are critical for general intelligence[122]. Research on brain-inspired AGI explores scaling, reasoning, and in-context learning, revealing both progress and limitations[123,124]. The role of deep learning provides foundational insights into high-dimensional processing and statistical modeling crucial for AGI[110]. Practical applications include detecting neurodegenerative disorders such as Parkinson's disease and Alzheimer's disease through IoT-based frameworks, highlighting diagnostic and treatment advancements[125]. However, the societal implications of AGI, including automation, privacy, and decision-making impacts, necessitate governance frameworks that ensure ethical use and equitable outcomes. The convergence of brain-inspired systems and computational advancements underscores the importance of balancing innovation with proactive regulation to address emerging risks.

Key challenges surrounding AGI

Achieving AGI involves overcoming several significant challenges. These challenges span technical, theoretical, ethical, societal, and economic domains, each presenting unique obstacles that must be addressed to realize AGI (Fig. 9). Embedded within each of these challenges are potential societal impacts, ethical considerations, economic implications, legal and regulatory challenges, and the psychological and sociological effects of achieving artificial general intelligence (AGI). Ensuring the security and safety of AGI systems will require robust mechanisms to address adversarial attacks, cyber threats, and autonomous decision-making risks. Additionally, scalability challenges must be addressed through energy-efficient models and hardware advancements. Addressing these challenges and navigating potential impacts requires a multidisciplinary approach that combines insights from computer science, cognitive science, ethics, and social sciences to develop AGI systems that are not only powerful but also safe and beneficial for humanity[117].

Technical challenges: AGI systems must support interactive, adaptive, and lifelong learning while effectively transferring knowledge across domains[126]. This necessitates advanced algorithms capable of processing diverse and dynamic data inputs. Additionally, the immense computational complexity and energy demands of AGI call for efficient algorithms and hardware to manage high resource usage[127–129]. Ensuring robustness and generalization is equally critical, as AGI systems must perform reliably across various tasks and handle noisy or incomplete data effectively[128,129]. These challenges highlight the need for transformative innovations to advance AGI development. AGI's technological trajectories and impact assessment underscore the need for a holistic approach—one that addresses scalability, ethical integrity, and inclusivity as AGI evolves. The convergence of machine learning advancements with ethical governance will be pivotal to ensuring that AGI serves humanity equitably, fostering innovation while respecting the complexities of human society. The studies collectively illuminate a future where the developmental pathways of AGI must align with societal goals, balancing progress with caution to navigate its profound and multifaceted implications.

Theoretical challenges: AGI development faces critical challenges in information processing and embodied cognition. Overcoming the combinatorial complexities in information ecosystems and managing selection pressures are pivotal to enhancing the information-processing capabilities of AGIs[130]. Additionally, incorporating embodied cognition, which enables AGI systems to understand and interact with the physical world as humans do, is crucial for ensuring that these systems can grasp the real-world implications of their actions[30,131]. Addressing these aspects is essential for creating truly capable and context-aware AGI systems.

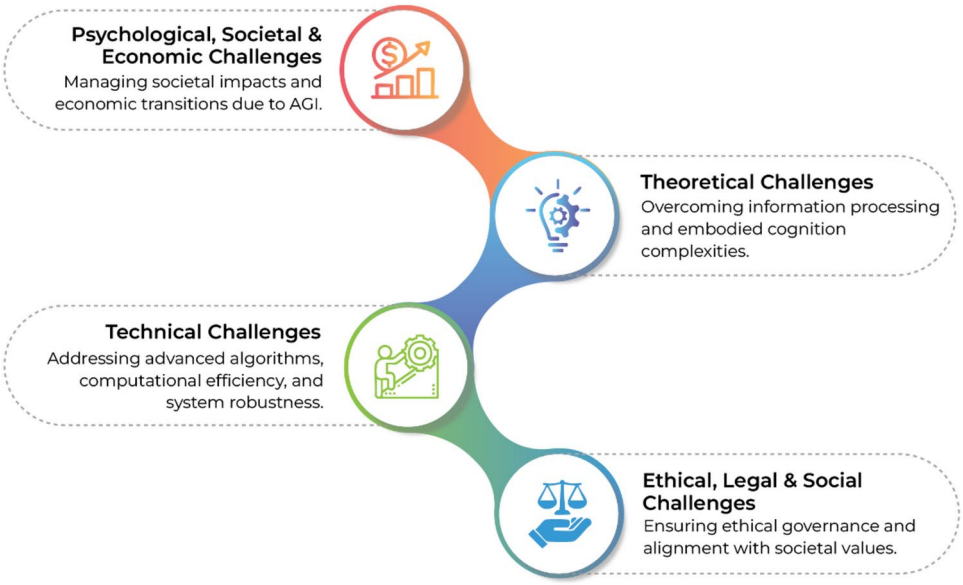

Fig. 9. Key challenges and impacts surrounding AGI.

Ethical, legal, and social challenges

The suddenness, timing, and potential danger of AGI necessitate careful ethical, legal, and regulatory considerations to guide its development responsibly[22]. Governance frameworks must prioritize scientific and technological ethics under the rule of law, emphasizing robust liability and data security systems to mitigate risks[32,132,133]. Additionally, the societal and political implications of AGI must be critically analyzed, particularly its effects on democratic values, equality, and electoral institutions, ensuring that its deployment aligns with broader societal goals[134]. Robust ethical and legal frameworks must address societal impacts, including privacy, employment, and accountability[135,136]. AGI development demands a focus on alignment, safety, regulation, and human–machine interaction[137,22]. Furthermore, effective human–machine interaction requires systems capable of understanding and responding to human emotions and social cues, ensuring safe and intuitive collaboration[126,135].

Psychological, societal, and economic challenges: The adoption of AGI could lead to profound psychological and sociological changes, including the possibility of humans becoming a minority in a population of autonomous AGIs. This scenario underscores the need for democratic systems to coordinate and ensure coexistence between humans and AGIs, safeguarding societal harmony and equity[33]. AGI's emergence has the potential to disrupt social order, presenting risks such as ethical dilemmas, liability attribution, intellectual property monopolies, and data security concerns[132,138]. Moreover, AGI could accumulate power and influence in society, increasing the risk of a "hard take off" scenario, where rapid advancements could surpass human control and governance structures[139]. Economically, the adoption of AGI could profoundly impact employment, requiring strategies to manage workforce transitions and maintain social stability[135]. Additionally, AGI systems must prioritize sustainable resource management to address global challenges, including health, education, and economic inequalities[140]. The creation of AGI could generate trillions in earnings for investors but poses risks such as a "technological singularity," which could devalue money and incentivize suboptimal behaviors among AGI businesses[141]. The widespread adoption of AGI also has significant economic, social, and political consequences, requiring detailed analyses of its effects on democratic systems and governance to ensure that it benefits society as a whole[134].

Implications

Implications for theory

This study contributes to advancing theoretical understanding in the field of AGI by explicitly engaging with and extending key theoretical frameworks across the technical, ethical, cognitive, and societal domains. This highlights critical areas where existing theories can evolve or be adapted to address the unique challenges and opportunities presented by AGI.

First, this research emphasizes the need to advance explainable AI (XAI) frameworks, reinforcing their importance in enhancing transparency, interpretability, and user trust in AGI systems. By applying XAI principles to AGI, this study addresses gaps in current theories related to decision-making, accountability, and alignment with societal goals. This expansion is crucial for ensuring that AGI systems operate ethically and responsibly in diverse real-world contexts.

Second, the study draws on global workspace theory and cognitive architectures, such as the learning intelligent distribution agent (LIDA), to explore the potential of AGI for mimicking human cognition. It extends these theories by incorporating elements of embodied cognition, emphasizing AGI's ability to interact with and adapt to physical environments. This integration addresses theoretical gaps in AGI adaptability and

common-sense reasoning. Additionally, this research leverages reinforcement learning frameworks and theories of neuroevolution to explore the developmental pathways of AGI. It advances these theories by introducing brain-inspired and hybrid architectures that combine symbolic reasoning with neural network learning, thus addressing limitations in scalability, generalization, and dynamic decision-making in current systems.

Third, the study applies collective intelligence frameworks to AGI, exploring how these systems can harness collaborative problem-solving capabilities across domains. This theoretical extension opens pathways for integrating AGI into global challenges such as climate change and disaster management, offering new perspectives on interdisciplinary applications. From an ethical standpoint, this research builds on existing moral and ethical decision-making theories by proposing frameworks for governance, inclusivity, and safety in AGI development. This highlights the theoretical necessity of aligning AGI objectives with human values to mitigate risks such as automation bias and existential threats, thereby advancing the discourse on ethical AI systems.

Finally, the study contributes to the theoretical discourse on human-like AI and consciousness by engaging in philosophical debates around AGI's potential sentience and autonomy. This calls for the refinement of theories such as Imitationalism and Brain-Inspired AI to address questions about moral agency, societal integration, and the ethical oversight of AGI.

Implications for practice

This study contributes to our understanding of AGI's implications for practitioners, particularly in workforce integration and upskilling, the design of transparent and trustworthy systems, contextual adaptability, and cross-disciplinary collaboration. While AGI offers significant opportunities, its implementation also presents risks such as job displacement, decision-making opacity, and socio-cultural biases, which must be proactively managed.

First, practitioners must prepare for AGI systems that integrate with human labor while mitigating job losses and skill redundancy. Although AGI can enhance collaborative decision-making in fields like education and renewable energy, its automation capabilities could disrupt traditional job markets. To address this, organizations must invest in worker retraining programs, hybrid workforce models, and AGI oversight roles. For example, in healthcare, AGI systems such as the Tianjic chip, which performs real-time diagnostics, require medical professionals to be trained in validating AGI-driven recommendations to ensure safe and ethical implementation. Similarly, in logistics, while AGI can optimize operations, it must be accompanied by upskilling initiatives to prevent mass displacement and support worker transitions into higher-value roles.

Second, practitioners must build AGI systems that are interpretable and aligned with societal expectations. While neurosymbolic AGI models can create personalized learning experiences, failure to explain decision-making processes could erode trust and exacerbate biases. In finance, AGI-driven decision-making tools must not only be transparent to regulators but also incorporate bias detection frameworks to prevent discriminatory lending or investment practices.

Third, AGI must be designed to function effectively across diverse sociocultural settings while preventing algorithmic bias and accessibility disparities. In education, AGI tutors must be tailored to linguistic and cultural variations to avoid reinforcing biases and widening educational inequalities. Similarly, in healthcare, AGI diagnostic tools must be adaptable to under-resourced settings, ensuring equitable access and preventing disparities in medical outcomes caused by training data biases favoring high-income regions.

Fourth, interdisciplinary collaboration is essential for refining AGI's capabilities while addressing its risks. Neuroscientists, computer scientists, and ethicists must work together to develop safeguards that prevent unintended consequences. For example, brain-inspired AGI systems used in telemedicine could improve empathy in virtual healthcare but may also raise concerns about data privacy, cognitive manipulation, or over-reliance on AI-driven diagnoses. Therefore, ethical oversight mechanisms must be embedded into AGI design to prevent misuse and ensure alignment with human values.

Implications for policy

Finally, this study has policy implications related to AGI. Five areas are particularly notable: regulatory frameworks for fair deployment; promoting inclusive access, ethical governance, and safety protocols; long-term risk mitigation; and incentivizing sustainability.

First, policymakers must establish clear and enforceable regulations to mitigate the adverse impacts of AGI on employment while ensuring economic growth. In manufacturing, policies should mandate human-in-the-loop oversight in AGI-driven automation to prevent mass job displacement while improving productivity. Governments could introduce workforce transition programs, tax incentives for companies that retain human workers alongside AGI systems, and mandatory reskilling initiatives to support displaced workers. In logistics, regulations should ensure the equitable distribution of AGI-driven efficiency gains by requiring profit-sharing models, retraining grants, and fair wage protections for employees affected by automation. Additionally, labor laws should be updated to address new employment models emerging from AGI integration, ensuring fair working conditions and preventing workforce exploitation in highly automated environments.

Second, policymakers must implement targeted policies to reduce disparities in AGI deployment, ensuring equitable access across education, healthcare, and other critical sectors. In education, government-backed subsidies and public–private partnerships can facilitate the adoption of AGI-powered tutoring systems in underprivileged schools, narrowing learning gaps and improving student outcomes. Additionally, national AI literacy programs should be established to equip students and teachers with the skills to integrate AGI tools into the learning process effectively. In healthcare, rural and underserved communities should benefit from subsidized AGI-driven diagnostic systems, telemedicine platforms, and predictive analytics, ensuring access to early disease detection, personalized treatment plans, and remote healthcare services. Regulatory frameworks

should also mandate fairness in AGI model training, preventing biases that could disproportionately disadvantage marginalized groups in both educational and healthcare applications.

Third, policymakers must enforce robust ethical guidelines to ensure that AGI systems align with human values, fairness, and accountability. In healthcare, AGI-driven systems, such as the Tianjic chip, should operate under strict ethical protocols that mandate bias auditing, explainability in diagnostic decisions, and compliance with global medical ethics standards to prevent discriminatory outcomes. Regulatory bodies should implement continuous oversight mechanisms, including third-party audits and real-time monitoring, to ensure AGI remains transparent and patient-centric. Similarly, in education, AGI-powered platforms must prioritize student privacy, adhering to strict data protection regulations (e.g., GDPR, FERPA) while ensuring algorithmic transparency and accountability in personalized learning decisions. Frameworks should also require AGI systems in education to offer human oversight options, allowing educators to intervene in automated recommendations to maintain ethical integrity and safeguard against unintended biases.

Fourth, policymakers must develop proactive regulations to address the existential risks and long-term consequences of AGI, particularly in high-stakes domains. In critical infrastructure, governance frameworks should mandate built-in fail-safes, redundancy mechanisms, and human override capabilities to prevent catastrophic failures caused by autonomous decision-making errors in sectors such as energy grids, defense systems, and financial markets. Regular stress-testing protocols and AI safety audits should be enforced to assess AGI's reliability in unpredictable scenarios. Additionally, policies should support the research and controlled development of AI systems, which are designed to provide constrained, advisory intelligence without direct autonomy, reducing the risks associated with the emergence of superintelligent AGI. Governments should also establish international AI safety coalitions to coordinate risk mitigation strategies, ensuring AGI deployment remains aligned with global security and ethical considerations.

Finally, policymakers should actively promote AGI applications that align with global sustainability goals, ensuring these technologies contribute to environmental and social well-being rather than exacerbating inequalities. In climate change modeling, AGI can be leveraged to optimize renewable energy deployment, enhance carbon footprint tracking, and improve climate resilience strategies by analyzing vast datasets to predict extreme weather patterns and natural disasters with greater accuracy. Governments should establish funding programs, tax incentives, and regulatory frameworks that encourage the development of AGI-driven solutions for sustainable agriculture, circular economy models, and biodiversity conservation. Additionally, international collaborations on AGI governance should prioritize equitable access to these sustainability-focused applications, ensuring that developing regions also benefit from AGI-driven innovations in climate adaptation and resource management.

Limitations and future directions

This study is not without limitations. Relying solely on Scopus within the PRISMA framework may introduce selection biases, omitting relevant studies from other databases like Web of Science, IEEE Xplore, and arXiv, potentially limiting the comprehensiveness of the review[142]. While the PRISMA framework ensures a systematic and transparent review, its structured approach may overlook emerging, interdisciplinary, or non-traditional studies[143], potentially limiting the breadth of AGI perspectives. Although BERT-based topic modeling excels in analyzing extensive datasets and extracting key themes, its performance is contingent upon the quality and scope of the input data. Notably, the HDBSCAN clustering algorithm it employs often classifies a significant portion of documents as outliers, excluding them from analysis[144]. By using semantic similarity to create cohesive topic representations, the method minimizes human bias in topic categorization, improving objectivity and consistency. Nonetheless, certain assumptions embedded within the algorithm design may also impose constraints. Relying on citation counts as impact indicators has inherent limitations, including affirmative citation bias, which can reinforce misconceptions[145], and the potential for misattributed citations, leading to inaccuracies in impact assessment[146].

As AGI continues to evolve, research has unveiled a spectrum of interconnected themes shaping its trajectory and societal implications (Fig. 10). Prevailing themes reflect longstanding interests and established directions, such as ethical governance frameworks and healthcare applications, which emphasize the need for responsible deployment and impactful use of AGI technologies. Moreover, rising themes capture emerging focal points, including the integration of collective intelligence principles and brain-inspired designs, which push the boundaries of AGI capabilities and alignment with human values. Finally, novel themes explore uncharted territories, such as the role of AGI in personalized education and its intersection with consciousness, paving the way for transformative advancements. These themes collectively provide a roadmap for understanding AGI development, its opportunities, and the challenges that lie ahead.

Prevailing theme: ethical guidelines and governance of AGI

The focus on ethical guidelines and governance highlights the need for structured frameworks guiding AGI development while addressing sustainability goals. As AGI becomes more autonomous, ethical and governance structures must mitigate risks such as job displacement and privacy breaches while aligning with the SDGs. AGI offers opportunities to support SDG priorities, such as renewable energy optimization and climate resilience strategies. Relevant frameworks include AGI ethics principles (e.g., transparency, accountability, and fairness), global governance models (such as the EU AI Act and IEEE's Ethically Aligned Design), and sustainability-driven AGI policies that integrate AGI into SDG-aligned initiatives. Future directions should first develop ethical guidelines that ensure accountability, inclusivity, and sustainability in AGI applications, particularly in areas such as renewable energy and climate modeling. Second, there should be the establishment of international governance structures promoting equitable AGI use to bridge digital divides and advance sustainability goals

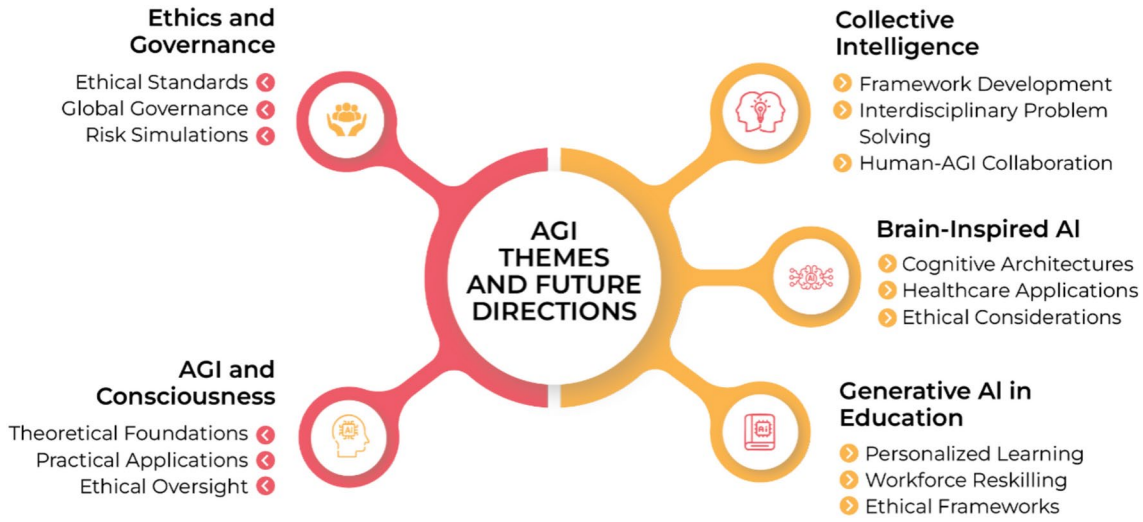

Fig. 10. AGI themes and future research directions.

such as clean energy and reduced inequalities. Third, risk simulations are needed to forecast the societal impacts of AGI and its potential contributions to SDG targets, such as resource optimization and educational access.

Rising theme: AGI and collective intelligence

Collective intelligence frameworks such as human-in-the-loop systems, swarm intelligence, and hybrid intelligence models offer opportunities to integrate real-time human expertise into AGI systems for addressing global challenges, such as disaster response, climate resilience, and epidemiological forecasting. By combining AGI's computational power with decentralized human insights, these frameworks can enhance adaptability, enabling faster and more context-aware decision-making in crisis scenarios. Leveraging collective intelligence can improve AGI's capacity to tackle complex, high-stakes challenges, from climate change mitigation strategies to public health crisis management, including pandemic response and resource allocation. For instance, in disaster management, AGI systems integrated with swarm intelligence can dynamically analyze large-scale sensor data and human reports to optimize relief efforts. Additionally, studies should explore how AGI can facilitate collaborative problem-solving in governance, urban planning, and global policy development, fostering innovative and inclusive decision-making that aligns with societal needs.

Rising theme: Brain-inspired AGI

The growing emphasis on biologically plausible AI models reflects a rising interest in designing AGI systems that replicate human cognitive processes, including learning, memory, and decision-making. This trend highlights the potential of brain-inspired architectures, such as neuromorphic computing and spiking neural networks (SNNs), to overcome traditional AI limitations in generalizability, adaptability, and reasoning. Future research should prioritize the development of hybrid AGI systems that integrate brain-like neural architectures with symbolic reasoning, enhancing contextual understanding and logical inference. Additionally, brain-inspired AGI holds transformative potential in healthcare and education, where it could revolutionize personalized medicine by enabling real-time diagnostics and treatment recommendations, as well as adaptive learning systems that respond dynamically to student behavior and cognitive states. Moreover, research should examine the ethical implications of AGI systems modeled on human cognition, particularly concerning privacy risks, cognitive manipulation, and psychological impact, ensuring that these advancements align with ethical and societal values.

Novel theme: AGI and consciousness

The intersection of AGI and consciousness represents an emerging frontier in understanding human-like cognition, self-awareness, and ethical reasoning within intelligent systems. This shift moves beyond purely functional AI toward AGI capable of introspection, adaptive decision-making, and moral judgment, with potential applications in eldercare, mental health support, and conflict resolution. Key advancements driving this area include cognitive architectures that simulate aspects of self-awareness, machine theories of consciousness aligned with cognitive science, and frameworks for ethical reasoning. Research should focus on developing theoretical foundations that guide the design of AGI with introspective capabilities while addressing moral and societal implications, including questions of agency, rights, and governance frameworks for responsible deployment. Future studies should explore methodologies such as computational modeling of consciousness, ethical simulations, and interdisciplinary policy frameworks to assess AGI's role in society. Additionally, practical applications must be explored, as AGI systems with conscious-like features could excel in emotionally intelligent domains such as eldercare, mental health counseling, and conflict resolution, offering nuanced interactions that enhance human well-being.

Novel theme: AGI in education

The convergence of generative AI and artificial general intelligence (AGI) has the potential to transform education and professional training by enabling adaptive, intelligent, and human-like learning systems. Future research should focus on three key areas. First, the development of AGI-driven generative AI platforms that dynamically adapt to individual learning styles and cognitive processes, delivering personalized educational experiences that evolve in real-time. Second, the creation of skill-based training programs that leverage AGI to simulate complex real-world scenarios, preparing professionals to collaborate with AGI-enabled systems in industries such as healthcare, engineering, and finance. Third, the establishment of ethical and regulatory guidelines to ensure the responsible and equitable deployment of AGI in education, addressing concerns related to data privacy, algorithmic bias, and accessibility to prevent exacerbating existing educational inequalities. By integrating AGI into education, learning environments can become more interactive, inclusive, and aligned with the evolving demands of the digital economy.

Conclusion

This study makes several significant contributions to AGI research by identifying underexplored themes using machine learning-based BERTopic modeling and establishing a conceptual framework that distinguishes key AI concepts while linking AGI to practical industry applications and ethical considerations. Our study critically addresses the pathways needed to enable scalable, adaptable, and explainable AGI across diverse environments by outlining strategic directions, including advancements in hybrid architectures, neuromorphic computing, and adaptive learning approaches, which enhance AGI's ability to generalize across contexts while maintaining interpretability. Additionally, we explore how AGI systems can be developed to align with ethical principles, societal needs, and equitable access, emphasizing the necessity for transparent governance, trust-building methodologies, and frameworks such as explainable AI (XAI) to ensure AGI serves human-centric objectives rather than exacerbating societal inequalities. Addressing the need for effective collaboration, trust, and transparency between humans and AGI systems, this study highlights the role of human-in-the-loop models, interdisciplinary oversight, and regulatory mechanisms that foster responsible AGI deployment. Finally, we examine how AGI can contribute to advancements through interdisciplinary integration, demonstrating its transformative potential in healthcare, education, sustainability, and decision-making systems. By linking AGI development to societal progress while acknowledging its risks, our findings offer a comprehensive roadmap for ensuring AGI's responsible and equitable evolution, meeting both current challenges and future global imperatives.

Data availability

All data generated or analyzed during this study are included in this published article.

Received: 10 January 2025; Accepted: 25 February 2025
Published online: 11 March 2025

References

1. Fjelland, R. Why general artificial intelligence will not be realized. *Hum. Soc. Sci. Commun.* **7**(1), 1–9 (2020).
2. Yampolskiy, R. On the Differences between Human and Machine Intelligence. In AISafety@ IJCAI (2021).
3. Hellström, T. & Bensch, S. Apocalypse now: no need for artificial general intelligence. *AI Soc.* **39**(2), 811–813 (2024).
4. Barrett, A. M. & Baum, S. D. A model of pathways to artificial superintelligence catastrophe for risk and decision analysis. *J. Exp. Theor. Artif. Intell.* **29**(2), 397–414 (2017).
5. Gill, K. S. Artificial super intelligence: beyond rhetoric. *AI Soc.* **31**, 137–143 (2016).
6. Pueyo, S. Growth, degrowth, and the challenge of artificial superintelligence. *J. Clean. Prod.* **197**, 1731–1736 (2018).
7. Saghiri, A. M., Vahidipour, S. M., Jabbarpour, M. R., Sookhak, M. & Forestiero, A. A survey of artificial intelligence challenges: Analyzing the definitions, relationships, and evolutions. *Appl. Sci.* **12**(8), 4054 (2022).
8. Ng, G. W. & Leung, W. C. Strong artificial intelligence and consciousness. *J. Artif. Intell. Conscious.* **7**(01), 63–72 (2020).
9. Yampolskiy, R. V. Agi control theory. In *Artificial General Intelligence: 14th International Conference, AGI 2021, Palo Alto, CA, USA, October 15–18, 2021, Proceedings 14* (pp. 316–326). Springer International Publishing (2022a).
10. Yampolskiy, R. V. On the controllability of artificial intelligence: An analysis of limitations. *J. Cyber Secur. Mob.* 321–404 (2022b).
11. Ramamoorthy, A. & Yampolskiy, R. Beyond mad? The race for artificial general intelligence. *ITU J.* **1**(1), 77–84 (2018).
12. Cao, L. Ai4tech: X-AI enabling X-Tech with human-like, generative, decentralized, humanoid and metaverse AI. *Int. J. Data Sci. Anal.* **18**(3), 219–238 (2024).
13. Goertzel, B. Artificial general intelligence and the future of humanity. The transhumanist reader: Classical and contemporary essays on the science, technology, and philosophy of the human future 128–137 (2013).
14. Li, X., Zhao, L., Zhang, L., Wu, Z., Liu, Z., Jiang, H. & Shen, D. Artificial general intelligence for medical imaging analysis. *IEEE Rev. Biomed. Eng.* (2025).
15. Rathi, S. Approaches to artificial general intelligence: An analysis (2022). arXiv preprint arXiv:2202.03153.
16. Arora, A. & Arora, A. The promise of large language models in health care. *The Lancet* **401**(10377), 641 (2023).
17. Bikkasani, D. C. Navigating artificial general intelligence (AGI): Societal implications, ethical considerations, and governance strategies. *AI Ethics* 1–16 (2024).
18. Li, J. X., Zhang, T., Zhu, Y. & Chen, Z. Artificial general intelligence for the upstream geoenergy industry: A review. *Gas Sci. Eng.* 205469 (2024).
19. Nedungadi, P., Tang, K. Y. & Raman, R. The transformative power of generative artificial intelligence for achieving the sustainable development goal of quality education. *Sustainability* **16**(22), 9779 (2024).
20. Zhu, X., Chen, S., Liang, X., Jin, X. & Du, Z. Next-generation generalist energy artificial intelligence for navigating smart energy. *Cell Rep. Phys. Sci.* **5**(9), 102192 (2024).
21. Faroldi, F. L. Risk and artificial general intelligence. *AI Soc.* 1–9 (2024).
22. Simon, C. J. Ethics and artificial general intelligence: technological prediction as a groundwork for guidelines. In *2019 IEEE International Symposium on Technology and Society (ISTAS)* 1–6 (IEEE, 2019).
23. Chouard, T. The Go files: AI computer wraps up 4–1 victory against human champion. *Nat. News* **20**, 16 (2016).

24. Morris, M. R., Sohl-Dickstein, J. N., Fiedel, N., Warkentin, T. B., Dafoe, A., Faust, A., Farabet, C. & Legg, S. Position: Levels of AGI for operationalizing progress on the path to AGI. In *International Conference on Machine Learning* (2023).
25. Weinbaum, D. & Veitas, V. Open ended intelligence: The individuation of intelligent agents. *J. Exp. Theor. Artif. Intell.* **29**(2), 371–396 (2017).
26. Sublime, J. The AI race: Why current neural network-based architectures are a poor basis for artificial general intelligence. *J. Artif. Intell. Res.* **79**, 41–67 (2024).
27. Wickramasinghe, B., Saha, G. & Roy, K. Continual learning: A review of techniques, challenges and future directions. *IEEE Trans. Artif. Intell.* (2023).
28. Fei, N. et al. Toward artificial general intelligence via a multimodal foundation model. *Nat. Commun.* **13**(1), 3094 (2022).
29. Kelley, D. & Atreides, K. AGI protocol for the ethical treatment of artificial general intelligence systems. *Procedia Comput. Sci.* **169**, 501–506 (2020).
30. McCormack, J. Autonomy, intention, Performativity: Navigating the AI divide. In *Choreomata* (pp. 240–257). Chapman and Hall/CRC (2023).
31. Khamassi, M., Nahon, M. & Chatila, R. Strong and weak alignment of large language models with human values. *Sci. Rep.* **14**(1), 19399 (2024).
32. McIntosh, T. R., Susnjak, T., Liu, T., Watters, P., Ng, A. & Halgamuge, M. N. A game-theoretic approach to containing artificial general intelligence: Insights from highly autonomous aggressive malware. *IEEE Trans. Artif. Intell.* (2024).
33. Salmi, J. A democratic way of controlling artificial general intelligence. *AI Soc.* **38**(4), 1785–1791 (2023).
34. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E. & Zhang, Y. Sparks of artificial general intelligence: Early experiments with gpt-4 (2023). arXiv preprint arXiv:2303.12712.
35. Faraboschi, P., Frachtenberg, E., Laplante, P., Milojicic, D. & Saracco, R. Artificial general intelligence: Humanity's downturn or unlimited prosperity. *Computer* **56**(10), 93–101 (2023).
36. Wei, Y. Several important ethical issues concerning artificial general intelligence. *Chin. Med. Ethics* **37**(1), 1–9 (2024).
37. McLean, S. et al. The risks associated with artificial general intelligence: A systematic review. *J. Exp. Theor. Artif. Intell.* **35**(5), 649–663 (2023).
38. Shalaby, A. Digital sustainable growth model (DSGM): Achieving synergy between economy and technology to mitigate AGI risks and address Global debt challenges. *J. Econ. Technol.* (2024a).
39. Croeser, S. & Eckersley, P. Theories of parenting and their application to artificial intelligence. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 423–428) (2019).
40. Lenharo, M. What should we do if AI becomes conscious? These scientists say it is time for a plan. *Nature* (2024). https://doi.org/10.1038/d41586-024-04023-8
41. Shankar, V. Managing the twin faces of AI: A commentary on "Is AI changing the world for better or worse?". *J. Macromark.* **44**(4), 892–899 (2024).
42. Wu, Y. Future of information professions: Adapting to the AGI era. *Sci. Tech. Inf. Process.* **51**(3), 273–279 (2024).
43. Sukhobokov, A., Belousov, E., Gromozdov, D., Zenger, A. & Popov, I. A universal knowledge model and cognitive architecture for prototyping AGI (2024). arXiv preprint arXiv:2401.06256.
44. Salmon, P. M. et al. Managing the risks of artificial general intelligence: A human factors and ergonomics perspective. *Hum. Fact. Ergon. Manuf. Serv. Ind.* **33**(5), 366–378 (2023).
45. Bory, P., Natale, S. & Katzenbach, C. Strong and weak AI narratives: An analytical framework. *AI Soc.* 1–11 (2024).
46. Gai, F. When artificial intelligence meets Daoism. In *Intelligence and Wisdom: Artificial Intelligence Meets Chinese Philosophers* 83–100 (2021).
47. Menaga, D. & Saravanan, S. Application of artificial intelligence in the perspective of data mining. In *Artificial Intelligence in Data Mining* (pp. 133–154). Academic Press (2021).
48. Adams, S., Arel, I., Bach, J., Coop, R., Furlan, R., Goertzel, B., Sowa, J. (2012). Mapping the landscape of human-level artificial general intelligence. AI Magazine 33(1), 25–42.
49. Beerends, S. & Aydin, C. Negotiating the authenticity of AI: how the discourse on AI rejects human indeterminacy. *AI Soc.* 1–14 (2024).
50. Besold, T. R. Human-level artificial intelligence must be a science. In *Artificial General Intelligence: 6th International Conference, AGI 2013, Beijing, China, July 31–August 3, 2013 Proceedings* 6 (pp. 174-177). Springer Berlin Heidelberg (2013).
51. Eth, D. The technological landscape affecting artificial general intelligence and the importance of nanoscale neural probes. *Informatica* **41**(4) (2017).
52. Nvs, B. & Saranya, P. L. Water pollutants monitoring based on Internet of Things. In *Inorganic Pollutants in Water* (pp. 371–397) (2020). Elsevier.
53. Flowers, J. C. Strong and Weak AI: Deweyan Considerations. In *AAAI spring symposium: Toward conscious AI systems* (Vol. 2287, No. 7) (2019).
54. Mitchell, M. Debates on the nature of artificial general intelligence. *Science* **383**(6689), eado7069 (2024).
55. Grech, V. & Scerri, M. Evil doctor, ethical android: Star Trek's instantiation of conscience in subroutines. *Early Hum. Dev.* **145**, 105018 (2020).
56. Noller, J. Extended human agency: Towards a teleological account of AI. *Hum. Soc. Sci. Commun.* **11**(1), 1–7 (2024).
57. Nominacher, M. & Peletier, B. Artificial intelligence policies. The digital factory for knowledge: Production and validation of scientific results 71–76 (2018).
58. Isaac, M., Akinola, O. M., & Hu, B. Predicting the trajectory of AI utilizing the Markov model of machine learning. In *2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI)* (pp. 30–34). IEEE (2023).
59. Stewart, W. The human biological advantage over AI. *AI & Soc.* 1–10 (2024).
60. Vaidya, A. J. Can machines have emotions? *AI Soc.* 1–16 (2024).
61. Triguero, I., Molina, D., Poyatos, J., Del Ser, J. & Herrera, F. General purpose artificial intelligence systems (GPAIS): Properties, definition, taxonomy, societal implications and responsible governance. *Inf. Fus.* **103**, 102135 (2024).
62. Bécue, A., Gama, J. & Brito, P. Q. AI's effect on innovation capacity in the context of industry 5.0: A scoping review. *Artif. Intell. Rev.* **57**(8), 215 (2024).
63. Yue, Y. & Shyu, J. Z. A paradigm shift in crisis management: The nexus of AGI-driven intelligence fusion networks and blockchain trustworthiness. *J. Conting. Crisis Manag.* **32**(1), e12541 (2024).
64. Chiroma, H., Hashem, I. A. T. & Maray, M. Bibliometric analysis for artificial intelligence in the internet of medical things: Mapping and performance analysis. *Front. Artif. Intell.* **7**, 1347815 (2024).
65. Wang, Z., Chen, J., Chen, J. & Chen, H. Identifying interdisciplinary topics and their evolution based on BERTopic. *Scientometrics* https://doi.org/10.1007/s11192-023-04776-5 (2023).
66. Wang, J., Liu, Z., Zhao, L., Wu, Z., Ma, C., Yu, S. & Zhang, S. Review of large vision models and visual prompt engineering. *Meta-Radiol.* 100047 (2023).
67. Daase, C. & Turowski, K. Conducting design science research in society 5.0—Proposal of an explainable artificial intelligence research methodology. In *International Conference on Design Science Research in Information Systems and Technology* (pp. 250–265). Springer, Cham (2023).
68. Yang, L., Gong, M. & Asari, V. K. Diagram image retrieval and analysis: Challenges and opportunities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 180–181) (2020).

69. Krishnan, B., Arumugam, S. & Maddulety, K. 'Nested'disruptive technologies for smart cities: Effects and challenges. *Int. J. Innov. Technol. Manag.* **17**(05), 2030003 (2020).
70. Krishnan, B., Arumugam, S. & Maddulety, K. Impact of disruptive technologies on smart cities: challenges and benefits. In *International Working Conference on Transfer and Diffusion of IT* (pp. 197–208). Cham: Springer International Publishing (2020).
71. Long, L. N. & Cotner, C. F. A review and proposed framework for artificial general intelligence. In *2019 IEEE Aerospace Conference* (pp. 1–10). IEEE (2019).
72. Everitt, T., Lea, G. & Hutter, M. AGI safety literature review (2018). arXiv preprint arXiv:1805.01109.
73. Wang, P. & Goertzel, B. Introduction: Aspects of artificial general intelligence. In *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms* (pp. 1–16). IOS Press (2007).
74. Yampolskiy, R. & Fox, J. Safety engineering for artificial general intelligence. *Topoi* **32**, 217–226 (2013).
75. Dushkin, R. V. & Stepankov, V. Y. Hybrid bionic cognitive architecture for artificial general intelligence agents. *Procedia Comput. Sci.* **190**, 226–230 (2021).
76. Nyalapelli, V. K., Gandhi, M., Bhargava, S., Dhanare, R. & Bothe, S. Review of progress in artificial general intelligence and human brain inspired cognitive architecture. In *2021 International Conference on Computer Communication and Informatics* (ICCCI) (pp. 1–13). IEEE (2021).
77. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71. https://doi.org/10.1136/bmj.n71 (2021).
78. Donthu, N., Kumar, S., Pandey, N., Pandey, N. & Mishra, A. Mapping the electronic word-of-mouth (eWOM) research: A systematic review and bibliometric analysis. *J. Bus. Res.* **135**, 758–773 (2021).
79. Comerio, N. & Strozzi, F. Tourism and its economic impact: A literature review using bibliometric tools. *Tour. Econ.* **25**(1), 109–131 (2019).
80. Raman, R., Gunasekar, S., Dávid, L. D. & Nedungadi, P. Aligning sustainable aviation fuel research with sustainable development goals: Trends and thematic analysis. *Energy Rep.* **12**, 2642–2652 (2024).
81. Raman, R., Gunasekar, S., Kaliyaperumal, D. & Nedungadi, P. Navigating the nexus of artificial intelligence and renewable energy for the advancement of sustainable development goals. *Sustainability* **16**(21), 1–25 (2024).
82. Egger, R. & Yu, J. A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Front. Sociol.* **7**, 886498 (2022).
83. Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019—2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies—Proceedings of the Conference*, 1(Mlm), 4171–4186 (2019).
84. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TFIDF procedure (2022). http://arxiv.org/abs/2203.05794
85. Alamsyah, A. & Girawan, N. D. Improving clothing product quality and reducing waste based on consumer review using RoBERTa and BERTopic language model. *Big Data Cogn. Comput.* **7**(4), 168 (2023).
86. Raman, R. et al. Green and sustainable AI research: An integrated thematic and topic modeling analysis. *J. Big Data* **11**(1), 55 (2024).
87. Um, T. & Kim, N. A study on performance enhancement by integrating neural topic attention with transformer-based language model. *Appl. Sci.* **14**(17), 7898 (2024).
88. Gan, L., Yang, T., Huang, Y., Yang, B., Luo, Y. Y., Richard, L. W. C. & Guo, D. Experimental comparison of three topic modeling methods with LDA, Top2Vec and BERTopic. In *International Symposium on Artificial Intelligence and Robotics* (pp. 376–391). Singapore: Springer Nature Singapore (2023).
89. Yi, J., Oh, Y. K. & Kim, J. M. Unveiling the drivers of satisfaction in mobile trading: Contextual mining of retail investor experience through BERTopic and generative AI. *J. Retail. Consum. Serv.* **82**, 104066 (2025).
90. Oh, Y. K., Yi, J. & Kim, J. What enhances or worsens the user-generated metaverse experience? An application of BERTopic to Roblox user eWOM. *Internet Res.* (2023).
91. Kim, K., Kogler, D. F. & Maliphol, S. Identifying interdisciplinary emergence in the science of science: Combination of network analysis and BERTopic. *Hum. Soc. Sci. Commun.* **11**(1), 1–15 (2024).
92. Khodeir, N. & Elghannam, F. Efficient topic identification for urgent MOOC Forum posts using BERTopic and traditional topic modeling techniques. *Educ. Inf. Technol.* 1–27 (2024).
93. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction (2018).
94. Douglas, T., Capra, L. & Musolesi, M. A computational linguistic approach to study border theory at scale. *Proc. ACM Hum. Comput. Interact.* **8**(CSCW1), 1–23 (2024).
95. Pelau, C., Dabija, D. C. & Ene, I. What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. *Comput. Hum. Behav.* **122**, 106855 (2021).
96. Yang, Y., Liu, Y., Lv, X., Ai, J. & Li, Y. Anthropomorphism and customers' willingness to use artificial intelligence service agents. *J. Hosp. Mark. Manag.* **31**(1), 1–23 (2022).
97. Kaplan, A. & Haenlein, M. Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Bus. Horizons* **63**(1), 37–50 (2020).
98. Sallab, A. E., Abdou, M., Perot, E. & Yogamani, S. Deep reinforcement learning framework for autonomous driving (2017). arXiv preprint arXiv:1704.02532.
99. Taecharungroj, V. "What can ChatGPT do?" Analyzing early reactions to the innovative AI chatbot on Twitter. *Big Data Cogn. Comput.* **7**(1), 35 (2023).
100. Hohenecker, P. & Lukasiewicz, T. Ontology reasoning with deep neural networks. *J. Artif. Intell. Res.* **68**, 503–540 (2020).
101. Došilović, F. K., Brčić, M. & Hlupić, N. Explainable artificial intelligence: A survey. In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 0210–0215). IEEE (2018).
102. Silver, D., Singh, S., Precup, D. & Sutton, R. S. Reward is enough. *Artif. Intell.* **299**, 103535 (2021).
103. McCarthy, J. From here to human-level AI. *Artif. Intell.* **171**(18), 1174–1182 (2007).
104. Wallach, W., Franklin, S. & Allen, C. A conceptual and computational model of moral decision making in human and artificial agents. *Top. Cogn. Sci.* **2**(3), 454–485 (2010).
105. Franklin, S. *Artificial Minds*. MIT Press, p. 412 (1995).
106. Yampolskiy, R. V. Turing test as a defining feature of AI-completeness. Artificial Intelligence, Evolutionary Computing and Metaheuristics: In the Footsteps of Alan Turing, 3–17 (2013).
107. Arcas, B. A. Do large language models understand us?. *Daedalus* **151**(2), 183–197 (2022).
108. Pei, J. et al. Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature* **572**(7767), 106–111 (2019).
109. Stanley, K. O., Clune, J., Lehman, J. & Miikkulainen, R. Designing neural networks through neuroevolution. *Nat. Mach. Intell.* **1**(1), 24–35 (2019).
110. Sejnowski, T. J. The unreasonable effectiveness of deep learning in artificial intelligence. *Proc. Natl. Acad. Sci.* **117**(48), 30033–30038 (2020).
111. Wang, J. & Pashmforoosh, R. A new framework for ethical artificial intelligence: Keeping HRD in the loop. *Hum. Resour. Dev. Int.* **27**(3), 428–451 (2024).
112. Boltuc, P. Human-AGI Gemeinschaft as a solution to the alignment problem. In *International Conference on Artificial General Intelligence* (pp. 33–42). Cham: Springer Nature Switzerland (2024).

113. Naudé, W. & Dimitri, N. The race for an artificial general intelligence: Implications for public policy. *AI Soc.* **35**, 367–379 (2020).
114. Cichocki, A. & Kuleshov, A. P. Future trends for human-AI collaboration: A comprehensive taxonomy of AI/AGI using multiple intelligences and learning styles. *Comput. Intell. Neurosci.* **2021**(1), 8893795 (2021).
115. Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. Building machines that learn and think like people. *Behav. Brain Sci.* **40**, e253 (2017).
116. Shalaby, A. Classification for the digital and cognitive AI hazards: urgent call to establish automated safe standard for protecting young human minds. *Digit. Econ. Sustain. Dev.* **2**(1), 17 (2024).
117. Liu, C. Y. & Yin, B. Affective foundations in AI-human interactions: Insights from evolutionary continuity and interspecies communications. *Comput. Hum. Behav.* **161**, 108406 (2024).
118. Dong, Y., Hou, J., Zhang, N. & Zhang, M. Research on how human intelligence, consciousness, and cognitive computing affect the development of artificial intelligence. *Complexity* **2020**(1), 1680845 (2020).
119. Muggleton, S. Alan turing and the development of artificial intelligence. *AI Commun.* **27**(1), 3–10 (2014).
120. Liu, Y., Zheng, W. & Su, Y. Enhancing ethical governance of artificial intelligence through dynamic feedback mechanism. In *International Conference on Information* (pp. 105–121). Cham: Springer Nature Switzerland (2024).
121. Qu, P. et al. Research on general-purpose brain-inspired computing systems. *J. Comput. Sci. Technol.* **39**(1), 4–21 (2024).
122. Nadji-Tehrani, M. & Eslami, A. A brain-inspired framework for evolutionary artificial general intelligence. *IEEE Trans. Neural Netw. Learn. Syst.* **31**(12), 5257–5271 (2020).
123. Huang, T. J. Imitating the brain with neurocomputer a "new" way towards artificial general intelligence. *Int. J. Autom. Comput.* **14**(5), 520–531 (2017).
124. Zhao, L. et al. When brain-inspired ai meets agi. *Meta-Radiol.* **1**, 100005 (2023).
125. Qadri, Y. A., Ahmad, K. & Kim, S. W. Artificial general intelligence for the detection of neurodegenerative disorders. *Sensors* **24**(20), 6658 (2024).
126. Kasabov, N. K. Neuroinformatics, neural networks and neurocomputers for brain-inspired computational intelligence. In *2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics* (SACI) (pp. 000013–000014). IEEE (2023).
127. Stiefel, K. M. & Coggan, J. S. The energy challenges of artificial superintelligence. *Front. Artif. Intell.* **6**, 1240653 (2023).
128. Yang S. & Chen, B. Effective surrogate gradient learning with high-order information bottleneck for spike-based machine intelligence. *IEEE Trans. Neural Netw. Learn. Syst.* (2023).
129. Yang, S. & Chen, B. SNIB: Improving spike-based machine learning using nonlinear information bottleneck. *IEEE Trans. Syst. Man Cybern. Syst.* **53**, 7852 (2023).
130. Walton, P. Artificial intelligence and the limitations of information. *Information* **9**(12), 332 (2018).
131. Ringel Raveh, A. & Tamir, B. From homo sapiens to robo sapiens: the evolution of intelligence. *Information* **10**(1), 2 (2018).
132. Chen, B. & Chen, J. China's legal practices concerning challenges of artificial general intelligence. *Laws* **13**(5), 60 (2024).
133. Mamak, K. AGI crimes? The role of criminal law in mitigating existential risks posed by artificial general intelligence. *AI Soc.* 1–11 (2024).
134. Jungherr, A. Artificial intelligence and democracy: A conceptual framework. *Soc. Media+ Soc.* **9**(3), 20563051231186353 (2023).
135. Guan, L. & Xu, L. The mechanism and governance system of the new generation of artificial intelligence from the perspective of general purpose technology. *Xitong Gongcheng Lilun yu Shijian/Syst. Eng. Theory Pract.* **44**(1), 245–259 (2024).
136. Pregowska, A. & Perkins, M. Artificial intelligence in medical education: Typologies and ethical approaches. *Ethics Bioethics* **14**(1–2), 96–113 (2024).
137. Bereska, L. & Gavves, E. Taming simulators: Challenges, pathways and vision for the alignment of large language models. In *Proceedings of the AAAI Symposium Series* (Vol. 1, No. 1, pp. 68–72) (2023).
138. Chen, G., Zhang, Y. & Jiang, R. A novel artificial general intelligence security evaluation scheme based on an analytic hierarchy process model with a generic algorithm. *Appl. Sci.* **14**(20), 9609 (2024).
139. Sotala, K. & Yampolskiy, R. Risks of the journey to the singularity. *Technol. Singular. Manag. J.* 11–23 (2017).
140. Mercier-Laurent, E. The future of AI or AI for the future. *Unimagined Futures–ICT Opportunities and Challenges* 20–37 (2020).
141. Miller, J. D. Some economic incentives facing a business that might bring about a technological singularity. In Singularity hypotheses: A scientific and philosophical assessment (pp. 147–159). Berlin, Heidelberg: Springer Berlin Heidelberg (2013).
142. Bramer, W. M., Rethlefsen, M. L., Kleijnen, J. & Franco, O. H. Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Syst. Rev.* **6**, 1–12 (2017).
143. Mishra, V. & Mishra, M. P. PRISMA for review of management literature–method, merits, and limitations—An academic review. *Advancing Methodologies of Conducting Literature Review in Management Domain*, 125–136 (2023).
144. de Groot, M., Aliannejadi, M. & Haas, M. R. Experiments on generalizability of BERTopic on multi-domain short text (2022). arXiv preprint arXiv:2212.08459.
145. Letrud, K. & Hernes, S. Affirmative citation bias in scientific myth debunking: A three-in-one case study. *PLoS ONE* **14**(9), e0222213 (2019).
146. McCain, K. W. Assessing obliteration by incorporation in a full-text database: JSTOR, Economics, and the concept of "bounded rationality". *Scientometrics* **101**, 1445–1459 (2014).

Author contributions

R.R.: Conceptualization: methodology, data curation, original draft, writing, review and editing. R.K.: writing—original draft; review and editing. K.A.: writing—original draft; review and editing. A.I.: writing—original draft; review and editing. P.N.: writing—original draft; review and editing. All authors reviewed the manuscript.

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Competing interests

The authors declare no competing interests.

Consent for publication

During the preparation of this work, the authors utilized ChatGPT 4o for editing and grammar checks. Subsequent to using this tool, the authors reviewed and edited the content as needed, assuming full responsibility for the content of the publication.

Additional information

Correspondence and requests for materials should be addressed to R.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.