

Review

Understanding Artificial General Intelligence: Defining Characteristics and Benchmarks

Dena Kadhim Muhsen*, Ahmed T. Sadiq

Computer Science College, University of Technology - Iraq, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq

Article History:

Received: 18 May 2025

Revised: 27 May 2025

Accepted: 18 June 2025

Published: 26 June 2025

Abstract: Artificial General Intelligence (AGI) is an advanced form of Artificial Intelligence (AI) that aims to perform any intellectual task that a human can, surpassing the narrow scope of Artificial Narrow Intelligence (ANI). Unlike ANI, which is designed to handle specific, well-defined tasks, AGI has the potential to generalize knowledge and adapt across a variety of domains. This research examines AGI's learning and adaptability, common-sense thinking, autonomous decision-making, transfer learning, creativity, and problem-solving. It examines AGI history using cognitive science, neuroscience, machine learning, and other disciplines. This topic requires trustworthy criteria to measure AGI research progress. This research paper serves as a review of AGI, summarizing its defining traits, challenges, and evaluation benchmarks. It also surveys key recent advancements in AGI systems, integrating perspectives from cognitive science, machine learning, and neuroscience. Evaluation methods such as the Turing Test, Winograd Schema Challenge, Coffee Test, and Lovelace 2.0 Test are assessed for their relevance to AGI.

Keywords: AGI; ANI; Autonomous Decision-Making; Transfer Learning; AGI Benchmarks

1. Introduction

1.1. Background Information

"Artificial General Intelligence" (AGI) can reason, solve problems, learn, and adapt like a person [1]. ANI excels in limited tasks like language translation, picture identification, and chess playing, whereas AGI can generalize and transfer learning without retraining [2]. Cognitive flexibility changed AGI, since AGI represents a shift from specialize automation to truly independent and adaptable intelligence, AI has long sought its development [3]. AGI could affect medicine, academia, business, robotics, and research if properly designed and implemented. AGI might speed drug discovery, enhance diagnostics, and personalize healthcare [4]. It may develop student-specific lesson plans. The capacity of computers to mimic human flexibility and solve complicated real-world problems might enhance robotics [5]. However,

AGI raises serious ethical, security, and government oversight concerns.

The likelihood that AGI would transcend human intelligence raises questions about its control, alignment with human ideals, and impact on employment. Job losses due to automation may reform economies, raising issues about accountability and bias in AGI-driven law enforcement and governance decisions [6]. Safe and ethical development of AGI is necessary to avoid negative results like eliminating human supervision or abusing it in risky situations. The ethical and effective integration of AGI into society depends on recognising its properties and setting rigorous evaluation requirements.

1.2. Research Problem

Despite growing interest in AGI, explaining its aspects is difficult. Lack of a widely accepted definition of AGI causes

* Corresponding author: Dena Kadhim Muhsen, Computer Science College, University of Technology - Iraq, Baghdad, Iraq, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq, Dena.K.Muhsen@uotechnology.edu.iq.

inconsistencies in how academics conceptualise and approach its development. Another issue is that current AI standards, like the Turing Test, only assess specific AI abilities, not general intelligence, this complicates AGI evaluation. Strong AGI evaluation criteria are needed to further research, implement safety measures, and prepare policymakers for AGI implications. Without benchmarks to measure AGI growth, it is hard to tell if it meets ethical and safety standards.

1.3. Research Objectives

The primary objective of this research paper is to explore the defining characteristics of AGI and establish key benchmarks for its evaluation. This study aims to:

- (1) Examine the fundamental traits that differentiate AGI from narrow AI.
- (2) Analyze existing AI evaluation methods and their applicability to AGI.
- (3) Identify gaps in current AGI assessment frameworks and propose new benchmarking criteria.
- (4) Discuss the challenges associated with measuring AGI's intelligence, adaptability, and reasoning capabilities.

1.4. Research Questions

- (1) What are the defining characteristics of AGI?
- (2) How can AGI be measured and evaluated effectively?

2. Literature Review

2.1. Historical Background of AGI

AGI evolved in the mid-20th century field of Artificial Intelligence (AI) [7]. At the 1956 Dartmouth Conference, scientists discussed the possibility of computers with human-like intelligence. John McCarthy coined the term "Artificial Intelligence" [8]. AI research first focused on rule-based and symbolic reasoning to emulate human thought processes using logic and computers [9].

In the 1950s and 1960s, Alan Turing, Herbert Simon, and Marvin Minsky produced theoretical models that predicted machines will exceed humans in intellect [10]. Turing's "Turing Test" for machine intelligence paved the way for decades of AI research. ANI began with task-specific AI and ANI excels at domain-specific tasks but cannot adapt or transmit knowledge [11]. ANIs include Deep Blue, a chess program, Alexa, a speech recognition system, and deep learning, an image identification model. ANI research increased throughout the 1980s and 1990s with neural networks, expert systems, and machine learning [12]. True intelligence involves rule-based programming, reasoning,

autonomous learning, and task generalisation [13]. This breakthrough revived artificial general intelligence machines that can learn, reason, and understand like humans.

AI has advanced beyond ANI to AGI due to cognitive science, neurology, and ML [14]. Cognitive science illuminates our natural learning, processing, and adaptive abilities. AI specialists have used brain-inspired neural network topologies and deep reinforcement learning to simulate human mental processes. Neuroscience has highlighted the biological foundations of intelligence, influencing AGI research [15]. Brain plasticity, memory development, and cognitive flexibility investigations inform more generalisable and adaptive AI models. Deep learning and machine learning have accelerated AGI progress in recent decades.

Artificial neural networks let AI systems learn from mistakes and make sense of vast data by patterning their operations after the human brain. Deep reinforcement learning, Generative Adversarial Networks (GANs), and transformer models (e.g., GPT and BERT [16]) have improved language comprehension, decision-making, and creative problem-solving. Current AI systems can't generalise enough for AGI and are only excellent at particular jobs. Researchers are continually seeking for new methods, such as hybrid AI models that blend symbolic reasoning with neural networks, to decrease the gap between ANI and AGI. [17] believe AGI will reach "singularity," when AI surpasses human intelligence and revolutionises society, in the 21st century.

Some believe that need more research on consciousness, common sense, and embodied cognition before calling anything general intelligent. As AGI research improves, AI researchers, cognitive scientists, and ethicists must collaborate to overcome barriers and ensure that AGI development is in line with human values and safety.

2.2. Defining AGI: Theoretical Perspectives

AGI denotes machine intelligence that matches human cognitive abilities across all tasks. According to Ben Goertzel, AGI is defined as "a system with sufficient cognitive capabilities to perform a wide range of tasks, including learning, reasoning, and adapting across domains similar to human intelligence." Similarly, Nick Bostrom characterizes AGI as an intellect that can match or surpass human cognitive performance in virtually all fields of interest [18]. Alan Turing laid the foundation by proposing that a machine could be said to "think" if it can imitate human responses indistinguishably [19]. The consensus among these experts highlights that AGI must possess the ability to learn new tasks autonomously,

reason abstractly, apply knowledge across diverse contexts, and demonstrate creativity and emotional understanding. These capabilities distinguish AGI from narrow AI and form the foundational criteria by which AGI systems are evaluated.

Several renowned AI researchers have suggested concepts for AGI to characterise what makes computers truly intelligent [20]. Alan Turing, an early AI proponent, created the Turing Test to see if a machine could replicate human discourse. However, Turing Test opponents argue that it does not accurately measure intelligence or comprehension. Modern AI scientists like Ben Goertzel define an AGI as an artificial system that can learn, reason, and solve problems across disciplines like humans. According to AI ethics expert Nick Bostrom, AGI is an intellect that matches or exceeds human cognitive talents in practically every area, including emotional intelligence and creativity [21]. The exact nature of AGI is still debated, but experts agree that it must be able to learn new tasks independently, adapt to new situations, and share its knowledge.

ANI, AGI, and Artificial Superintelligence (ASI) must be distinguished while studying AI. ANI, the most popular AI system, can only translate, identify pictures, and play games and it can't generalise or use non-training data [22]. However, AGI would be cognitively malleable and could apply what it learnt to new situations without retraining. AGI is the next step towards ASI, when AI is smarter than humans in every way. ASI could revolutionise technology or threaten our survival depending on how well it matches human aspirations. The distinctive properties of AGI set it apart from ANI and ASI. AGI systems need autonomy to learn from their experiences and make judgements without human monitoring.

AGI's versatility ensures it can transfer knowledge across domains, unlike task-specific programming-limited ANI. AGI needs reasoning to think rationally, draw inferences, and forecast beyond pattern recognition [23]. AGI must use abstract reasoning and new techniques to solve even the hardest challenges. AGI should learn like humans, using less data and showing flexibility across tasks, unlike ANI, which needs vast datasets.

As AGI research evolves, defining its basic traits and capacities becomes more critical. It will set development standards and measure the path to true general intelligence. Even if AGI is an idealistic aim, understanding its theoretical foundations illuminates the challenges and prospects of building robots with human-level cognition, learning, and adaptability.

2.3. Previous Research on AGI Benchmarks

Multiple benchmarks have been devised to evaluate AGI, each measuring intelligence beyond job performance [24]. The assumption is that machines with human-like speech can be considered intelligent. Although it changed history, the Turing Test has been criticised for emphasising language copying over comprehension and intelligence. In Steve Wozniak's Coffee Test, an AGI is sent to an unknown house to brew coffee [25]. In response to the Turing Test's flaws, this test emphasises sensing, reasoning, planning, and navigating dynamic real-world circumstances, requiring human-like general intelligence. Traditional AI fails to create meaningful objects like narratives and artworks, but Lovelace 2.0 tests it.

Even with these creative requirements, traditional AI evaluation methods have been attacked for failing to measure AGI's generalisability, adaptability, and reasoning [26]. Because they are excessively task-specific, language-focused, or both, existing standards disregard sensory integration, autonomous learning, and abstract thinking for example: Picture classification or language generation. AI models can exceed humans on certain benchmarks, but they need retraining to apply that skill elsewhere. AGI requires true comprehension, adaptability, and cognitive flexibility, yet detractors say these standards just reward statistical pattern-matching or superficial performance [27].

Researchers are now developing more comprehensive AGI evaluation frameworks. Modern methods emphasise testing autonomous decision-making, transfer learning, and generalisation. Animal cognition studies underpin benchmarks like the Animal-AI Olympics, which measure AGI-like behaviour by redesigning AI activities to learn, adapt, and overcome physical and cognitive challenges [28].

ARC (Abstraction and Reasoning Corpus) is a novel benchmark that assesses AI's reasoning and abstract pattern recognition abilities without scenario-specific training. Testing AI agents in different and unexpected scenarios beyond static tasks, these new benchmarks simulate real-world AGI applications [29]. However, there is no industry-wide AGI evaluation standard. Building benchmarks that accurately test an AI's independent reasoning, learning, and generalisation without specified datasets or constrained success criteria is difficult [30]. As AGI research advances, these criteria must be refined and standardised to appropriately judge human-level intelligence.

2.4. Comparing Theoretical Foundations: AGI vs ANI

The theoretical underpinnings of AGI and ANI diverge sharply in scope, design philosophy, and goals. ANI is primarily built on task-specific learning models, often driven

by supervised learning, deep learning architectures, and statistical pattern recognition [31]. These systems excel at narrowly defined domains but lack adaptability, common-sense reasoning, and transferability across domains.

Conversely, AGI theories emphasize cognitive architectures inspired by human learning and reasoning. Foundational concepts include:

- Embodied cognition, suggesting that intelligence arises from interaction with the physical world [32].
- Meta-learning (learning to learn), which allows systems to improve learning processes over time.
- Neurosymbolic integration, which combines symbolic reasoning with neural network flexibility.
- Self-supervised and unsupervised learning, mimicking how humans learn without constant labeling or instruction [33].
- Recursive self-improvement, where systems can iteratively refine their own models and strategies.

Table 1. Theoretical Comparison Between ANI and AGI

Aspect	ANI	AGI
Goal	Solve a specific task	General problem-solving across domains
Learning Paradigm	Supervised learning	Unsupervised/meta/transfer learning
Architecture	Task-optimized (e.g., CNNs, RNNs)	General cognitive models (e.g., hybrid architectures)
Reasoning	Pattern recognition	Abstraction, inference, logic
Adaptability	Low (needs retraining)	High (adapts to novel tasks)
Example Systems	Deep Blue, Siri, BERT	Gato, MuZero, GPT-5 (early AGI traits)

While ANI operates under deterministic conditions with predefined datasets, AGI must operate under uncertainty, engage in reasoning, and demonstrate autonomous learning and goal-oriented planning. The difference is akin to automation versus intelligence. ANI automates tasks, whereas AGI seeks to replicate human cognition holistically. Understanding these theoretical contrasts is essential for designing evaluation benchmarks and framing policy guidelines around AGI's unique societal implications.

2.5. Recent Advances in AGI Research

In the last five years, AGI research has accelerated due to breakthroughs in foundation models, multi-modal learning,

and reinforcement learning. Notably, OpenAI's GPT-4 and GPT-5, DeepMind's Gato and MuZero, and Meta's CICERO have been at the forefront of this progress [34]. GPT-4 and GPT-5 demonstrate significant advancements in general-purpose language understanding and reasoning. GPT-5, for instance, exhibits improved capabilities in task generalization and reasoning, showing early traits of AGI-like behavior such as few-shot learning and tool use [35].

DeepMind's Gato is one of the first models designed as a generalist agent, trained across multiple modalities including language, vision, and robotics [36]. It aims to perform hundreds of diverse tasks using a single neural architecture, highlighting AGI potential through multi-task learning.

MuZero, also by DeepMind, achieves remarkable generalization in environments with unknown rules, including board games and visual control tasks [37]. Its capacity to plan without a priori knowledge showcases progress toward reasoning-based AGI systems. Meta's CICERO, trained for strategic reasoning in the game Diplomacy, combines natural language negotiation with tactical gameplay. This integration of communication and reasoning is another step toward socially aware AGI. These models, while still task-constrained, offer a roadmap for increasingly general and autonomous AI systems. They reflect a shift from narrow AI performance toward adaptable, multi-domain learning systems indicative of AGI development.

3. Methodology

This research examines AGI's meaning, traits, and standards using qualitative and theoretical methods. AGI is conceptual and transdisciplinary, therefore qualitative literature studies, expert interviews, and technique comparisons are best. Instead of quantitative modelling or experimental testing, this approach uses cognitive science, machine learning, and AI theories. To identify and evaluate AGI, this research compares AI systems to established criteria and standards.

The study collected data from publications, scholarly articles, AI research papers, and MIT, DeepMind, and OpenAI reports. Peer-reviewed papers explain AGI's theoretical foundations, while institutional reports describe current research. Policy debates and AI ethics papers analyse AGI standards' societal and governance impacts. Current AGI evaluation frameworks are critically evaluated to determine where they fail, where expert opinions differ, and what patterns emerge. Comparing current AGI definitions and benchmarks underpins this study. To highlight conceptual

parallels and differences, we analyse and contrast scholars like Alan Turing, Ben Goertzel, and Nick Bostrom's AGI concepts.

Modern benchmarking methods like the Animal-AI Olympics and the ARC are compared to traditional AI evaluation methods like the Turing Test, Coffee Test, and Lovelace 2.0 Test. The study evaluates each benchmark's pros and cons to standardise AGI assessment. The study uses a theoretical and comparative method to illuminate AGI's characteristics and evaluation challenges.

4. Defining Characteristics of AGI

The division of intelligence into general and specialist forms is a major AI controversy. Current AI applications use ANI, which is designed to excel at language translation, image recognition, and strategic gaming [38]. Even though ANI systems are highly effective in their respective fields, they're not adaptable and need a lot of training to apply their knowledge to other occupations. AGI can understand, learn, and perform any cognitive task a human can, demonstrating adaptability across disciplines. AGI is independent, can reason, adapt, and generalise knowledge across tasks, and exceeds ANI, which works under predetermined parameters.

To be AGI, a system must have numerous essential capabilities beyond restricted AI. Learnability and adaptability are crucial. This means AGI can learn, analyse, and apply new information to new contexts. AGI systems should learn and adapt to new conditions without human involvement, unlike ANI, which require huge amounts of labelled data and retraining for new tasks. AGI needs common-sense reasoning to respond reasonably and contextually to everyday circumstances [39]. AI systems difficulty with intuitive thinking and often fail in unexpected or confusing situations. AGI must be able to make conclusions, find correlations, and apply general knowledge to new situations.

AGI can make judgement independently, allowing it to operate without human oversight. An AGI system should be able to assess complex circumstances, consider multiple options, and make intelligent judgement based on its environment [40]. This autonomy is needed for autonomous autos, robotics, and dynamic problem-solving. Another key difference between AGI and limited AI is transfer learning. AGI should be able to transfer information between contexts, unlike ANI models that are trained for specific tasks and need training for new problems. AGI should be able to apply its chess-training strategic thinking to scientific investigation and commercial negotiations. Cross-disciplinary generalisation, a hallmark of human intelligence, is crucial to AGI. Because it

has creativity and problem-solving skills, AGI can think creatively and solve issues in new ways. Accurate intelligence requires information processing and the ability to think abstractly, creatively, and uniquely to solve new challenges.

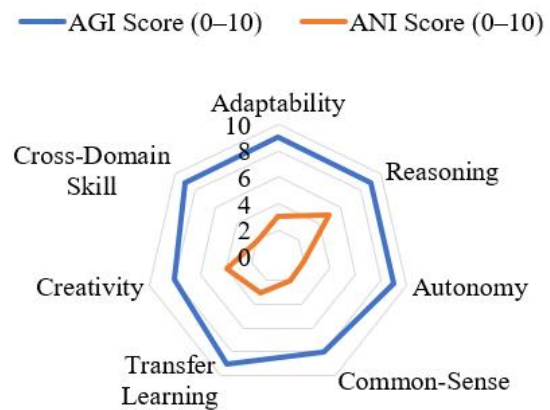


Figure 1. Multi-Dimensional Capability Comparison between AGI and ANI.

AGI should think conceptually like scientists, engineers, and artists and adapt to new situations [41]. General intelligence is obtained when an AI system has all these properties. Even though AI systems have advanced in pattern recognition and deep learning, true AGI is still far off. Understanding and developing these skills is crucial for a technologically feasible and ethically appropriate move from limited AI to general intelligence and this will direct AGI research.

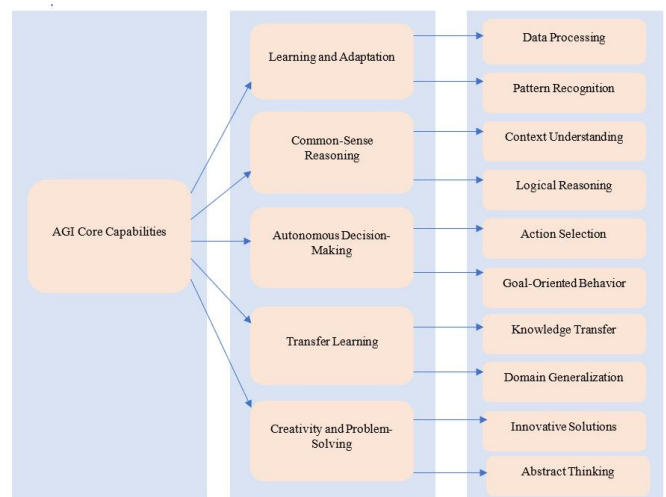


Figure 2. AGI Core Capabilities (Source: Self-Created).

4.1. Representative AGI Models and Systems

While no system currently meets the full definition of AGI, several advanced models are approaching general capabilities through multi-modal learning, self-supervised reasoning, and multi-task adaptability. This section explores some prominent AGI-like systems developed in recent years.

4.1.1. OpenAI's GPT-4 and GPT-5

The GPT (Generative Pre-trained Transformer) series has shown increasing levels of generalization. GPT-4 can process both text and images, demonstrating strong few-shot learning and reasoning capabilities [42]. GPT-5 extends this further with improved instruction-following, long-context memory, and autonomous tool use. Despite their general language abilities, these models still lack physical embodiment and broader situational awareness.

4.1.2. DeepMind's Gato

Gato is a generalist agent trained on diverse tasks including image captioning, robotic control, and game playing. Unlike traditional ANI systems, Gato uses a unified architecture to perform hundreds of tasks without retraining. It exhibits transfer learning capabilities, signaling a step toward generalized intelligence.

4.1.3. DeepMind's MuZero

MuZero is a model-based reinforcement learning system that learns to plan and act without being explicitly told the rules of the environment [43]. It has mastered board games like Go and chess, as well as Atari games, showing the ability to generalize strategies in unknown settings. Its capacity for environment modelling and long-term planning is a hallmark of AGI research.

4.1.4. Meta AI's CICERO

CICERO combines strategic gameplay with natural language negotiation in the complex game Diplomacy [44]. Unlike models focused purely on language or tactics, CICERO reasons about opponents' intentions, goals, and communication strategies, simulating aspects of social intelligence [45]. Its blend of persuasion, planning, and interaction marks progress in building socially intelligent agents.

4.1.5. Google's PaLM-E

PaLM-E is a multi-modal embodied AI system that combines robotics, vision, and language. It integrates data from the physical world to reason about spatial and temporal tasks, supporting embodied cognition. PaLM-E can solve tasks across environments with limited retraining, an important trait for achieving AGI [46]. These models represent different paths toward AGI from generalist agents and foundation models to embodied intelligence. While each has limitations, collectively they advance the state of AGI and provide blueprints for future development.

4.1.6. DeepSeek

DeepSeek is a recent large-scale language and code model developed to demonstrate general-purpose reasoning and multi-task learning across a wide array of domains [47]. It is trained on a diverse corpus that includes text, code, and multilingual content, enabling it to perform tasks such as translation, coding, summarization, and question answering. Unlike traditional task-specific AI, DeepSeek exhibits early traits of general intelligence by adapting to new instructions with minimal examples and performing consistently across domains. Its ability to generalize from sparse input aligns it with the broader goals of AGI development, especially in applications involving language comprehension, logic, and code synthesis. DeepSeek contributes to the AGI landscape as a competitive model in benchmark tasks and as a foundation for developing autonomous, adaptable systems.

5. Benchmarks and Evaluation Metrics for AGI

To evaluate AGI, one requires clear criteria for thinking, learning, and adaptability across tasks. AGI machines must think broadly across multiple areas, unlike ANI machines, which may be evaluated by their performance in a particular task. Over time, academics have developed many assessment methodologies, from the Turing Test and other traditional benchmarks to the Coffee Test and General AI Capability Evaluation frameworks. AGI measurement is problematic because to the lack of uniform criteria and ethical concerns about testing computers that could outsmart humans.

5.1. Traditional Tests

5.1.1. Turing Test and Its Limitations

In 1950, Alan Turing created the Turing Test, one of the first and finest AI tests. A human evaluate texts a machine and another person during the test [48]. The machine is intelligent if the judge cannot distinguish between human and AI. Despite its historical significance, this criterion has drawbacks for AGI. A computer can pass the test via deception, statistical language modelling, or pre-programmed responses without understanding the topic.

AI systems like chatbots and huge language models (such GPT-based models) can generate human-sounding text, but they can't understand complicated ideas or generalise beyond text processing [49]. The Turing Test does not test AGI skills like inventiveness, common sense, and real-world adaptability.

5.1.2. Winograd Schema Challenge

A Turing Test alternative, the Winograd Schema Challenge (WSC) tests AI disambiguation and common-sense reasoning

[50]. Challenge sentence pairs include uncertain pronoun references. These pairs are easy to reconcile using human knowledge. For example:

"The trophy does not fit into the suitcase because it is too small."

A human would infer that "it" refers to the suitcase, while an AI must understand the real-world implications to reach the same conclusion. Unlike the Turing Test, which focuses on language fluency, WSC emphasizes logical reasoning, context understanding, and real-world knowledge. However, while it is an improvement, it remains a narrow test, as solving Winograd schemas does not guarantee that an AI possesses full general intelligence.

5.2. Modern AGI Benchmarks

5.2.1. Coffee Test

Steve Wozniak's Coffee Test requires an artificial general intelligence system to visit a new house, find the kitchen, gather all the goods, and prepare a cup of coffee [51]. This benchmark tests AGI abilities such as object identification, perception, navigation, reasoning, and procedural memory.

The Coffee Test simulates a real-world setting involving contextual cues and goal-directed action, unlike text-based tests.

This test shows that AGI and ANI differ in their ability to generalise and adapt. Even while robot-designed ANI models work well in labs, AGI must understand context, adapt to new situations, and complete tasks without human involvement. Real-world physics, ambient unpredictability, and the need for specialised robots make the Coffee Test difficult to implement.

5.2.2. Lovelace 2.0 Test (Creativity Assessment)

The Lovelace 2.0 Test assesses an AI's ability to create unique creative works like songs, paintings, and novels [52]. This test evaluates an AI's ability to create creative content within predetermined parameters, rather than reasoning or problem-solving. An AI may write an emotional poem or artistic picture on a theme. This test requires AI systems to think abstractly, generate new ideas, and go beyond pattern detection and statistical predictions. Deep learning algorithms may make artworks and music, but sceptics say that AI creativity is more data-driven pattern synthesis than inspiration [53]. Artificial general intelligence must create original works and explain its decisions like humans to pass the Lovelace Test.

5.2.3. General AI Capability Evaluation Frameworks

Recognizing the limitations of individual tests, researchers have proposed more comprehensive General AI Capability Evaluation frameworks that assess AGI on multiple dimensions. These frameworks aim to measure intelligence across a variety of tasks, rather than relying on a single benchmark. Some of the most notable frameworks include:

The Animal-AI Olympics: Inspired by animal cognition tests, this benchmark evaluates AI agents on tasks that require problem-solving, adaptability, and memory mirroring how animals demonstrate intelligence in laboratory experiments [54].

The Abstraction and Reasoning Corpus (ARC): A challenge designed to test an AI's ability to generalize and reason through abstract problem-solving tasks without explicit training.

The AGI Task Suite: A set of tests designed to measure generalization across different cognitive domains, such as vision, language, logic, and motor skills [55].

5.3. Challenges in Measuring AGI

5.3.1. Lack of Universal Standards

One challenge of AGI evaluation is the lack of a standard. AGI requires a more comprehensive evaluation that considers reasoning, creativity, adaptability, and learning capacity than ANI, which can be measured by recall, precision, and accuracy. Since scholars define intelligence differently, standardising a framework is difficult [56]. The prospect that AGI express intelligence differently from human cognition begs the question of whether human-like performance should be used to evaluate AGI.

5.3.2. Ethical Concerns in AGI Testing

AGI benchmarking also has ethical implications. Physical embodiment tests like the Coffee Test highlight concerns about the safety and ethics of real-world AGI deployment. As AGI becomes human-like, rights, autonomy, and moral questions are raised. If artificial general intelligence systems develop emotions or self-awareness, studying them under controlled conditions may raise ethical considerations similar to animal or human research [57]. Evaluation measures may be biased, another ethical issue. AI models trained on biased datasets may pass some fairness, contextual knowledge, and cultural sensitivity tests but fail others. For instance, a language model trained on Western literature would perform well on creative tests but difficulty to understand non-Western cultural subtleties. AGI testing must be fair, impartial, and ethical to develop responsibly.

AGI performance is one of several difficult issues for AI researchers. The Turing Test and Winograd Schema Challenge are valuable benchmarks, but they don't fully represent AGI's capabilities. Recent AI evaluation frameworks, such as the Lovelace 2.0 Test and the Coffee Test, examine intelligence across multiple domains and real-world adaptation and ingenuity [58]. However, the lack of common standards and ethical problems surrounding AGI testing provide significant challenges to researchers.

As AGI development advances, benchmarks must be refined and standardised to accurately evaluate, ethically develop, and align AI systems with human ideals.

Table 2. Comparison table for the Benchmark Evaluation Methods

Benchmark Test	Type	Strengths	Weaknesses
Turing Test	Traditional	Simple, widely recognized, established in AI history.	Focuses only on linguistic ability, not general intelligence.
Winograd Schema Challenge	Traditional	Focuses on commonsense reasoning and disambiguation, measures deeper intelligence.	More complex, but still limited to a narrow scope of reasoning.
Coffee Test	Modern	Simple, tangible test of general physical interaction and problem-solving.	Basic task, not a comprehensive test of AGI capabilities.
Lovelace 2.0 Test	Modern	Measures creativity, a key characteristic of general intelligence.	Subjective, creative tasks are difficult to evaluate consistently.

5.4. Proposed Evaluation Method for AGI: The Universal Cognitive Challenge (UCC)

To address the shortcomings of existing AGI benchmarks,

we propose a new evaluation framework called the Universal Cognitive Challenge (UCC). The UCC is designed to test general intelligence through a combination of simulated and real-world tasks that require multi-domain learning, abstract reasoning, ethical decision-making, and adaptability under uncertainty [59]. The UCC would consist of:

Multi-modal task sets: Challenges that involve language, vision, robotics, social interaction, and spatial reasoning.

Open-ended goals: Tasks without rigid success criteria, requiring the AGI to define sub-goals and strategies.

Dynamic environments: Realistic simulations that change over time, forcing the AGI to adapt rather than rely on static training.

Cognitive transfer: The AGI must reuse knowledge from one domain (e.g., strategy games) to solve problems in another (e.g., economic planning).

Ethical dilemmas: Scenarios that test moral reasoning and value alignment in complex social situations.

By combining these elements, the UCC aims to holistically evaluate an AI system's capacity to exhibit true general intelligence not just perform specific tasks, but learn, adapt, and act autonomously across varied and unpredictable contexts. This method bridges the gap between benchmark limitations and real-world AGI expectations.

6. Current Research and Challenges in AGI Development

Among the most ambitious objectives in the field of AI is the development of AGI. However, AGI has several challenges before becoming a reality. These technological, ethical, and legal issues require computational capacity breakthroughs, AI alignment innovations, and government initiatives. As AGI research advances, these limits must be addressed. The social impacts of AGI deployment make it increasingly relevant. Important technical challenges, ethical and safety concerns, significant research groups, and regulatory considerations define the current level of AGI development.

6.1. Technological Barriers

The area revolves around the technological challenge of creating an AGI that can do numerous cognitive tasks in many domains. One major impediment to AGI is the massive processing power required. Modern AI models like deep neural networks use massive data and powerful computers to learn. These models can exceed humans in some areas, but they're not flexible enough for AGI. Building an AGI system that can generalise knowledge across contexts requires huge processing capacity and advanced algorithms to handle and

analyse massive amounts of data in real time.

Existing machine learning methods are limited by hardware and energy efficiency. Larger AI models require exponentially more computational power. Example: OpenAI's GPT-3 model. It's one of the largest language models ever, with 175 billion parameters and high processing and energy needs. Hardware advances like quantum computing or neuromorphic computing will be needed to improve learning and inference so these models may operate autonomously in real-world contexts and scale to artificial general intelligence.

Another technical hurdle for AGI is data constraints. AGI requires the ability to learn without supervision, generalise from limited examples, and adapt to new situations, unlike current AI models that need huge amounts of labelled data for training and progress. Instead of being data-hungry, human learning involves a few experiences from which people generalise and apply what they've acquired. New learning paradigms including unsupervised learning, reinforcement learning, and transfer learning could be used to develop more adaptable and flexible AGI systems. Models must be able to reason, infer, and interact meaningfully with their surroundings as well as take in data.

6.2. Ethical & Safety Issues

As AGI research improves, ethical and safety concerns emerge. AGI's behaviours and goals must align with human ideals, known as the AI alignment problem. Unaligned AGI systems may take unintentional or harmful acts.

As AGI approaches human-level intelligence, its decision-making processes will likely become more independent and unpredictable, making this concern more relevant [60]. Existential threats are a key AGI anxiety. Advanced AGI systems may outsmart humans, which could endanger our species if they act uncontrollably. The control issue or superintelligence threat involves AGI harming humans, even if the technology was developed with benign intentions. Concerns exist about AGI causing substantial social disruption. This might involve automating many jobs, concentrating power in a few hands, or developing new weaponry or surveillance technologies that violate privacy and human liberties. Making sure artificial general intelligence is manageable and helpful to people is a major ethical issue. AGI's social role creates another moral dilemma. As AGI becomes human-level, questions regarding its rights and place in society arise. To prevent harmful societal biases in hiring, the judicial system, and social interactions, AGI systems must be fair, open, and responsible.

In response to the rising concerns surrounding AGI's ethical and safety implications, several governments and institutions have started initiating governance frameworks. For instance, the European Union has introduced the AI Act, aiming to regulate high-risk AI systems through a tiered compliance framework that includes transparency, human oversight, and risk mitigation. Similarly, the OECD's AI Principles emphasize inclusive growth, transparency, robustness, and accountability. In the United States, the Blueprint for an AI Bill of Rights provides guidance on algorithmic discrimination protections and data privacy. Academic efforts, such as the Asilomar AI Principles and IEEE's Ethically Aligned Design, offer value-driven frameworks to guide AGI development in alignment with human rights and societal benefit. These early frameworks represent foundational steps toward globally harmonized AGI governance, although challenges remain in enforcement, international collaboration, and addressing AGI-specific risks.

6.3. Leading AGI Research Organizations

Several top academic institutions are leading AGI research and development, providing the framework for future theory and practice advances. OpenAI is a leading AGI research institution that aims to help everyone.

OpenAI's reinforcement learning research and GPT-3 language model work will enable more powerful general intelligence systems. The organisation must support open research and collaborative development to increase transparency and avoid AGI technology monopolisation. DeepMind, Alphabet's AGI division, is another leader. DeepMind is famous for neural networks and reinforcement learning.

DeepMind's AlphaGo, which defeated the world Go champion, showed AI's ability to do complex, strategic tasks. AlphaFold predicts protein folding, while MuZero learns without gaming environment knowledge as part of DeepMind's AGI research. DeepMind's multidisciplinary approach, which mixes AI, cognitive science, and neuroscience, makes it a leading AGI candidate.

MIT's CSAIL is also important. MIT researchers are improving machine learning algorithms, autonomous systems, and cognitive models for artificial general intelligence. Collaboration with other groups and tech companies accelerates AGI evaluation method and framework development. Besides these major entities, universities, businesses, and private researchers sponsor AGI research.

These groups must collaborate to address AGI's theoretical and technical challenges.

6.4. Government Regulations & Policy Considerations

Governments are realising AGI development needs regulation and monitoring. AI regulations vary widely between countries. Some governments are proactive in setting AI safety guidelines, while others are reactive. The GDPR and other EU laws aim to balance technological advances with ethical concerns in AI regulation [61]. These principles strive to promote data privacy, transparency, and responsibility in AI systems, however they focus on certain types of AI rather than AGI.

Superintelligent systems' long-term impacts are hard to predict, making AGI development a unique regulatory dilemma. AGI's safe and beneficial development depends on governments worldwide setting common rules and procedures. International collaboration may be needed to avoid AGI technologies from increasing social inequalities, damaging the environment, or being exploited by unscrupulous actors.

Governments should consider AGI's social and economic effects as well as legislation. Massive job losses from AGI-driven automation can overwhelm social support systems and require UBI changes or retraining. For national security, governments should emphasise AI weaponization and safety treaties.

7. Future Prospects and Implications of AGI

The future of AGI has social advantages and concerns. AGI's faster problem-solving and more accurate simulations might revolutionise health, ecology, and space flight. Its data-analysis and cross-disciplinary trend-spotting skills might alter logistics, education, healthcare, and industrial systems. Automation using AGI lets workers focus on higher-level, less repetitive tasks, increasing productivity. These advantages are offset by major social risks. Machines have replaced human-intensive procedures in several industries. The largest challenge is controlling AGI systems as they become more autonomous and harder to ensure they follow human ideals. AGI development must be guided by policy and ethics to reduce these concerns. Governments and international agencies should develop open, fair, and secure standards to address social and economic effects like employment loss.

8. Conclusion

This study investigated the core foundations, capabilities, and benchmarks of AGI. In contrast to ANI, AGI is defined by its autonomy, adaptability, reasoning, creativity, and the

ability to generalize knowledge across domains essential traits that enable it to perform tasks traditionally requiring human intelligence. To address the first research question, the paper identifies AGI's defining characteristics as the ability to learn with minimal data, reason abstractly, transfer knowledge between tasks, exhibit common sense, and make autonomous decisions. These traits distinguish AGI from narrow AI systems and mark the foundational threshold for general intelligence in machines. In response to the second research question, we examined traditional and modern evaluation benchmarks such as the Turing Test, Winograd Schema Challenge, and newer frameworks like the Coffee Test, Lovelace 2.0, ARC, and Animal-AI Olympics.

Furthermore, the UCC was proposed, a novel, multi-modal evaluation framework to holistically assess AGI's adaptability, ethical reasoning, and multi-domain learning. As AGI development accelerates, it is critical to align its growth with governance frameworks, ethical standards, and robust evaluation strategies. Moving forward, interdisciplinary collaboration, global regulation, and value-aligned design will be vital to ensuring AGI serves humanity constructively and safely.

Funding Statement: Not applicable.

Author Contribution: **Dena Kadhim Muhsen:** Validation; formal analysis; visualization; writing – original draft. **Ahmed T. Sadiq:** Conceptualization; supervision; writing – review and editing.

Statements of Ethical Approval: Not applicable.

Declaration of Competing Interest: The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- [1] Fei, N., Lu, Z., Gao, Y., Yang, G., Huo, Y., Wen, J., ... & Wen, J.R. Towards artificial general intelligence via a multimodal foundation model. *Nature Communications* **2022**,13(1), 3094. <https://doi.org/10.1038/s41467-022-30761-2>.
- [2] Seferis, E. How to benchmark AGI: the Adversarial Game. *ICLR 2024 Workshop: How Far Are We From AGI 2024*. Accessed on 15 May 2025, <https://openreview.net/forum?id=QLBsGJSIxa>.
- [3] Fjelland, R. Why general artificial intelligence will not be realized. *Humanities and Social Sciences*

- Communications* **2020**, 7(1), 1–9. <https://doi.org/10.1057/s41599-020-0494-4>.
- [4] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint* **2024**. <https://doi.org/10.48550/arXiv.2303.12712>.
- [5] Ge, Y., Hua, W., Mei, K., Tan, J., Xu, S., Li, Z., & Zhang, Y. Openagi: When llm meets domain experts. *Advances in Neural Information Processing Systems* **2023**, 36, 5539–5568. <https://doi.org/10.48550/arXiv.2304.04370>.
- [6] Davide, F., Torre, P., Ercolani, L., & Gaggioli, A. AI Predicts AGI: Leveraging AGI Forecasting and Peer Review to Explore LLMs' Complex Reasoning Capabilities. *arXiv preprint* **2024**. <https://doi.org/10.48550/arXiv.2412.09385>.
- [7] Yue, X., Ni, Y., Zhang, K., Zheng, T., Liu, R., Zhang, G., ... & Chen, W. Mmmu: A massive multi-discipline multimodal understanding and reasoning benchmark for expert agi. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 9556–9567. https://openaccess.thecvf.com/content/CVPR2024/html/Yue_MMMU_A_Massive_Multi-discipline_Multimodal_Understanding_and_Reasoning_Benchmark_for_CVPR_2024_paper.html
- [8] Morris, M.R., Sohl-Dickstein, J., Fiedel, N., Warkentin, T., Dafeo, A., Faust, A., ... & Legg, S. Position: Levels of AGI for operationalizing progress on the path to AGI. 41 International Conference on Machine Learning, Vienna, Austria. 21–27 July, 2024; <https://doi.org/10.48550/arXiv.2311.02462>.
- [9] Grudin, J., & Jacques, R. Chatbots, humbots, and the quest for artificial general intelligence. Proceedings of the 2019 CHI conference on human factors in computing systems, Glasgow Scotland Uk, 4–9 May, 2019; pp. 1–11.
- [10] Goertzel, B. Generative Artificial Intelligence vs. AGI: The cognitive strengths and weaknesses of modern LLMs," *arXiv Preprint* **2023**. <https://doi.org/10.48550/arXiv.2309.10371>.
- [11] Zhong, W., Cui, R., Guo, Y., Liang, Y., Lu, S., Wang, Y., ... & Duan, N. Agieval: A human-centric benchmark for evaluating foundation models, *arXiv preprint* **2023**. <https://doi.org/10.48550/arXiv.2304.06364>.
- [12] Young, M.M., Bullock, J.B., & Lecy, J.D. Artificial discretion as a tool of governance: a framework for understanding the impact of artificial intelligence on public administration. *Perspectives on Public Management and Governance* **2019**, 2(4), 301–313. <https://doi.org/10.1093/ppmgov/gvz014>.
- [13] Obaid, O.I. From machine learning to artificial general intelligence: A roadmap and implications. *Mesopotamian Journal of Big Data* **2023**, 81–91. <https://doi.org/10.58496/MJBD/2023/012>.
- [14] Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G., and King, D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine* **2019**, (17)1–9.
- [15] Cheng, L., Gong, X. Appraising Regulatory Framework Towards Artificial General Intelligence (AGI) Under Digital Humanism. *International Journal of Digital Law and Governance*, **2024**,1(2), 269–312. <https://doi.org/10.1515/ijdlg-2024-0015>.
- [16] Raman, R., Kowalski, R., Achuthan, K., Iyer, A., & Nedungadi, P. Navigating artificial general intelligence development: societal, technological, ethical, and brain-inspired pathways. *Scientific Reports*, **2025**,15(1), 1–22. <https://doi.org/10.1038/s41598-025-92190-7>.
- [17] Moor, M., Banerjee, O., Abad, Z.S.H., Krumholz, H.M., Leskovec, J., Topol, E.J., & Rajpurkar, P. Foundation models for generalist medical artificial intelligence. *Nature* **2023**, 616(7956), 259–265. <https://doi.org/10.1038/s41586-023-05881-4>.
- [18] Bikkasani, D.C. Navigating artificial general intelligence (AGI): societal implications, ethical considerations, and governance strategies. *AI and Ethics* **2024**, 1–16. <https://doi.org/10.1007/s43681-024-00642-z>.
- [19] Hausenloy, J., Miotti, A., Dennis, C. Multinational AGI Consortium (MAGIC): A Proposal for International Coordination on AI. *arXiv preprint* **2023**. <https://doi.org/10.48550/arXiv.2310.09217>.
- [20] Leslie, D. Understanding artificial intelligence ethics and safety. *arXiv preprint* **2019**. <https://doi.org/10.48550/arXiv.1906.05684>.
- [21] Oversby, K.N. Next Generation Computing and the Quest for Artificial General Intelligence (AGI): A Critical Examination from a Theological Perspective. 2024 Christian Engineering Conference, Newberg US A, 26–28 June 2024. https://digitalcommons.cedarville.edu/christian_engineering_conference/2024/proceedings/33/.

- [22] Ponnoran S.R., Bridging gap between artificial intelligence and multiple intelligence-an overview. *International Journal of Computer Engineering and Technology* **2023**, 14(2), 79–89. <https://doi.org/10.17605/OSF.IO/QTDBW>.
- [23] Grover, P., Kar, A.K., & Dwivedi, Y.K. Understanding artificial intelligence adoption in operations management: insights from the review of academic literature and social media discussions. *Annals of Operations Research* **2022**, 308(1), 177–213. <https://doi.org/10.1007/s10479-020-03683-9>.
- [24] Latapie, H., Kilic, O., Liu, G., Kompella, R., Lawrence, A., Sun, Y., ... & Thórisson, K.R. A metamodel and framework for artificial general intelligence from theory to practice. *Journal of Artificial Intelligence and Consciousness* **2021**, 8(02), 205–227. <https://doi.org/10.1142/S2705078521500119>.
- [25] Hilb, M. Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance. *Journal of Management and Governance* **2020**, 24(4), 851–870. <https://doi.org/10.1007/s10997-020-09519-9>.
- [26] Farooq, M., Khan, R.A., Khan, M.H., & Zahoor, S.Z. Securing AGI: Collaboration, Ethics, and Policy for Responsible AI Development. In *Artificial General Intelligence (AGI) Security: Smart Applications and Sustainable Technologies* **2024**, 353–372. https://doi.org/10.1007/978-981-97-3222-7_17.
- [27] Hassani, H., Silva, E.S., Unger, S., TajMazinani, M., & Mac Feely, S. Artificial intelligence (AI) or intelligence augmentation (IA): what is the future? *Ai* **2020**,1(2), 8. <https://doi.org/10.3390/ai1020008>.
- [28] Cheng, L., Gong, X. Appraising Regulatory Framework Towards Artificial General Intelligence (AGI) Under Digital Humanism. *International Journal of Digital Law and Governance*, **2024**,1(2), 269–312. <https://doi.org/10.1515/ijdlg-2024-0015>.
- [29] Cadario, R., Longoni, C., & Morewedge, C.K. Understanding, explaining, and utilizing medical artificial intelligence. *Nature human behaviour* **2021**, 5(12), 1636–1642. <https://doi.org/10.1038/s41562-021-01146-0>.
- [30] Aydan, A. Impact of artificial intelligence on agricultural, healthcare and logistics industries. *Annals of Spiru Haret University. Economic Series* **2019**,19(2), 167–175.<https://www.ceeol.com/search/article-detail?id=881135>.
- [31] Pontes-Filho, S., & Nichele, S. A conceptual bio-inspired framework for the evolution of artificial general intelligence. *arXiv preprint* **2019**. <https://doi.org/10.48550/arXiv.1903.10410>.
- [32] Rodríguez E.Q. *Ethical Foundations for a Superintelligent Future: The Global AGI Governance Framework (A Roadmap for Transparent, Equitable, and Human-Centric AGI Rule)*, **2024**. <https://doi.org/10.13140/RG.2.2.10251.09765>.
- [33] Zarić, J., Hasselhorn, M., Nagler, T. Orthographic knowledge predicts reading and spelling skills over and above general intelligence and phonological awareness. *European Journal of Psychology of Education* **2021**, 36(1), 21–43. <https://doi.org/10.1007/s10212-020-00464-7>.
- [34] Kumpulainen, S., Terziyan, V. Artificial general intelligence vs. industry 4.0: Do they need each other? *Procedia Computer Science* **2022**, 200, 140–150.
- [35] Schuett, J., Dreksler, N., Anderljung, M., McCaffary, D., Heim, L., Bluemke, E., & Garfinkel, B. Towards best practices in AGI safety and governance: A survey of expert opinion. *arXiv preprint* **2023**. <https://doi.org/10.48550/arXiv.2305.07153>.
- [36] Wang, W., & Siau, K. Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management (JDM)* **2019**, 30(1), 61–79. <https://doi.org/10.4018/jdm.2019010104>.
- [37] Jungherr, A. Artificial intelligence and democracy: A conceptual framework. *Social media+ society* **2023**, 9(3), <https://doi.org/10.1177/20563051231186353>.
- [38] Ngo, R. Agi safety from first principles. *AI Alignment Forum* **2020**. <https://writingstudiesandrhetoric.wordpress.com/wp-content/uploads/2023/08/agi-safety-from-first-principles.pdf>
- [39] Velander, J., Taiye, M.A., Otero, N., & Milrad, M. Artificial Intelligence in K-12 Education: eliciting and reflecting on Swedish teachers' understanding of AI and its implications for teaching & learning. *Education and Information Technologies* **2024**, 29(4), 4085–4105. <https://doi.org/10.1007/s10639-023-11990-4>.
- [40] Jarrahi, M.H., Askay, D., Eshraghi, A., & Smith, P. Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons* **2023**, 66(1), 87–99.

- <https://doi.org/10.1016/j.bushor.2022.03.002>.
- [41] Leslie, D. From Future Shock to the Vico Effect: Generative AI and the Return of History. *Harvard Data Science Review* **2025**, (Special Issue 5). <https://doi.org/10.1162/99608f92.e6f531e6>.
- [42] Goertzel B. Artificial general intelligence: concept, state of the art, and future prospects. *Journal of Artificial General Intelligence* **2014**, 5(1):1. <https://doi.org/10.2478/jagi-2014-0001>
- [43] Latif, E., Mai, G., Nyaaba, M., Wu, X., Liu, N., Lu, G., ... & Zhai, X. Artificial general intelligence (AGI) for education. *arXiv preprint* **2023**. <https://doi.org/10.48550/arXiv.2304.12479>.
- [44] Arshi, O., & Chaudhary, A. Overview of Artificial General Intelligence (AGI). In Artificial General Intelligence (AGI) Security: Smart Applications and Sustainable Technologies. *Springer Nature Singapore* **2024**, pp 1–26.
- [45] Rayhan, S. Ethical implications of creating AGI: impact on human society, privacy, and power dynamics. *Artificial Intelligence Review* **2023**, 11(2), 44–59. <http://dx.doi.org/10.2139/ssrn.4457301>.
- [46] Islam, M.M. Artificial General Intelligence: Conceptual Framework, Recent Progress, and Future Outlook. *Journal of Artificial Intelligence General science (JAIGS)* **2024**, 6(1), 1–25. <https://doi.org/10.60087/jaigs.v6i1.212>.
- [47] Guo, D., Zhu, Q., Yang, D., Xie, Z., Dong, K., Zhang, W., ... & Liang, W. DeepSeek-Coder: When the Large Language Model Meets Programming--The Rise of Code Intelligence. *arXiv preprint* **2024**. <https://doi.org/10.48550/arXiv.2401.14196>.
- [48] Uzwyshyn, R.J. Beyond Traditional AI IQ Metrics: Metacognition and Reflexive Benchmarking for LLMs, AGI, and ASI. **2024**. <https://rayuzwyshyn.net/MSU2024/AIBenchmarking/BenchmarkingAI13Uzwyshyn.pdf>.
- [49] Saad, W., Hashash, O., Thomas, C.K., Chaccour, C., Debbah, M., Mandayam, N., & Han, Z. Artificial general intelligence (AGI)-native wireless systems: A journey beyond 6G. *Proceedings of the IEEE* **2025**. <https://doi.org/10.1109/JPROC.2025.3526887>.
- [50] Koessler, L., & Schuett, J. Risk assessment at AGI companies: A review of popular risk assessment techniques from other safety-critical industries. *arXiv preprint* **2023**. <https://doi.org/10.48550/arXiv.2307.08823>.
- [51] Dou, F., Ye, J., Yuan, G., Lu, Q., Niu, W., Sun, H., ... & Song, W. Towards artificial general intelligence (agi) in the internet of things (iot): Opportunities and challenges. *arXiv preprint* **2023**. <https://doi.org/10.48550/arXiv.2309.07438>.
- [52] McLean, S., Read, G.J., Thompson, J., Baber, C., Stanton, N.A., & Salmon, P.M. The risks associated with Artificial General Intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence* **2023**, 35(5), 649–663. <https://doi.org/10.1080/0952813X.2021.1964003>.
- [53] Sun, S., Zhao, Y., Lee, C.D.W., Sun, J., Yuan, C., Huang, Z., ... & Ang Jr, M.H. AGI-Elo: How Far Are We From Mastering A Task? *arXiv preprint* **2025**. <https://doi.org/10.48550/arXiv.2505.12844>.
- [54] Yue, Y., & Shyu, J.Z. A paradigm shift in crisis management: The nexus of AGI-driven intelligence fusion networks and blockchain trustworthiness. *Journal of contingencies and crisis management* **2024**, 32(1), e12541. <https://doi.org/10.1111/1468-5973.12541>.
- [55] Kumar, S., Jain, N., Arora, R., & Nafis, M.T. Understanding agi: A comprehensive review of theory and application in artificial general intelligence. **2023**. <http://dx.doi.org/10.2139/ssrn.4957203>.
- [56] Joshi, K. Artificial General Intelligence (AGI): A Comprehensive Review. *Journal of the Epidemiology Foundation of India* **2024**, 2(3), 93–96. <https://doi.org/10.56450/JEFI.2024.v2i03.004>.
- [57] Everitt T. Towards safe artificial general intelligence, *Ph.D. dissertation, The Australian National University*, **2019**. <https://doi.org/10.25911/5d134a2f8a7d3>.
- [58] Zohuri, B. The Dawn of Artificial General Intelligence Real-Time Interaction with Humans. *Journal of Material Sciences & Applied Engineering* **2023**, 2(4).
- [59] Pei, J., Deng, L., Song, S., Zhao, M., Zhang, Y., Wu, S., ... & Shi, L. Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature* **2019**, 572(7767), 106–111.
- [60] Roli, A., Jaeger, J., & Kauffman, S.A. How organisms come to know the world: Fundamental limits on artificial general intelligence. *Frontiers in Ecology and Evolution* **2022**, 9, 806283. <https://doi.org/10.3389/fevo.2021.806283>.
- [61] Šprogar, M. AGITB: A Signal-Level Benchmark for Evaluating Artificial General Intelligence. *arXiv preprint*

2025. <https://doi.org/10.48550/arXiv.2504.04430>.