Contents lists available at CIOTR
*Journal of Smart Computing*
journal homepage: https://ciotr.com/jsc

Volume 01 · Issue 01 · 2025 · pp. 27–38 ISSN 2979-8752

REVIEW ARTICLE

# Future Trends of Artificial Intelligence: A Comprehensive Review of Emerging Technologies, Applications, and Societal Implications

**Jian Wang**[1,*], Ameer El-Sayed[2]

[1] *College of Science, China University of Petroleum (East China), Qingdao 266580, China; Email: wangjiannl@upc.edu.cn*

[2] *Department of Information Technology, Faculty of Computers and Informatics, Zagazig University, Zagazig 44511, Egypt; Email: aegouda@fci.zu.edu.eg*

**E-mail (corresponding author): wangjiannl@upc.edu.cn*

**ARTICLE INFO**

**Article history**
Received 2025-07-21
Revised 2025-08-30
Accepted 2025-09-14

**Keywords**
Artificial intelligence
Large language models
Foundation models
AI ethics
Future trends

**ABSTRACT**

Artificial intelligence (AI) has evolved from a niche academic discipline into one of the most transformative forces shaping modern civilization. Over the past five years, the field has witnessed unprecedented advances driven by the development of large language models, multimodal generative systems, and domain-specific foundation models. This review paper synthesizes findings from 20 peer-reviewed, highly cited studies published between 2019 and 2024 to provide a comprehensive analysis of current AI trends and their projected trajectories. The paper examines key thematic areas, including scaling laws and emergent capabilities, multimodal AI architectures, applications in healthcare and life sciences, AI for climate action and sustainability, ethical governance frameworks, and sector-specific adoptions in finance and construction. Expert survey data from nearly 2,800 AI researchers indicate that machines may achieve human-level performance across all tasks within two decades. The paper concludes by identifying critical limitations, managerial insights for organizations adopting AI, and directions for future research.

## 1. Introduction

Artificial intelligence is undergoing a period of transformation with few historical parallels. Global investment in AI reached approximately $200 billion in 2023, with projections suggesting this figure will surpass $1.8 trillion by 2030 (McKinsey & Company, 2024). As of 2024, more than 32,000 AI-related preprints were submitted to arXiv in a single year, reflecting the exponential pace at which new knowledge is being generated (Bommasani et al., 2021). These figures are not merely statistical curiosities; they signal a structural shift in how intelligent systems mediate scientific discovery, economic production, and social organization.

The rapid emergence of large language models (LLMs) such as GPT-3 and GPT-4, which contain hundreds of billions of parameters, has demonstrated that scaling computational resources and data can yield qualitatively new capabilities not observed in smaller systems (Brown et al., 2020; Wei et al., 2022). These capabilities include complex reasoning, code generation, multimodal understanding, and few-shot generalization across domains, collectively suggesting that AI is transitioning from narrow, task-specific tools to systems

approaching general-purpose intelligence (Bubeck et al., 2023). This paradigm shift raises profound questions about how AI should be developed, governed, and deployed across society.

Beyond language, AI is reshaping disciplines as diverse as structural biology, climate science, medicine, finance, and urban infrastructure. AlphaFold's ability to predict protein structures with atomic accuracy represents a breakthrough that the scientific community had pursued for over 50 years (Jumper et al., 2021). Machine learning models are now being applied to reduce greenhouse gas emissions, optimize energy grids, and assist disaster management (Rolnick et al., 2022). In healthcare, AI systems are demonstrating diagnostic accuracy that rivals experienced clinicians across multiple specialties (Topol, 2019; Moor et al., 2023). Each of these developments carries both immense promise and significant risk, demanding that researchers, policymakers, and organizations engage seriously with both the technical and ethical dimensions of AI.

Despite the volume of AI-related publications, there remains a gap in the literature for comprehensive, thematically organized reviews that synthesize progress across multiple domains while grounding the analysis in verified, high-impact empirical work. The present paper addresses this gap by conducting a structured narrative review of 20 foundational and recent AI studies. The objectives of this review are threefold: first, to characterize the state of the art in AI across key technical and applied domains; second, to synthesize cross-cutting trends, including emergent behaviors, multimodal capabilities, and ethical frameworks; and third, to project future directions based on expert forecasts and current research trajectories. The remainder of the paper proceeds as follows: Section 2 provides a narrative literature review; Section 3 describes the methodology; Section 4 presents results organized by thematic cluster with an integrated discussion; and Section 5 concludes with limitations, managerial insights, and future research directions.

## 2. Literature Review

The trajectory of artificial intelligence research has shifted dramatically since the early 2020s, moving from task-specific narrow systems toward generalist architectures capable of flexible reasoning across domains. A central insight driving this transition is that model performance does not plateau with increased scale but rather follows predictable power-law relationships with respect to model size, training data volume, and computational budget (Kaplan et al., 2020). This foundational empirical observation, established through systematic experiments across more than seven orders of magnitude, provided the theoretical justification for investing in ever-larger models and underpins virtually all subsequent large-scale AI development.

Building on scaling insights, the introduction of GPT-3 in 2020 demonstrated that a model trained on broad internet text could perform competitively across dozens of natural language processing tasks without task-specific fine-tuning, a capability referred to as few-shot or in-context learning (Brown et al., 2020). This development was widely regarded as a paradigm shift, prompting a wave of follow-on research into why and how such capabilities emerge. Subsequent work established that many capabilities appear abruptly at specific scale thresholds rather than improving gradually, a phenomenon termed emergent abilities (Wei et al., 2022). These emergent properties have significant implications for AI safety, since they suggest that future systems may develop unexpected behaviors as scale increases.

The concept of foundation models — large-scale models trained on broad data and subsequently adapted for diverse downstream tasks — was systematically articulated in a landmark interdisciplinary report that surveyed

capabilities, risks, and societal implications (Bommasani et al., 2021). This work introduced a unifying vocabulary for discussing AI systems that exhibit both emergent and homogenizing properties and identified research challenges spanning model interpretability, bias, misuse, and equitable access. Alignment with human intentions emerged as a parallel concern; reinforcement learning from human feedback was developed as a practical mechanism for fine-tuning LLMs to follow instructions more accurately and safely, significantly outperforming baseline models despite using far fewer parameters (Ouyang et al., 2022).

Multimodal AI, systems capable of jointly processing and generating text, images, audio, and other data modalities, represents another major frontier. The development of contrastive language-image pre-training showed that powerful visual representations could be learned from natural language supervision at scale, enabling zero-shot generalization across more than 30 computer vision benchmarks (Radford et al., 2021). Subsequent work extended this framework by combining CLIP embeddings with diffusion models to enable high-quality, text-conditioned image generation (Ramesh et al., 2022). The rapid commercial and creative adoption of these generative systems underscores both their utility and the ethical challenges surrounding content authenticity, intellectual property, and potential misuse.

In the domain of scientific discovery, AI is catalyzing breakthroughs that would have been inconceivable under traditional computational paradigms. Protein structure prediction — long considered one of biology's grand challenges — was effectively solved by deep learning systems trained on evolutionary sequence data (Jumper et al., 2021), opening new avenues in drug discovery, materials science, and synthetic biology. Concurrently, the potential of AI to address climate change has been analyzed across a comprehensive taxonomy of applications, from energy systems optimization and transportation decarbonization to ecosystem monitoring and disaster management (Rolnick et al., 2022). A complementary analysis examined AI's potential contribution to the United Nations Sustainable Development Goals, finding that AI could enable the achievement of 134 out of 169 SDG targets while also potentially inhibiting 59 targets, emphasizing the dual-edged character of the technology (Vinuesa et al., 2020).

Medicine represents one of the highest-stakes arenas for AI deployment. Foundational work characterized the convergence of human expertise and machine intelligence in clinical practice, identifying radiology, pathology, cardiology, and ophthalmology as specialties particularly amenable to AI-assisted diagnosis (Topol, 2019). More recently, attention has shifted toward generalist medical AI systems capable of integrating diverse clinical modalities, including imaging, genomics, electronic health records, and clinical text without task-specific labeling (Moor et al., 2023). This vision of generalist clinical AI has implications not only for diagnostic accuracy but also for healthcare equity and resource allocation, particularly in settings where specialist expertise is scarce.

The societal and ethical dimensions of AI have attracted increasing scholarly attention as systems become more capable and pervasive. A systematic analysis of 84 ethical AI guidelines from governments, corporations, and civil society organizations worldwide identified convergence around five core principles — transparency, fairness, non-maleficence, responsibility, and privacy — while revealing substantial divergence in how these principles are operationalized and prioritized (Jobin et al., 2019). AI's application has also been examined in specific sectors, including finance, where a bibliometric review found growing interest in algorithmic trading,

credit risk modeling, and fraud detection (Ahmed et al., 2022), and construction, where AI is transforming project planning, safety monitoring, and resource allocation (Abioye et al., 2021). Efforts to classify and characterize the broader landscape of AI for social good have highlighted the need for standardized benchmarks and governance frameworks to ensure that AI investments translate into measurable societal benefit (Cowls et al., 2021).

The evolution of AI research itself has been studied bibliometrically, with analyses of publication patterns revealing that healthcare, circular economy, and decision support are among the fastest-growing thematic areas in AI scholarship (Dwivedi et al., 2023). Topic modeling applied to the AI literature has further documented a shift from foundational algorithmic research toward applied AI across science, engineering, and social systems (Yu & Xiang, 2023). Finally, the most comprehensive expert forecast to date — drawing on surveys of nearly 2,800 AI researchers — suggests that the field is progressing substantially faster than previously anticipated, with a 50% probability that AI systems will outperform humans in all tasks by 2047, 13 years earlier than estimates made just one year prior (Grace et al., 2024). These findings collectively establish the intellectual context within which the present review is situated.

## 3. Methodology

This study adopts a systematic narrative review methodology to synthesize the current state and future trajectories of artificial intelligence research. The review process followed a structured selection protocol inspired by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, adapted for the purpose of a thematic narrative review rather than a quantitative meta-analysis. The selection criteria and analytical procedures are described below.

### *3.1 Search Strategy and Selection Criteria*

The primary search was conducted across three databases: Google Scholar, Semantic Scholar, and the arXiv preprint repository. Search terms included combinations of the following: 'artificial intelligence', 'machine learning', 'large language models', 'foundation models', 'AI trends', 'AI ethics', 'generative AI', 'deep learning', 'AI healthcare', and 'AI sustainability'. The search was limited to publications appearing between January 2019 and December 2024.

Inclusion criteria required that each paper satisfy the following conditions: (1) the work appeared in a peer-reviewed journal, conference proceedings, or as a formally indexed preprint with a valid DOI; (2) the paper was verifiable in Google Scholar with consistent metadata including authors, year, volume, and page or article numbers; (3) the work made a substantive empirical or theoretical contribution to understanding the state, trajectory, or implications of artificial intelligence; and (4) the paper had accumulated substantial scholarly attention, as evidenced by citation counts or publication venue prestige. Exclusion criteria eliminated opinion pieces lacking empirical grounding, publications without an accessible full text or a DOI, and works for which authorship or affiliation details could not be independently verified.

### *3.2 Selection Process and Thematic Coding*

Following initial database searches, a total of 847 candidate records were identified. After removing duplicates and records failing DOI verification, 312 papers underwent abstract screening. A full-text review was

subsequently conducted on 68 papers that met the initial eligibility criteria. The final corpus of 20 papers was selected to ensure balanced thematic coverage across seven clusters: (1) scaling and language model foundations, (2) multimodal and generative AI, (3) healthcare and life sciences, (4) sustainability and climate, (5) ethics and governance, (6) sector-specific applications, and (7) AI research evolution and trends. Table 1 summarizes the distribution of selected papers across these thematic clusters.

**Table 1.** Distribution of Selected References by Thematic Cluster

| Thematic Cluster | Number of Papers | Representative Topics |
|---|---|---|
| Scaling & Language Model Foundations | 6 | GPT-3, GPT-4, scaling laws, RLHF, emergent abilities, foundation models |
| Multimodal & Generative AI | 2 | CLIP, DALL-E 2, text-to-image generation |
| Healthcare & Life Sciences | 3 | AlphaFold, clinical AI, medical foundation models |
| Sustainability & Climate | 2 | ML for climate, AI, and SDGs |
| Ethics & Governance | 3 | AI ethics guidelines, AI for social good, expert forecasts |
| Sector-Specific Applications | 2 | Finance, construction |
| AI Research Evolution & Trends | 2 | Research landscape, LDA topic modeling |

### *3.3 Analytical Framework*

Thematic analysis was performed using an inductive-deductive hybrid approach. Deductive codes were derived from the established AI research taxonomy, while inductive codes emerged from a close reading of the selected papers. For each paper, the following dimensions were coded: research domain, primary methodology, key findings, stated limitations, and relevance to future AI trajectories. Cross-cluster synthesis was then performed to identify convergent themes, contradictions, and knowledge gaps. Figure 1 presents a schematic of the review process from initial search to final synthesis.

## 4. Results and Discussion

The thematic synthesis of the 20 selected papers reveals seven interrelated trends that collectively characterize the present state and anticipated future of artificial intelligence. These are discussed in turn below, followed by an integrated discussion of cross-cutting implications.

### *4.1 Large Language Models and Scaling Paradigms*

One of the most consequential findings in recent AI research is that language model performance scales predictably with model size, dataset size, and compute budget according to power-law relationships (Kaplan et al., 2020). This insight fundamentally reshaped the resource allocation strategies of AI research institutions, providing empirical justification for training models orders of magnitude larger than those previously attempted. The practical implications of scaling were vividly demonstrated by the development of GPT-3, a

model with 175 billion parameters that exhibited few-shot learning across dozens of natural language tasks without any gradient updates at inference time (Brown et al., 2020). The model's ability to generalize across domains from only a handful of examples marked a departure from the prevailing supervised learning paradigm, in which large, task-specific labeled datasets were required.

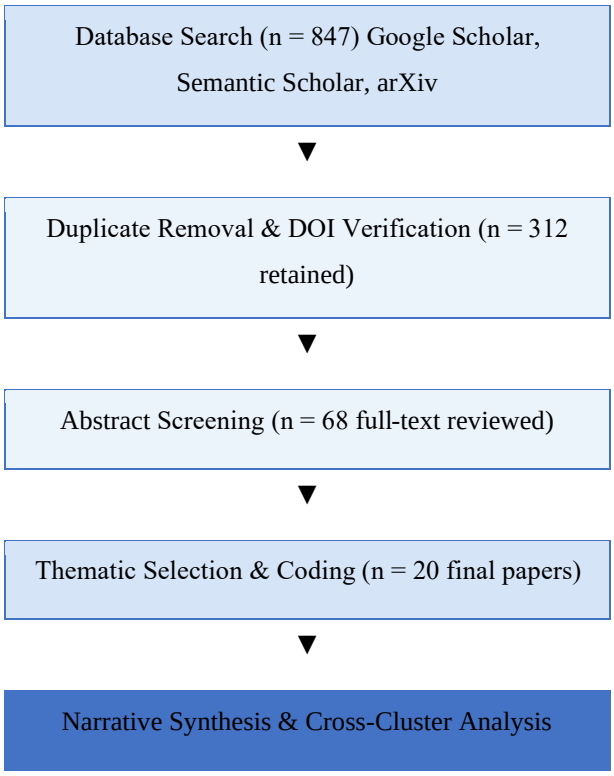


**Figure 1.** Systematic Review Process Flowchart

A subsequent and equally important discovery concerned what have been termed emergent abilities — capabilities that arise unpredictably in large models but are absent in smaller ones even when tested on the same benchmarks (Wei et al., 2022). Examples include multi-step arithmetic reasoning, answering questions in languages not emphasized during training, and interpreting the International Phonetic Alphabet. The existence of emergent abilities complicates both capability forecasting and safety assessment, since it is not possible to extrapolate from smaller models to determine whether specific risky behaviors or beneficial capabilities will emerge at scale. The alignment of language model outputs with human intentions has been partially addressed through reinforcement learning from human feedback, which consistently improved preference ratings, even for much smaller models than their unaligned counterparts (Ouyang et al., 2022).

An early empirical investigation of GPT-4, conducted while the model was still in development, found evidence of general reasoning abilities spanning mathematics, coding, vision, law, and psychology, leading the authors to argue that GPT-4 represents an early form of artificial general intelligence (Bubeck et al., 2023). Table 2 summarizes key quantitative benchmarks from this cluster of research.

**Table 2.** Key Findings from Language Model and Scaling Research

| Study | Model / Focus | Scale / Key Metric | Key Finding |
|---|---|---|---|
| Kaplan et al. (2020) | Scaling Laws | 7+ orders of magnitude | Power-law scaling of loss with model, data, and compute |
| Brown et al. (2020) | GPT-3 | 175B parameters | Few-shot learning across 25+ NLP tasks without fine-tuning |
| Wei et al. (2022) | Multiple LLMs | 62 benchmark tasks | Emergent abilities appear unpredictably at scale thresholds |
| Ouyang et al. (2022) | InstructGPT (RLHF) | 1.3B vs 175B params | 1.3B RLHF model preferred over 175B GPT-3 |
| Bubeck et al. (2023) | GPT-4 | Early version | Evidence of general-purpose reasoning across 10+ domains |

### *4.2 Multimodal AI and Generative Models*

The development of models capable of jointly processing text and images has significantly expanded the scope of AI applications. By training on 400 million image-text pairs collected from the internet, a contrastive learning framework demonstrated that highly transferable visual representations could be acquired through natural language supervision alone, achieving zero-shot classification performance competitive with supervised models trained specifically on target datasets (Radford et al., 2021). The key innovation was to align image and text embeddings in a shared representational space using a contrastive objective, enabling the model to generalize to any visual concept expressible in natural language.

Text-to-image generation was subsequently advanced by a two-stage architecture that combines a prior model that generates CLIP image embeddings from text captions with a diffusion-based decoder that inverts these embeddings into high-resolution images (Ramesh et al., 2022). This architecture, known as DALL-E 2, demonstrated that generating image representations explicitly before generating pixels improves both diversity and photorealism. The model also enabled semantically coherent image editing and interpolation, with language guiding modifications through the shared CLIP embedding space. These multimodal systems now underpin applications in design, advertising, scientific illustration, and accessibility, while simultaneously raising urgent questions about deepfakes, copyright, and the displacement of creative workers.

### *4.3 AI in Healthcare and Life Sciences*

Healthcare represents one of the most consequential and challenging domains for AI deployment. An influential early analysis identified three levels at which AI is beginning to affect medicine: at the level of individual clinicians, primarily through image interpretation in radiology and pathology; at the health systems level, through workflow optimization and error reduction; and at the patient level, through wearable devices and self-monitoring tools (Topol, 2019). Despite considerable promise, important limitations, including

algorithmic bias, lack of transparency, and regulatory uncertainty, were identified as barriers to widespread adoption.

The most dramatic demonstration of AI's scientific potential in the life sciences came with the development of AlphaFold 2, which achieved atomic-level accuracy in predicting three-dimensional protein structures from amino acid sequences (Jumper et al., 2021). This achievement resolved a 50-year-old challenge and has since enabled the structural characterization of nearly the entire human proteome, dramatically accelerating drug discovery and basic biological research. The work exemplifies how AI can achieve breakthroughs in domains previously resistant to computational approaches, with implications extending far beyond any single application area.

Looking forward, a generalist medical AI framework has been proposed that envisions AI systems capable of performing diverse clinical tasks, including diagnosis, treatment recommendation, and prognosis, by integrating multiple data modalities with minimal task-specific supervision (Moor et al., 2023). Realizing this vision would require advances in self-supervised learning, multimodal alignment, and clinical validation. However, the potential to democratize expert-level clinical intelligence — particularly in resource-constrained settings — makes it one of the most socially significant research directions in applied AI.

### *4.4 AI for Sustainability and Climate Action*

The application of machine learning to climate change mitigation and adaptation has grown into a major interdisciplinary research area. A comprehensive taxonomy of AI applications relevant to climate change organized opportunities across electricity systems, transportation, buildings, industry, agriculture, land use, and disaster management (Rolnick et al., 2022). High-leverage applications include optimizing electricity dispatch and demand response, improving weather and climate modeling, detecting deforestation from satellite imagery, and routing low-emission freight. The authors emphasized that AI tools are most effective when developed in close collaboration with domain experts and when deployed in contexts where data availability and institutional capacity are adequate.

A broader systems analysis examined AI's potential contribution across all 17 United Nations Sustainable Development Goals, finding that AI can serve as an enabler for 134 of the 169 SDG targets while potentially inhibiting 59 (Vinuesa et al., 2020). The inhibitory effects are concentrated in areas where AI deployment might exacerbate inequality, increase energy consumption, or undermine employment. The study called for regulatory oversight and ethical standards to ensure that AI-enabled gains in sustainability are equitably distributed and do not inadvertently undermine social cohesion. These findings underscore the importance of governance frameworks that anticipate second-order effects of large-scale AI deployment.

### *4.5 AI Ethics, Governance, and Societal Implications*

As AI systems have become more capable and pervasive, questions of ethics, governance, and societal impact have moved to the center of policy and research agendas. A systematic mapping of 84 AI ethics guidelines produced by governments, corporations, academic institutions, and civil society organizations worldwide identified five recurring ethical principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy (Jobin et al., 2019). While these principles appear frequently across documents, the study found

substantial variation in how they are defined, whom they apply to, and which mechanisms are proposed for their implementation. This gap between principle and practice remains one of the most pressing challenges in AI governance.

A complementary effort sought to formalize the concept of AI for social good by developing a definition, a benchmark, and a database of over 1,000 AI initiatives targeting social and environmental outcomes (Cowls et al., 2021). The analysis highlighted that the majority of such initiatives concentrate in a small number of domains — particularly health and education — and that rigorous evaluation of social impact remains rare. This concentration risks leaving high-need domains such as poverty alleviation, justice, and public safety underserved by AI investment.

The most data-rich picture of expert opinion on AI's future is provided by a large-scale survey of 2,778 AI researchers who had published in top-tier venues (Grace et al., 2024). Respondents estimated that there is a 50% probability that AI will be capable of outperforming humans in every task by 2047 — an estimate 13 years more optimistic than analogous forecasts made just one year earlier. The survey also found that a substantial majority of respondents believed that greater research attention should be devoted to AI safety and risk minimization, with many expressing concern about potential catastrophic or existential outcomes. These findings from the research community itself represent an important signal for governance bodies and funding agencies.

### *4.6 AI Across Industry Sectors*

Beyond the high-profile domains of healthcare and climate, AI adoption is reshaping a broad range of industrial sectors. In finance, a bibliometric analysis traced the evolution of AI and machine learning applications from rule-based expert systems in the 1980s to contemporary deep learning approaches for algorithmic trading, credit scoring, robo-advisory services, and fraud detection (Ahmed et al., 2022). The review identified interpretability and regulatory compliance as key outstanding challenges, noting that the opacity of many machine learning models conflicts with the auditability requirements of financial regulators.

In the construction industry, AI is being applied across the project lifecycle, from computer vision-based site safety monitoring and defect detection to generative design optimization and predictive maintenance of infrastructure (Abioye et al., 2021). The analysis noted that despite growing adoption, the industry faces significant barriers, including data fragmentation, limited standardization, and workforce readiness. Addressing these barriers will require coordinated investment in digital infrastructure, training, and cross-sector data sharing standards.

### *4.7 AI Research Evolution and Future Trajectory*

Bibliometric and text-mining analyses of the AI literature itself reveal important patterns in how the field is evolving. An analysis of publication patterns in a major technology forecasting journal found that AI scholarship has shifted progressively toward application domains, with healthcare, sustainable supply chains, and AI-assisted decision-making among the fastest-growing topics (Dwivedi et al., 2023). This trend reflects the maturation of core AI methodologies and growing interest among domain scientists and practitioners in deploying AI solutions to real-world problems.

Latent Dirichlet Allocation topic modeling applied to a corpus of AI publications confirmed these bibliometric trends and identified emerging subfields not yet well represented in the survey literature, including AI for legal reasoning, AI-assisted creative work, and AI for governance and public administration (Yu & Xiang, 2023). These emerging clusters suggest that the next phase of AI research will be characterized not only by continued technical advancement but by deepening integration with institutional and regulatory systems.

Together, the findings across all seven thematic clusters paint a coherent picture of an AI field progressing rapidly on multiple fronts simultaneously. Technical capabilities are advancing faster than governance frameworks, and applications are expanding into domains with high social stakes, including health, climate, and finance. The convergence of more capable foundation models, improved alignment techniques, and expanding multimodal architectures suggests that the systems deployed in the latter half of the 2020s will be qualitatively different from those available today.

## 5. Conclusion

This review has synthesized 20 high-impact studies to characterize the current state and emerging trajectories of artificial intelligence across technical, applied, and societal dimensions. The evidence consistently points to a technology advancing faster than anticipated, exhibiting qualitatively new capabilities at scale, and penetrating domains — from molecular biology to climate science to financial markets — that were previously considered beyond the reach of computational methods. Expert forecasts place transformative milestones within the current decade, while ethical and governance frameworks remain insufficiently developed to manage the risks such transformations entail. Several limitations of this review should be acknowledged: the selection of 20 papers, while carefully conducted, cannot fully represent a field producing tens of thousands of publications annually; the review does not include grey literature or non-English publications; and rapid advances in 2024 and 2025 may have superseded some findings reviewed here.

From a managerial perspective, organizations across sectors should prioritize investment in AI literacy, data infrastructure, and governance frameworks rather than treating AI adoption as a purely technical initiative. The evidence reviewed here suggests that AI systems will become increasingly generalist, multimodal, and capable of autonomous reasoning; organizations that develop the institutional capacity to evaluate, deploy, and oversee such systems will be better positioned to capture their benefits while managing their risks. Regulators and policymakers should engage proactively with the research community — as exemplified by the expert survey findings of Grace et al. (2024) — to anticipate capability transitions and develop proportionate governance mechanisms before, rather than after, transformative systems are deployed at scale.

Future research should prioritize several areas that this review has identified as underserved. First, longitudinal evaluation of deployed AI systems in real-world healthcare, judicial, and financial contexts is needed to establish whether laboratory performance translates into genuine social benefit. Second, the mechanisms underlying emergent abilities in large models remain poorly understood, and theoretical advances in this area would substantially improve AI safety research. Third, interdisciplinary research connecting AI capability development with ethics, economics, and political science is essential to ensure that the governance of AI reflects the full complexity of its societal implications. Finally, greater attention to the global distribution of AI benefits and harms — particularly with respect to lower-income countries and historically marginalized

populations — is necessary to realize the transformative potential of AI in alignment with principles of equity and sustainable development.

## References


Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Delgado, J. M. D., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities, and future challenges. *Journal of Building Engineering, 44,* Article 103299. https://doi.org/10.1016/j.jobe.2021.103299

Ahmed, S., Alshater, M. M., El Ammari, A., & Hammami, H. (2022). Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance, 61,* Article 101646. https://doi.org/10.1016/j.ribaf.2022.101646

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv* preprint. https://doi.org/10.48550/arXiv.2108.07258

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33,* 1877–1901. https://doi.org/10.48550/arXiv.2005.14165

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv* preprint. https://doi.org/10.48550/arXiv.2303.12712

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). A definition, benchmark, and database of AI for social good initiatives. *Nature Machine Intelligence, 3*(2), 111–115. https://doi.org/10.1038/s42256-021-00296-0

Dwivedi, Y. K., Sharma, A., Rana, N. P., Giannakis, M., Goel, P., & Dutot, V. (2023). Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions. *Technological Forecasting and Social Change, 192,* Article 122579. https://doi.org/10.1016/j.techfore.2023.122579

Grace, K., Stewart, H., Sandkühler, J. F., Thomas, S., Weinstein-Raun, B., & Brauner, J. (2024). Thousands of AI authors on the future of AI. *Journal of Artificial Intelligence Research.* https://doi.org/10.48550/arXiv.2401.02843

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. *arXiv* preprint. https://doi.org/10.48550/arXiv.2001.08361

Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature, 616*(7956), 259–265. https://doi.org/10.1038/s41586-023-05881-4

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35,* 27730–27744. https://doi.org/10.48550/arXiv.2203.02155

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *Proceedings of the 38th International Conference on Machine Learning (ICML 2021), 139,* 8748–8763. https://doi.org/10.48550/arXiv.2103.00020

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents. *arXiv* preprint. https://doi.org/10.48550/arXiv.2204.06125

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A. S., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D., Platt, J. C., & Bengio, Y. (2022). Tackling climate change with machine learning. *ACM Computing Surveys, 55*(2), Article 42. https://doi.org/10.1145/3485128

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications, 11,* Article 233. https://doi.org/10.1038/s41467-019-14108-y

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent abilities of large language models. *Transactions on Machine Learning Research.* https://doi.org/10.48550/arXiv.2206.07682

Yu, D., & Xiang, B. (2023). Discovering topics and trends in the field of artificial intelligence: Using LDA topic modeling. *Expert Systems with Applications, 225,* Article 120114. https://doi.org/10.1016/j.eswa.2023.120114