BRAIN. Broad Research in Artificial Intelligence and Neuroscience
e-ISSN: 2067-3957 | p-ISSN: 2068-0473
Covered in: Web of Science (ESCI); EBSCO; JERIH PLUS (hkdir.no); IndexCopernicus; Google Scholar; SHERPA/RoMEO; ArticleReach Direct; WorldCat; CrossRef; Peeref; Bridge of Knowledge (mostwiedzy.pl); abcdindex.com; Editage; Ingenta Connect Publication; OALib; scite.ai; Scholar9; Scientific and Technical Information Portal; FID Move; ADVANCED SCIENCES INDEX (European Science Evaluation Centre, neredataltics.org); ivySCI; exaly.com; Journal Selector Tool (letpub.com); Citefactor.org; fatcat!; ZDB catalogue; Catalogue SUDOC (abes.fr); OpenAlex; Wikidata; The ISSN Portal; Socolar; KVK-Volltitel (kit.edu)
2026, Volume 17, Issue 2, pages: 115-143
Submitted: March 17th, 2025 | *Accepted for publication:* May 23rd, 2025

Unified Understanding of Artificial Intelligence Systems: A Comprehensive Review and the Capability–Constraint–Infrastructure (CCI) Framework

Denisa-Daniela Frimu-Pascu
National University of Science and Technology POLITEHNICA Bucharest, Bucharest, Romania.
denisa.pascu@stud.acs.upb.ro
https://orcid.org/0009-0001-1792-3616

Ciprian Dobre
National University of Science and Technology POLITEHNICA Bucharest Romanian Academy of Sciences
National Institute for Research and Development in Informatics Bucharest, Romania.
ciprian.dobre@upb.ro
https://orcid.org/0000-0003-4638-7725

Edmond Gabriel Olteanu
Faculty of Law, University of Craiova, Craiova, Romania.
edmond.olteanu@edu.ucv.ro

Abstract: *This systematic review synthesises 229 scholarly records on contemporary artificial intelligence (AI), complemented by selected foundational works needed for theoretical interpretation. The evidence is organised across five analytical dimensions: (1) technologies and architectures, (2) infrastructure requirements, (3) reasoning capabilities, (4) technical and practical limitations, and (5) real-world use cases. The review follows PRISMA-aligned reporting principles and analyses the literature through an operational Capability–Constraint–Infrastructure (CCI) framework. The framework defines normalised component scores for capability, constraint burden, infrastructure adequacy, operational fit, and societal-regulatory readiness, and combines them in an effective AI performance index. This quantitative interpretation is complemented by a four-level CCI maturity model with explicit threshold logic. The synthesis shows that current AI progress is driven by substantial gains in perception, generation, and narrow-task performance, but remains constrained by reasoning limitations, catastrophic forgetting, interpretability challenges, data dependence, and increasing computational cost. To align the analysis with neuroscience-oriented AI research, the review examines neuro-symbolic AI, cognitive architectures, continual learning, and biologically inspired efficiency as bridges between engineering practice and brain-inspired intelligence. A legal-governance perspective is incorporated by treating documentation, auditability, privacy, human oversight, and compliance-by-design as practical ingredients of deployment readiness in regulated domains. Overall, the paper argues that sustainable AI advancement depends less on scale alone and more on measurable alignment among capability, constraint management, infrastructure, and socially acceptable deployment conditions.*

Keywords: *artificial intelligence; systematic review; neuro-symbolic AI; cognitive architectures; continual learning; explainable AI; AI governance; AI regulation; cloud computing; capability – constraint – infrastructure framework.*

How to cite: Frimu-Pascu, D.-D., Dobre, C., & Olteanu, E. G. (2026). Unified understanding of artificial intelligence systems: A comprehensive review and the Capability–Constraint–Infrastructure (CCI) framework. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, *17*(2), 115-143. https://doi.org/10.70594/brain/17.2/8

1. Introduction

Artificial intelligence has transitioned from theoretical exploration to practical implementation in virtually every sector of modern society. The rapid evolution from symbolic AI systems to contemporary deep learning architectures has fundamentally transformed how machines process information and recognise patterns to support human decision-making (Sharma & Garg, 2021). AI technologies underpin critical applications ranging from medical diagnostics and financial fraud detection to autonomous systems and personalised education (Shankar, 2024; Saghiri et al., 2022). However, widespread adoption has revealed significant technical, ethical, and practical challenges that constrain AI's effectiveness and raise important questions about its future development trajectory.

The current AI landscape is characterised by remarkable achievements in narrow, well-defined tasks, particularly those involving pattern recognition in large datasets (Kotyal et al., 2024; Safitra et al., 2024). Deep learning models have demonstrated superhuman-level performance in image classification, natural language processing, and game playing (Tiwari et al., 2025; Ahmed et al., 2023). However, these successes mask fundamental limitations in reasoning, generalisation, explainability, and resource efficiency that become apparent when AI systems are deployed in complex real-world environments (Karanam et al., 2025; Lu et al., 2023). The computational infrastructure required to train and deploy state-of-the-art models has expanded exponentially, raising concerns about energy consumption and environmental sustainability (Miya et al., 2025; Mukhamediev et al., 2022).

Understanding the current state of AI requires an examination of multiple interconnected dimensions. Technological foundations and model architectures that enable AI capabilities, changes in computational infrastructure that support training and deployment, and fundamental limitations in reasoning and algorithmic approaches are among these dimensions. On the practical side, challenges encountered in real-world applications and diverse use cases that demonstrate both AI's potential and its constraints warrant further investigation . This comprehensive review addresses these dimensions through a systematic analysis of recent literature to provide researchers, practitioners, and policymakers with an integrated perspective on where AI currently stands and where it is heading.

The significance of this review lies in its integrated approach to analysing AI technology. While previous surveys have examined specific aspects of AI—such as particular algorithms, application domains, or ethical concerns, few have provided an integrated analysis that connects technical capabilities with infrastructure requirements, algorithmic limitations, and practical deployment challenges (Nanjundan et al., 2025; Sengar et al., 2025). This gap is particularly problematic as organisations and policymakers make critical decisions about AI adoption, reflected in investments strategies and regulatory frameworks, without a comprehensive understanding of the technology's current state and fundamental constraints. In regulated settings, deployability depends not only on benchmark performance but also on traceability, data governance, accountability allocation, privacy protection, and the ability to justify system behaviour to auditors, affected users, and sectoral authorities. A review framework that ignores these legal-operational conditions risks overstating real-world maturity.

To address the fragmented understanding of AI systems across performance, infrastructure, and deployment dimensions, this paper proposes an operational *Capability–Constraint–Infrastructure (CCI) framework*. Unlike prior surveys that focus on isolated aspects such as algorithms, models, or application domains, the proposed framework provides a unified analytical model linking AI capabilities with inherent constraints and deployment environments. This perspective enables a deeper understanding of the trade-offs that govern real-world AI system performance, scalability, and sustainability.

The remainder of this article is organised as follows: Section II presents a literature review establishing the theoretical foundations and evolution of AI technologies. Section III describes the methodology for selecting and analysing the relevant literature. Sections IV through VII present detailed findings on AI technologies and models, infrastructure requirements, reasoning capabilities and limitations, and technical challenges. Section VIII examines real-world practical applications across multiple domains. Section IX introduces and formalises the Capability–Constraint–Infrastructure framework, including its operational variables, performance index, illustrative medical applications, and maturity model. Section X discusses the broader implications, limitations, and future validation requirements. Finally, Section XI by outlining implications for research, practice, and policy.

2. Literature Review

2.1. Scope and Methodology

The field of artificial intelligence has experienced substantial growth in both research output and practical applications in the past decade. In this dynamic landscape, comprehensive surveys of AI technologies have become essential for understanding the rapid evolution of the field (Sengar et al., 2025; Sharma & Garg, 2021). Recent literature reviews have examined AI from various perspectives, including technical architectures (Ahmed et al., 2023), application domains (Kotyal et al., 2024), ethical implications (Saghiri et al., 2022), and deployment challenges (Karanam et al., 2025). However, the interdisciplinary nature of AI research often results in fragmented knowledge, with technical advances in algorithms and models disconnected from practical considerations of infrastructure, deployment, and real-world constraints.

Several comprehensive surveys have attempted to address this gap. Sharma and Garg (2021) presented a broad survey of AI technologies, applications, and challenges, analysing the limitations and hardware requirements across major algorithms and models. Ahmed et al. (2023) focused specifically on deep learning modelling techniques, examining current progress, advantages, and challenges of neural network architectures. Mukhamediev et al. (2022) reviewed AI and machine learning technologies, with an emphasis on classification approaches, as well as their limitations, opportunities, and challenges. As shown in Figure 1, these foundational studies establish that understanding AI requires simultaneous considerations of technical capabilities, infrastructure requirements, and practical limitations.

2.2. Evolution of AI Technologies

As shown in Figure 2, the evolution of AI can be conceptualised through distinct developmental transitions, from early symbolic systems and expert systems to statistical machine learning and, most recently, to deep learning and large-scale neural networks (Radanliev, 2025; Sharma et al., 2025). Early AI research focused on rule-based systems and logical reasoning, achieving success in constrained domains but struggling with the complexity and ambiguity of real -world problems (Sharma et al., 2025). The emergence of machine learning in the 1990s and 2000s shifted the focus towards statistical pattern recognition, enabling systems to learn from data rather than relying solely on hand-crafted rules.

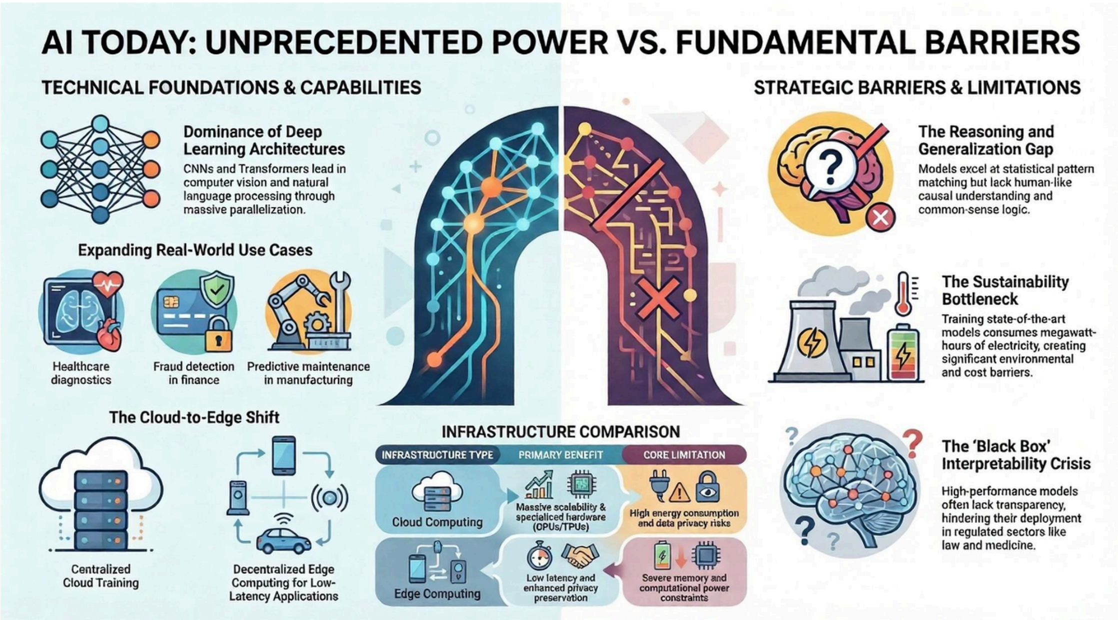

Figure 1. AI power vs. fundamental barriers

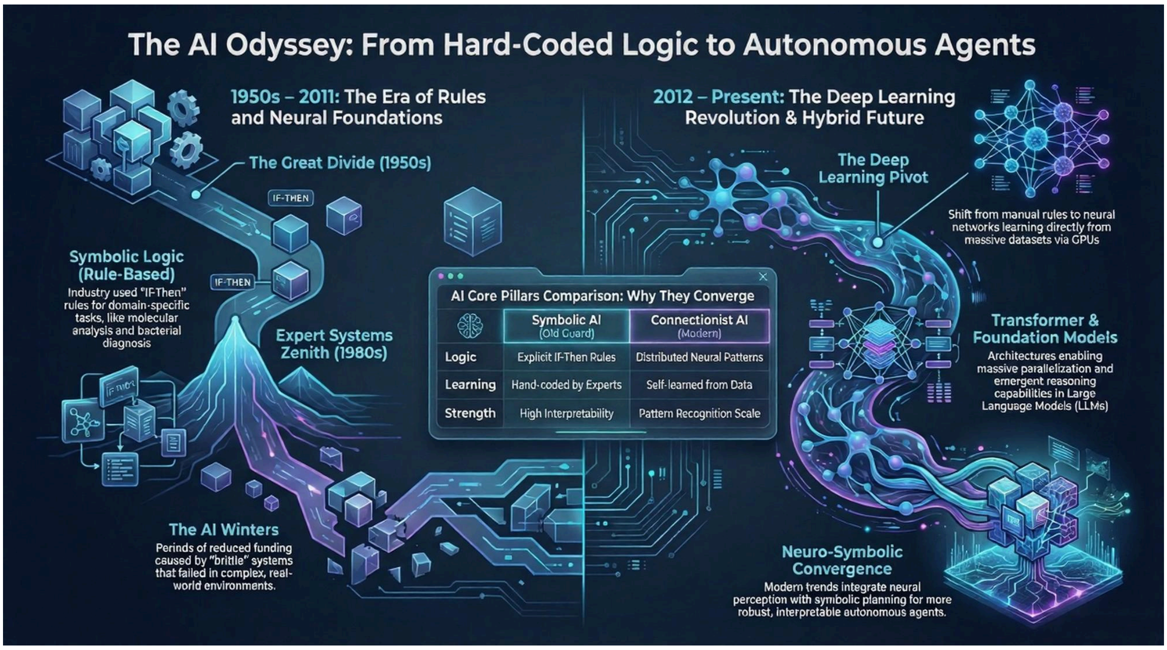

Figure 2. The evolution of Artificial Intelligence

By the 1990s and 2000s, researchers increasingly recognised that symbolic and connectionist approaches were complementary in nature rather than contradictory. Neuro-symbolic AI has evolved as a research paradigm explicitly designed to integrate neural networks with symbolic reasoning (Bhuyan et al., 2024; Medsker, 1994; Hassabis et al., 2017). This approach seeks to combine the learning capabilities of neural networks with the interpretability and reasoning capabilities of symbolic systems (Liang et al., 2025). Early hybrid systems demonstrated that neural networks could handle perception and pattern recognition, whereas symbolic components managed high-level reasoning and planning (Medsker, 1994).

Various integration models have been developed, including neural networks integrated within symbolic systems for specific tasks such as pattern recognition, symbolic knowledge used to constrain or guide neural network learning, and tightly coupled architectures in which symbolic and neural components interact bidirectionally (Bhuyan et al., 2024).

The convergence of paradigms was supported by theoretical work demonstrating the fundamental connections between symbolic and connectionist approaches. Researchers have demonstrated that neural networks can implement symbolic operations and that symbolic systems may be interpreted in terms of distributed representations (Bhuyan et al., 2024). This theoretical unification suggests that the historical divide was more about emphasis and methodology than fundamental incompatibility (Bhuyan et al., 2024; Ye, 2025). This led to the development of deep learning, a connectionist approach that has achieved unprecedented success across diverse domains (Lappin, 2025). Key enablers for the development of deep neural networks with many layers that can learn hierarchical representations of data include the availability of large datasets, increased computational power (especially GPUs), and algorithmic innovations such as dropout, batch normalisation, and advanced optimisation methods (Lappin, 2025). Deep learning has achieved superhuman-level performance in tasks previously thought to require symbolic reasoning, including image recognition, natural language processing, and game playing (Mundlamuri et al., 2025).

The deep learning revolution, catalysed by advances in computational power, large datasets, and algorithmic innovations, has dominated AI research and applications since 2012 (Sharma and Garg, 2021; Ahmed et al., 2023). Convolutional neural networks have transformed computer vision, recurrent neural networks have enabled sequential data processing, and attention mechanisms and transformer architectures have revolutionised natural language processing (Ahmed et al., 2023; Lu et al., 2023). Recently, large language models and foundation models have demonstrated remarkable capabilities in few-shot learning and cross-domain transfer, although they require significant computational and data resources (Bi et al., 2025; Heigl, 2025).

Large language models (LLMs) represent a significant development, exhibiting capabilities that blur the boundaries between connectionist and symbolic approaches. Neural networks trained on massive text corpora demonstrate emergent abilities in reasoning, planning, and knowledge manipulation—tasks traditionally associated with symbolic AI (Liang et al., 2025; Ye, 2025). Contemporary AI agents increasingly combine neural network components for perception and learning with symbolic or structured components for planning, reasoning, and knowledge management (Kota, 2025; Kotseruba & Tsotsos, 2020). This synthesis reflects the maturation of both paradigms and the recognition that different aspects of intelligence may require different computational approaches (Bhuyan et al., 2024; Kota, 2025).

2.3. Neuro-symbolic Convergence, Cognitive Architectures, and Biological Efficiency

Neuro-symbolic convergence is relevant not only as an engineering strategy but also as a computational hypothesis about cognition. Cognitive architectures typically distinguish perception, working memory, long-term memory, action selection, and deliberative reasoning as partially specialised but interacting functions (Kotseruba & Tsotsos, 2020). This organisation can be mapped onto neuro-symbolic AI: neural components handle perception, representation learning, and uncertainty-tolerant pattern extraction, while symbolic components support compositional structure, explicit memory, planning, constraint handling, and rule-governed inference (Bhuyan et al., 2024; Lake et al., 2017; Medsker, 1994).

The comparison with biological neural processing is instructive. Human cognition does not rely solely on large-scale pattern matching. It combines fast perceptual processing with abstraction, memory consolidation, causal expectations, and goal-directed behaviour. Neuroscience-inspired AI therefore emphasises structured world models, modular learning, attentional control, and efficient adaptation (Hassabis et al., 2017; Lake et al., 2017). These mechanisms are not direct copies of the brain, but they identify design principles that current deep learning often lacks: sample efficiency, compositional generalisation, robustness under distribution shift, and the ability to integrate new experience without erasing prior knowledge.

Biological efficiency also challenges the scale-first trajectory of contemporary AI. The human brain supports flexible behaviour under severe energy constraints, while many frontier AI systems improve by increasing parameters, data, and accelerator demand. From this perspective, sparse computation, event-driven processing, neuromorphic hardware, continual learning, and hardware-software co-design are not only deployment optimisations. They represent potential pathways toward AI systems with enhanced CCI profiles because they can increase capability while reducing energy, memory, and latency burdens (Friston, 2010; Hassabis et al., 2017; Parisi et al., 2019).

Within the CCI framework, neuro-symbolic and cognitive-architecture approaches occupy a distinctive position. They aim to increase reasoning-oriented capability without relying exclusively on parameter scaling, and they can reduce constraint burden by embedding prior knowledge, modular control, or explicit representations. This makes them especially relevant for high-stakes settings where interpretability, sample efficiency, and auditable behaviour matter as much as raw benchmark accuracy.

2.4. Theoretical Foundations

The theoretical foundations of contemporary AI draw on several key fields, including statistical learning theory for generalisation and model complexity, optimisation theory for training algorithms, information theory for representation learning, and computational complexity theory for limits on efficient computation (Ahmed et al., 2023; Mukhamediev et al., 2022). Deep learning architectures employ hierarchical feature learning, where multiple layers of nonlinear transformations progressively extract increasingly abstract representations from raw data (Ahmed et al., 2023).

However, significant theoretical gaps remain. The practical success of deep learning often outpaces theoretical understanding, and current models still lack strong formal guarantees for generalisation, robustness, and convergence (Ahmed et al., 2023; Lu et al., 2023). Their reasoning capabilities also remain limited compared with human cognition, especially in abstract reasoning, causal inference, and common-sense understanding (Liang et al., 2025; Lu et al., 2023). These limitations have direct consequences for AI deployment and operational reliability.

Current research therefore emphasises integration between symbolic and connectionist paradigms (Bhuyan et al., 2024; Kotseruba & Tsotsos, 2020; Liang et al., 2025). LLM-based autonomous agents exemplify this convergence: they use neural networks for language understanding and generation while incorporating symbolic structures for task planning, memory management, and tool use (Kotseruba & Tsotsos, 2020). The field is increasingly oriented toward systems that learn from data and reason with explicit knowledge, combining the strengths of both historical traditions (Bhuyan et al., 2024; Ye, 2025).

Researchers propose that future AI systems will require this integration to achieve robust and generalisable intelligence (Bhuyan et al., 2024; Lake et al., 2017). Contemporary AI research increasingly recognises that different aspects of intelligence may require different computational approaches and that powerful systems will integrate multiple paradigms (Bhuyan et al., 2024; Hassabis et al., 2017). The future of AI likely lies not in choosing between rule-based and agent-based approaches, but in understanding how to combine them to create systems that are both powerful and principled (Kotseruba & Tsotsos, 2020; Liang et al., 2025).

3. Research Methodology

This review was conducted as a structured qualitative synthesis of contemporary AI research and reported in accordance with PRISMA 2020 guidelines (Page et al., 2021). The objective was not only to catalogue architectures and applications, but also to analyse how capability, constraint burden, infrastructure dependence, reasoning capacity, and deployment context vary across the literature.

The search strategy combined broad AI survey terms with targeted terms for reasoning, infrastructure, neuro-symbolic AI, interpretability, continual learning, responsible AI, and domain-specific deployment. Searches targeted peer-reviewed journal articles, conference proceedings, systematic reviews, and high-quality technical syntheses published primarily between 2020 and early 2025. Foundational works published before 2020 were retained when they were necessary for theoretical framing, especially in relation to neuroscience-inspired AI, cognitive architectures, PRISMA reporting, explainability, and continual learning. Preprints and technical reports were treated as contextual evidence unless their claims were corroborated by peer-reviewed or institutional sources. Where recent preprints were included, they were used primarily to represent emerging technical directions rather than as sole evidence for established scientific claims.

The exact query families used for retrieval are presented in Table I. Search syntax was adjusted to the conventions of each database or publisher platform while preserving the same Boolean logic and conceptual scope.

Table 1. Search strings used for the PRISMA-aligned retrieval strategy

ID	Search string / Boolean family
S1	"artificial intelligence" AND (technologies OR applications OR challenges OR limitations) AND (review OR survey)
S2	("machine learning" OR "deep learning") AND (architectures OR models OR applications) AND (review OR survey)
S3	("foundation models" OR "large language models" OR "generative AI") AND (capabilities OR limitations OR applications)
S4	("AI reasoning" OR "machine reasoning" OR generalisation) AND ("deep learning" OR "neural networks" OR "large language models")
S5	("neuro-symbolic AI" OR "neural symbolic" OR "cognitive architectures") AND (reasoning OR learning OR cognition)
S6	("catastrophic forgetting" OR "continual learning" OR "lifelong learning") AND (neural OR artificial intelligence)
S7	("explainable AI" OR interpretability OR "mechanistic interpretability") AND (accuracy OR trust OR high-stakes)
S8	("edge AI" OR "cloud AI" OR "hardware acceleration" OR "neuromorphic") AND (infrastructure OR deployment OR efficiency)
S9	("responsible AI" OR "AI governance" OR fairness OR privacy OR regulation) AND (deployment OR healthcare OR finance OR public services)

The retrieval stage yielded 659 records across the nine searches. After merging search outputs and removing duplicate or clearly ineligible records, 229 studies remained for eligibility assessment and qualitative coding. AI-assisted relevance re-ranking was used only to support prioritisation during analysis; it was not used as an inclusion criterion. To be included, articles needed to provide substantive technical, empirical, or conceptual analysis of at least one of the five review dimensions. Exclusion criteria included purely opinion-based commentary, inaccessible full text, non-scholarly web material, and papers with no direct link to AI capability, limitations, infrastructure, reasoning, ethics, or deployment evidence.

The PRISMA diagram documents how records were identified, deduplicated, assessed for eligibility, and included for qualitative synthesis. The final corpus of 229 records was retained for full qualitative coding because each included record contributed substantive evidence to at least one analytical dimension. Records removed before eligibility assessment included duplicates, inaccessible items, clearly out-of-scope records, and sources without direct relevance to AI capability, limitations, infrastructure, reasoning, ethics, or deployment.

Because several studies addressed more than one analytical dimension, the category counts are not mutually exclusive and therefore do not sum to 229. This coding structure links the literature synthesis to the later CCI variables and clarifies how the reviewed evidence informed the framework.

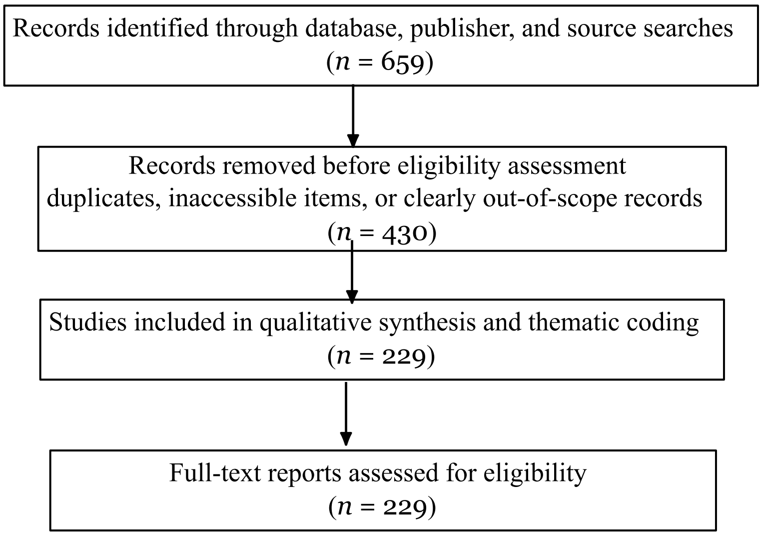

Figure 3. PRISMA-aligned overview of the study selection pipeline

For each paper, the extraction form recorded the AI paradigm under study, target task, model family, infrastructure assumptions, limitations reported by the authors, and any discussion relevant to deployment risk, ethics, regulation, or operational context. This coding schema directly supports the CCI formalisation: capability indicators were distilled from task performance and functional breadth, constraints from cost and failure modes, infrastructure from compute and deployment assumptions, operational fit from domain workflow evidence, and societal-regulatory readiness from evidence on privacy, fairness, auditability, oversight, and compliance.

The resulting method remains qualitative rather than meta-analytical because the included studies use heterogeneous tasks, datasets, benchmarks, and reporting standards. However, the PRISMA-aligned selection diagram, explicit search strings, dimension-level coding, and operational link between coding categories and CCI variables provide a reproducible bridge between the literature synthesis and the proposed analytical model.

4. AI Technologies, Models, and Architectures

4.1. Machine Learning Paradigms

Contemporary AI systems employ three primary machine learning paradigms: supervised, unsupervised, and reinforcement learning (Mukhamediev et al., 2022; Sharma & Garg, 2021). Supervised learning trains models on labelled data to predict outputs for new inputs and dominates practical applications in classification and regression tasks (Kotyal et al., 2024; Safitra et al., 2024). Deep neural networks have become the predominant approach in supervised learning, achieving state-of-the-art performance across computer vision, natural language processing, and speech recognition (Ahmed et al., 2023; Sharma & Garg, 2021).

Unsupervised learning techniques, including clustering, dimensionality reduction, and generative modelling, enable the discovery of patterns and structures in unlabelled data (Mukhamediev et al., 2022). Recent advances in self-supervised and contrastive learning show that models can learn powerful representations from unlabeled data, reducing dependence on expensive manual annotations (Ahmed et al., 2023). Reinforcement learning trains agents to make sequential decisions through interaction with the environment and has achieved success in game playing, robotics, and control systems, although practical deployment remains challenging because of sample inefficiency and safety concerns (Dulaj et al., 2025; Sharma & Garg, 2021).

4.2. Deep Learning Architectures

Deep learning architectures constitute the technical foundation of most contemporary AI systems. Convolutional neural networks remain the dominant approach for computer vision tasks because they use spatial hierarchies and local connectivity to process image data efficiently (Ahmed et al., 2023; Saghiri et al., 2022). Hybrid architectures that combine CNNs with recurrent neural networks address challenges in sequential data processing and temporal modelling (Safitra et al., 2024; Saghiri et al., 2022). Long short-term memory networks and gated recurrent units have also proven effective for time-series analysis, natural language processing, and speech recognition (Ahmed et al., 2023; Mukhamediev et al., 2022).

The transformer architecture, introduced in 2017, has transformed natural language processing and increasingly influenced other domains (Heigl, 2025; Lu et al., 2023). Transformers use self-attention mechanisms to capture long-range dependencies without the sequential processing constraints of RNNs, enabling massive parallelisation during training (Ahmed et al., 2023). Large language models built on transformer architectures, such as GPT and BERT variants, have demonstrated strong few-shot learning capabilities and cross-task generalisability (Bi et al., 2025; Heigl, 2025). However, these models require substantial computational resources and training data, raising questions about accessibility and sustainability (Bi et al., 2025; Miya et al., 2025).

Graph neural networks are an emerging class of architectures designed to process structured data represented as graphs, with applications in social network analysis, molecular chemistry, and knowledge representation (Ahmed et al., 2023; Mukhamediev et al., 2022). Generative models, including adversarial networks and diffusion models, synthesise realistic images, text, and other data types, with applications ranging from content creation to data augmentation (Ahmed et al., 2023; Safitra et al., 2024).

4.3. Emerging Model Structures

Recent research has focused on more efficient and capable model structures that address limitations in current architectures. Mixture-of-experts models dynamically route inputs to specialised sub-networks, improving the efficiency and scalability of large models (Bi et al., 2025). Neural architecture search automates the design of network architectures optimised for specific tasks and hardware constraints (Ahmed et al., 2023; Mukhamediev et al., 2022). Attention mechanisms continue to evolve, with variants designed to reduce computational complexity while maintaining modelling capacity (Liang et al., 2025; Lu et al., 2023).

Foundation models and large-scale pre-trained models represent a transition toward general-purpose AI systems that can be adapted to diverse downstream tasks with minimal task-specific training (Bi et al., 2025; Heigl, 2025). These models demonstrate emergent capabilities that are not explicitly programmed, including in-context learning and chain-of-thought reasoning (Lu et al., 2023). However, their size, training costs, and environmental impact raise concerns about accessibility and sustainability (Bi et al., 2025; Miya et al., 2025). Efficient model structures, including pruning, quantisation, and knowledge distillation, aim to compress large models while preserving performance (Mukhamediev et al., 2022; Shankar, 2024).

5. Processing Infrastructure and Computational Requirements

5.1. Cloud and Distributed Computing

The computational demands of modern AI systems have driven widespread adoption of cloud computing infrastructure for both training and deployment (Miya et al., 2025; Sharma and Garg, 2021). Cloud platforms provide access to specialised hardware accelerators, distributed training frameworks, and scalable inference services that would be too costly for many organisations to maintain independently (Dulaj et al., 2025; Sharma and Garg, 2021). Major cloud providers also offer AI-specific services, including pre-trained models, AutoML platforms, and managed machine learning pipelines that reduce barriers to adoption (Kotyal et al., 2024; Safitra et al., 2024).

Distributed training across multiple GPUs or TPUs has become essential for training large-scale AI models. Techniques such as data parallelism, model parallelism, and pipeline parallelism improve the utilisation of computational resources (Ahmed et al., 2023; Miya et al., 2025). However, distributed training introduces challenges involving data locality, communication overhead, and fault tolerance (Mukhamediev et al., 2022). The concentration of AI training capacity in large cloud providers also raises concerns about accessibility, cost, and environmental impact because training state-of-the-art models can consume large amounts of electricity and cost millions of dollars (Bi et al., 2025; Miya et al., 2025).

5.2. Edge Computing Architectures

Edge AI deploys models on resource-constrained devices at the network edge and has become increasingly important for applications requiring low latency, privacy preservation, and operation in disconnected environments (Miya et al., 2025; Shankar, 2024). Edge computing architectures that support federated learning distribute neural network training and inference across edge devices and cloud servers, balancing computational efficiency with communication costs (Dulaj et al., 2025; Shankar, 2024). Applications include autonomous vehicles, industrial IoT, mobile devices, and smart sensors (Kotyal et al., 2024; Shankar, 2024; Toma et al., 2025).

However, deployment at the edge faces challenges because such devices have limited computational power, restricted memory, and tight energy budgets (Miya et al., 2025; Shankar, 2024). Model compression techniques, including pruning, quantisation, and knowledge distillation, are essential for deploying neural networks on edge hardware (Mukhamediev et al., 2022; Shankar, 2024). Hardware-software co-design optimises both model architectures and hardware implementations for specific edge platforms (Mukhamediev et al., 2022; Shankar, 2024). Despite these advances, the trade-off between model capability and resource constraints remains a central challenge for edge AI (Miya et al., 2025; Shankar, 2024).

5.3. Hardware Acceleration and Resource Constraints

Specialised hardware accelerators are essential for AI workloads. Graphics processing units provide high-degree parallelism for matrix operations central to neural network training and inference (Ahmed et al., 2023; Sharma and Garg, 2021). Tensor processing units and other AI-specific accelerators further optimise deep learning workloads (Miya et al., 2025; Mukhamediev et al., 2022). Field-programmable gate arrays and application-specific integrated circuits enable customised hardware implementations tailored to specific models or applications (Dulaj et al., 2025; Shankar, 2024).

Despite these advances, hardware limitations remain significant. Memory bandwidth and capacity often acts as a bottleneck for performance, particularly for large models that exceed on-chip memory (Miya et al., 2025; Shankar, 2024). Energy consumption remains a critical concern for both data centre deployment and edge devices (Bi et al., 2025; Shankar, 2024). The high cost of specialised AI hardware limits accessibility, creating disparities between well-resourced organisations and smaller entities (Miya et al., 2025; Mukhamediev et al., 2022). Emerging hardware technologies, including neuromorphic and quantum computing, promise new capabilities but remain in early research stages (Mukhamediev et al., 2022; Saghiri et al., 2022).

6. Reasoning Capabilities and Algorithmic Limitations

6.1. Current Reasoning Approaches

Contemporary AI systems use several approaches to reasoning, although they remain limited compared with human cognitive capabilities. Neural-symbolic integration seeks to combine the pattern-recognition strengths of neural networks with the logical reasoning capabilities of symbolic AI systems (Lu et al., 2023; Mukhamediev et al., 2022). Knowledge graphs and structured knowledge representations support structured reasoning over factual information and relationships

(Dulaj et al., 2025; Mukhamediev et al., 2022). Attention mechanisms and transformer architectures have demonstrated emergent reasoning capabilities, including multi-hop and analogical reasoning in large language models (Heigl, 2025; Lu et al., 2023).

Mathematical reasoning has recently received particular attention, with specialised models and training approaches designed to solve mathematical problems (Lu et al., 2023). Chain-of-thought prompting and related techniques encourage models to generate intermediate reasoning steps, improving performance on complex reasoning tasks (Bi et al., 2025; Lu et al., 2023). However, these approaches remain fragile and often fail in problems requiring genuine abstraction or novel problem-solving strategies (Liang et al., 2025; Lu et al., 2023). Current AI reasoning remains largely tied to pattern matching and statistical associations learned from training data, lacking the causal understanding and commonsense reasoning that humans use flexibly (Lu et al., 2023; Mukhamediev et al., 2022).

6.2. Fundamental Algorithmic Constraints

Deep learning algorithms face several constraints that limit reasoning and generalisation. Their reliance on large-scale training data means that models perform poorly in tasks or domains where data are scarce or expensive to obtain (Karanam et al., 2025; Kotyal et al., 2024). Models can learn spurious correlations and dataset-specific patterns rather than underlying causal relationships, leading to poor generalisation when distributions shift (Karanam et al., 2025; Mukhamediev et al., 2022). Adversarial examples also reveal the fragility of learned representations by demonstrating that carefully crafted inputs can mislead models (Ahmed et al., 2023; Karanam et al., 2025).

The black-box nature of deep neural networks makes it difficult to understand or verify their reasoning processes, creating challenges for debugging, validation, and trust (Karanam et al., 2025; Nanjundan et al., 2025). Models often lack explicit representations of uncertainty and can produce confident predictions even when extrapolating far beyond their training distribution (Karanam et al., 2025; Mukhamediev et al., 2022). Compositional generalisation, or the ability to combine learned concepts in novel ways, remains a significant challenge because models struggle to apply learned knowledge to structurally different problems (Liang et al., 2025; Lu et al., 2023). These algorithmic limitations reflect the gap between current AI approaches and human-like intelligence (Lu et al., 2023; Mukhamediev et al., 2022).

6.3. Explainability and Interpretability Challenges

Explainability and interpretability have become critical concerns for AI deployment in high-stakes domains such as healthcare, finance, and criminal justice (Karanam et al., 2025; Nanjundan et al., 2025). Deep neural networks function as complex non-linear mappings with millions or billions of parameters, making it difficult to understand why a model produces a particular output (Ahmed et al., 2023; Karanam et al., 2025). Post-hoc explanation techniques, including saliency maps, attention visualisation, and feature importance methods, provide partial insight but often fail to capture the true reasoning process (Karanam et al., 2025; Nanjundan et al., 2025).

The trade-off between model performance and interpretability remains a fundamental challenge. Simpler and more interpretable models, such as decision trees or linear models, often underperform complex deep learning models, while the most capable models are frequently the least interpretable (Karanam et al., 2025; Mukhamediev et al., 2022). This tension is especially problematic in regulated domains, where explanations may be legally required or ethically necessary (Nanjundan et al., 2025; Toto et al., 2025). Recent research into inherently interpretable models and mechanistic interpretability aims to design models whose reasoning processes are transparent by construction, although these approaches have not yet matched the performance of standard deep learning in many settings (Karanam et al., 2025; Rudin, 2019). Limited explainability limits trust, accountability, and the ability to identify and correct model failures (Karanam et al., 2025; Nanjundan et al., 2025).

7. Technical and Practical Limitations

7.1. Data Requirements and Quality Issues

Modern AI systems, particularly deep learning models, require large quantities of high-quality labelled data for training (Karanam et al., 2025; Kotyal et al., 2024). Data collection, cleaning, and annotation are expensive, time-consuming and often become the main bottlenecks in AI development (Kotyal et al., 2024; Mukhamediev et al., 2022). Many domains lack sufficient data because of privacy constraints, rare events, or high annotation costs (Karanam et al., 2025; Kumar & Roy, 2025). Data quality problems, including label noise, missing values, and measurement errors, can also degrade model performance (Karanam et al., 2025; Kotyal et al., 2024).

Bias in training data can produce models that perpetuate or amplify societal inequities (Karanam et al., 2025; Nanjundan et al., 2025). Historical data often reflect discriminatory practices, and models trained on such data can reproduce these biases in their predictions (Karanam et al., 2025; Saghiri et al., 2022). Distribution shift, where deployment data differ from training data, can further reduce performance and cause unexpected failures (Karanam et al., 2025; Mukhamediev et al., 2022). Privacy concerns limit data sharing and aggregation, particularly in sensitive domains such as healthcare and finance (Kotyal et al., 2024; Kumar & Roy, 2025). Federated learning and differential privacy aim to enable model training while preserving privacy, but they introduce additional complexity and performance trade-offs (Dulaj et al., 2025; Mukhamediev et al., 2022).

7.2. Continual Learning and Catastrophic Forgetting

A central obstacle to robust AI deployment is the inability of many neural systems to learn continuously without overwriting previously acquired knowledge. This phenomenon, commonly described as *catastrophic forgetting*, emerges when gradient updates for a new task interfere with parameter configurations needed for earlier tasks (Parisi et al., 2019). In static benchmark evaluation, this weakness can remain hidden because training and testing are usually separated into fixed datasets. In long-lived deployment, however, the data distribution changes as patient populations, industrial processes, financial behaviours, language use, and environmental conditions evolve.

Three major mitigation families are commonly discussed. Replay-based methods preserve or generate examples from earlier tasks, but they introduce memory costs and may raise privacy concerns in sensitive domains. Regularisation-based methods constrain parameter updates to protect previously important knowledge, but they can limit plasticity when genuinely new conditions emerge. Dynamic-architecture methods allocate new capacity for new tasks, but they increase model complexity and infrastructure requirements. In CCI terms, continual learning is therefore not merely a capability upgrade. It is a coupled capability–constraint problem: adaptation can increase C and O, while added memory, monitoring, privacy, and validation burdens can increase K and reduce practical deployability if not explicitly managed.

7.3. Mechanistic Interpretability and Internal Representations

Interpretability should be distinguished from surface-level explanation. Post-hoc feature attribution, saliency maps, and local explanation tools can improve transparency, but they do not necessarily reveal the internal computations that drive model behaviuor (Rudin, 2019; Confalonieri et al., 2021). Mechanistic interpretability seeks a stronger objective: identifying the internal circuits, representations, subnetworks, or algorithmic motifs that implement specific functions. This goal matters because high-stakes AI systems require not only plausible explanations after a prediction, but also evidence that the system is using stable, clinically or operationally meaningful patterns rather than shortcuts, spurious correlations, or hidden proxies for protected attributes.

The scientific promise of mechanistic interpretability is substantial. It can support debugging, safety auditing, model editing, and a more cumulative understanding of learned representations. Its current limitation is scalability. Fine-grained mechanistic accounts are easier to produce for smaller models or restricted tasks than for large foundation models with billions of parameters and distributed internal representations. This gap is important for the CCI framework: benchmark performance may indicate high apparent capability, but weak mechanistic understanding increases constraint burden through opacity, audit difficulty, and residual safety risk. For this reason, interpretability is treated both as a capability-related property and as a societal-regulatory readiness condition.

7.4. Bias, Fairness, and Ethical Concerns

Algorithmic bias and fairness have emerged as critical ethical concerns in AI deployment (Karanam et al., 2025; Saghiri et al., 2022). AI systems can discriminate against protected groups based on race, gender, age, or other characteristics, either through biased training data or through proxy variables that correlate with protected attributes (Karanam et al., 2025; Nanjundan et al., 2025). Defining and measuring fairness is challenging because multiple competing fairness criteria cannot always be satisfied simultaneously (Karanam et al., 2025; Mukhamediev et al., 2022). The opacity of deep learning models also makes it difficult to detect and mitigate bias, and post-hoc debiasing techniques may reduce bias in training data while failing to generalise (Karanam et al., 2025; Nanjundan et al., 2025).

Beyond bias, AI systems raise broader ethical concerns, including privacy violations, manipulation, autonomous weapons, job displacement, and concentration of power (Nanjundan et al., 2025; Saghiri et al., 2022). The environmental impact of training large models has become a significant concern because carbon emissions from AI training can be substantial (Bi et al., 2025; Miya et al., 2025). Accountability and liability for AI system failures remain unclear, particularly when systems make consequential decisions that affect human lives (Nanjundan et al., 2025; Toto et al., 2025). Underdeveloped governance frameworks and regulatory uncertainty create additional challenges for organisations deploying AI systems (Nanjundan et al., 2025; Saghiri et al., 2022). Addressing these concerns requires collaboration among technologists, ethicists, policymakers, and users (Karanam et al., 2025; Saghiri et al., 2022).

From a legal-governance perspective, these concerns are no longer merely aspirational principles. Contemporary frameworks such as the NIST AI Risk Management Framework and the European Union's AI Act translate ethical concerns into operational expectations around risk classification, documentation, data governance, transparency, human oversight, robustness, and post-deployment monitoring (European Parliament and Council of the European Union, 2024; National Institute of Standards and Technology, 2023). In the language of the present review, deployment maturity depends partly on whether a system is auditable, contestable, and lawfully operable in its target setting, not only on whether it attains high predictive performance. This is especially relevant in high-impact domains such as healthcare, education, employment, finance, public administration, and law enforcement, where weak documentation or unclear accountability can become a barrier to deployment even when technical metrics appear satisfactory (European Parliament & Council of the European Union, 2024).

7.5. Scalability and Deployment Challenges

Deploying AI systems in real-world environments creates practical challenges beyond technical performance (Karanam et al., 2025; Miya et al., 2025). Integrating AI into existing IT infrastructure and legacy systems requires significant engineering effort (Kotyal et al., 2024; Saghiri et al., 2022). Models trained in controlled research environments often fail when confronted with real-world complexity, variability, and edge cases (He et al., 2020; Karanam et al., 2025). Maintaining and updating deployed models as data distributions evolve requires continuous monitoring and retraining (Karanam et al., 2025; Mukhamediev et al., 2022).

Inference costs can be prohibitive, particularly for large models or high-throughput applications (Bi et al., 2025; Miya et al., 2025). Latency requirements for real-time applications may be incompatible with the inference time of complex models. (Miya et al., 2025; Shankar, 2024). Ensuring reliability, safety, and robustness in critical applications requires extensive testing and validation, which current practices do not always address adequately (Karanam et al., 2025; Nanjundan et al., 2025). The shortage of skilled AI practitioners creates bottlenecks in development and deployment (Kotyal et al., 2024; Rahmaniar et al., 2025). Organisations also often lack the culture, processes, and governance structures needed to deploy and manage AI systems effectively (Kotyal et al., 2024; Kumar & Roy, 2025). These challenges explain why many AI projects fail to move from proof of concept to production deployment (He et al., 2020; Karanam et al., 2025).

8. Use Cases and Real-World Applications

8.1. Healthcare and Medical Diagnostics

Healthcare has emerged as one of the most promising application domains for AI, with implementations spanning medical imaging, diagnostics, drug discovery, and personalised treatment (Tiwari et al., 2025; Toto et al., 2025). AI-powered medical imaging systems have demonstrated diagnostic accuracy comparable to or exceeding that of human experts for selected tasks, including diabetic retinopathy screening, skin cancer detection, radiological image interpretation, and anatomical segmentation (Bunea et al., 2025; Tiwari et al., 2025; Toto et al., 2025). Deep learning models analyse medical images to identify pathologies, segment anatomical structures, and predict disease progression (Sharma et al., 2025; Tiwari et al., 2025).

Beyond imaging, AI systems support clinical decision-making by analysing electronic health records, predicting patient outcomes, and identifying high-risk patients (Sharma et al., 2025; Toto et al., 2025). Natural language processing extracts structured information from clinical notes and medical literature (Toto et al., 2025). Drug discovery applications use AI to identify promising molecular candidates, predict drug-target interactions, and optimise clinical trial designs (Kumar and Roy, 2025). Personalised medicine uses AI to tailor treatment based on individual patient characteristics, genetic profiles, and predicted responses (Sharma et al., 2025; Toto et al., 2025).

The accuracy–explainability trade-off is especially visible in medical AI. A model with high diagnostic accuracy may still be unsuitable for clinical use if it cannot offer an interpretable rationale that clinicians can inspect, if its outputs cannot be audited, or if its performance differs across patient subgroups. In image-based diagnosis, for example, a segmentation or classification system may perform well on diagnostic accuracy metrics while relying on scanner artefacts, site-specific acquisition patterns, or demographic proxies. Such behaviour raises clinical and legal concerns because clinicians must justify decisions, patients may contest automated outputs, and regulatory frameworks increasingly require documentation, risk management, human oversight, and post-market monitoring (European Parliament and Council of the European Union, 2024; National Institute of Standards and Technology, 2023; Rudin, 2019).

Healthcare AI also faces deployment barriers that extend beyond model design. Regulatory requirements for medical devices can create high barriers to deployment (He et al., 2020; Toto et al., 2025). Liability concerns and the need for clinical validation slow adoption (He et al., 2020). Bias in training data can lead to disparate performance across demographic groups, potentially worsening health inequities (He et al., 2020; Toto et al., 2025). Integration with clinical workflows and electronic health record systems requires substantial organizational effort (Toto et al., 2025). Despite these challenges, healthcare remains a high-priority domain for AI research and deployment because it makes the core CCI trade-off explicit: clinically useful AI must combine accuracy, interpretability, infrastructure reliability, privacy, and accountable human oversight.

8.2. Finance and Business Operations

Financial services have rapidly adopted AI for fraud detection, risk assessment, algorithmic trading, customer service, and process automation (Kotyal et al., 2024; Rahmaniar et al., 2025). Machine learning models analyse transaction patterns to identify fraudulent activity in real time and minimise financial losses (Kotyal et al., 2024; Sharma et al., 2025). Credit scoring and loan approval systems use AI to assess risk and support lending decisions, although fairness and discrimination concerns remain (Karanam et al., 2025; Kotyal et al., 2024). Algorithmic trading systems use reinforcement learning and predictive models to execute trades at high frequencies (Dulaj et al., 2025; Kotyal et al., 2024).

Chatbots and virtual assistants powered by natural language processing handle customer enquiries, cutting costs while increasing service availability (Abbasi et al., 2025; Kotyal et al., 2024). Robotic process automation uses AI to automate repetitive business processes, including data entry, invoice processing, and compliance reporting. Predictive maintenance systems in manufacturing and industrial settings use machine learning to forecast equipment failures and optimise maintenance schedules (Abbasi et al., 2025; Rahmaniar et al., 2025). Business intelligence applications use AI for demand forecasting, supply chain optimisation, and market analysis (Kotyal et al., 2024; Rahmaniar et al., 2025). Financial AI applications face regulatory scrutiny regarding fairness, transparency, and accountability (Karanam et al., 2025; Kotyal et al., 2024). The need for explainable decisions in regulated contexts conflicts with the opacity of complex models (Karanam et al., 2025; Nanjundan et al., 2025). Data privacy regulations limit data sharing and model training (Kotyal et al., 2024). Market volatility and distribution shift can also cause model failures with significant financial consequences (Karanam et al., 2025). Despite these challenges, the financial sector continues to invest heavily in AI technologies (Kotyal et al., 2024; Rahmaniar et al., 2025).

8.3. Manufacturing and Industrial Applications

Manufacturing and industrial sectors employ AI to support quality control, predictive maintenance, process optimisation, and the use of autonomous systems in general (Kotyal et al., 2024; Rahmaniar et al., 2025). Computer vision systems are used to inspect products for defects with greater consistency and speed than human inspectors (Kotyal et al., 2024; Safitra et al., 2024). Predictive maintenance models can be used to analyse equipment sensor data, forecast failures before they occur, and reduce downtime and maintenance costs (Abbasi et al., 2025; Rahmaniar et al., 2025). Finally, process optimisation uses reinforcement learning to improve efficiency, reduce waste, and minimise energy consumption (Kotyal et al., 2024; Dulaj et al., 2025).

Autonomous robots and cobots (collaborative robots) can work alongside humans in manufacturing environments, handling tasks ranging from assembly to material handling (Kotyal et al., 2024; Rahmaniar et al., 2025). Digital twins, which are virtual replicas of physical systems, enable the simulation and optimisation of manufacturing processes (Kotyal et al., 2024). Supply chain management systems further utilise AI for demand forecasting, inventory optimisation, and logistics planning (Rahmaniar et al., 2025; Abbasi et al., 2025). Industrial IoT platforms integrate AI with sensor networks to enable the real-time monitoring and control of complex, large-scale systems (Shankar, 2024; Kotyal et al., 2024).

In general, industrial AI deployments face challenges involving legacy-system integration, harsh operating environments, and safety requirements (Kotyal et al., 2024; Rahmaniar et al., 2025). The need for real-time processing and low latency drives adoption of edge computing architectures (Miya et al., 2025; Shankar, 2024). Data quality issues such as sensor noise and equipment variability affect model performance (Karanam et al., 2025; Kotyal et al., 2024). The shortage of workers with both domain expertise and AI skills also limits adoption (Rahmaniar et al., 2025). Nevertheless, manufacturing remains a major growth area for AI applications (Kotyal et al., 2024; Rahmaniar et al., 2025).

8.4. Education and Public Services

AI technologies are transforming education through personalised learning systems, automated grading, intelligent tutoring, and educational content generation (Kotyal et al., 2024; Saghiri et al., 2022). Adaptive learning platforms use AI to customise educational content and pacing based on student performance and learning style (Saghiri et al., 2022; Sharma et al., 2025).

Intelligent tutoring systems provide personalised feedback and guidance, supplementing human instruction (Saghiri et al., 2022). Automated essay scoring and grading systems reduce instructor workload, although fairness and accuracy concerns remain (Karanam et al., 2025; Saghiri et al., 2022).

Public services use AI for resource allocation, service delivery optimisation, and citizen engagement (Kotyal et al., 2024; Saghiri et al., 2022). Smart city applications use AI for traffic management, energy optimisation, and public safety (Dulaj et al., 2025; Kotyal et al., 2024). Government agencies deploy AI for fraud detection, eligibility determination, and case management (Kotyal et al., 2024; Saghiri et al., 2022). Natural language processing enables automated processing of citizen enquiries and documents (Saghiri et al., 2022).

Educational AI faces challenges involving equitable access, student privacy, and the risk of reinforcing educational inequities (Karanam et al., 2025; Saghiri et al., 2022). Another concern is over-reliance on automated systems and the potential deskilling of educators (Saghiri et al., 2022). Public sector AI deployment must navigate complex regulatory environments, procurement processes, and public accountability requirements (Nanjundan et al., 2025; Saghiri et al., 2022). Bias in public-service AI can perpetuate discrimination and erode public trust (Karanam et al., 2025; Saghiri et al., 2022). Despite these challenges, education and public services remain important application domains for AI (Kotyal et al., 2024; Saghiri et al., 2022).

8.5. Cross-Domain Applications

Several AI applications span multiple domains, demonstrating the general-purpose nature of modern AI technologies. Natural language processing powers machine translation, content generation, sentiment analysis, and information extraction (Ahmed et al., 2023; Kotyal et al., 2024). Computer vision enables applications such as autonomous vehicles, surveillance systems, augmented reality, and accessibility tools (Ahmed et al., 2023; Kotyal et al., 2024). Speech recognition and synthesis support virtual assistants, transcription services, and accessibility applications (Ahmed et al., 2023; Kotyal et al., 2024).

Recommendation systems used by e-commerce, streaming services, and social media platforms employ collaborative filtering and deep learning to personalise content and product suggestions (Kotyal et al., 2024; Safitra et al., 2024). Autonomous systems, including drones, robots, and vehicles, integrate perception, planning, and control using AI techniques (Dulaj et al., 2025; Kotyal et al., 2024). Cybersecurity applications use AI for threat detection, vulnerability assessment, and automated responses (Dulaj et al., 2025; Kotyal et al., 2024). Climate and environmental applications use AI for weather prediction, climate modelling, and environmental monitoring (Kotyal et al., 2024; Zuo et al., 2025).

These cross-domain applications demonstrate both the versatility and the limitations of current AI technologies. Success in one domain does not guarantee transferability to another, and domain-specific customisation is typically required (Karanam et al., 2025; Mukhamediev et al., 2022). The diversity of applications creates challenges for developing general-purpose AI systems that work reliably across contexts (Bi et al., 2025; Lu et al., 2023). Nevertheless, the breadth of successful applications demonstrates the practical significance and broad applicability of AI technology (Kotyal et al., 2024; Sharma & Garg, 2021).

9. Capability-Constraint-Infrastructure (CCI) Framework

9.1. Conceptual Overview

A central limitation of many AI surveys is that they evaluate systems primarily through isolated performance outcomes such as accuracy, task success, benchmark leadership, or model size. Such measures are useful, but they do not fully explain whether an AI system can be deployed, maintained, audited, trusted, or adapted in real environments. Real-world AI effectiveness emerges from an interaction among what a system can do, what limits or risks accompany this performance, and what technical and organisational infrastructure is available to support it. The

Capability–Constraint–Infrastructure (CCI) framework is therefore formulated as an operational analytical model for comparing AI systems across technical, infrastructural, cognitive, clinical, ethical, and governance dimensions.

The three core components of the CCI framework are *Capability*, *Constraint*, and *Infrastructure*. These components are not independent descriptors; rather, they represent an interacting system through which AI performance becomes meaningful in deployment conditions. Capability describes what an AI system is able to do. Constraint describes the technical, cognitive, ethical, legal, and operational limits that restrict this performance. Infrastructure describes the computational, organisational, and governance environment that enables or prevents the system from functioning reliably. The value of the framework lies in treating these dimensions jointly. A model with high capability but excessive computational cost, limited explainability, high bias risk, or insufficient deployment support may remain unsuitable for practice. Conversely, a less complex model with balanced capability, manageable constraints, and adequate infrastructure may be more mature for real-world use.

Capability refers to the functional competence of an AI system in relation to a defined task and context. It includes conventional performance indicators such as accuracy, precision, recall, robustness, generalisation, reasoning ability, multimodal integration, adaptability, and capacity for human-AI interaction. In narrow applications, capability may be measured through benchmark scores or task-specific validation. In more complex domains, such as medicine, education, autonomous systems, or legal decision support, capability also includes the ability to transfer knowledge across settings, handle ambiguous inputs, maintain performance under distributional shift, and support decisions requiring contextual reasoning. Capability therefore captures not only whether a model performs well under controlled test conditions, but also whether its competence remains meaningful when exposed to uncertain and variable environments.

Constraint refers to the factors that limit, degrade, or complicate the use of an AI system despite its apparent capability. These constraints include computational burden, energy consumption, latency, memory demand, data dependence, vulnerability to distribution shift, catastrophic forgetting, limited causal understanding, hallucination, adversarial sensitivity, opacity, bias, privacy risks, and regulatory uncertainty. Constraints also include social and institutional barriers, such as low user trust, insufficient documentation, lack of contestability, and difficulty assigning accountability for automated decisions. In this sense, constraint burden functions as a counterweight to capability. A highly accurate model can still have low practical value if it is too resource-intensive, insufficiently interpretable, unsafe for vulnerable groups, or incompatible with legal and ethical requirements.

Infrastructure refers to the technical and organisational substrate required to train, deploy, monitor, and maintain an AI system. At the technical level, it includes data pipelines, cloud or edge computing resources, specialised hardware accelerators, storage capacity, network stability, cybersecurity mechanisms, model monitoring tools, and integration with existing software systems. At the organisational level, infrastructure includes maintenance expertise, governance procedures, audit mechanisms, clinical or industrial workflow integration, and post-deployment feedback loops. Infrastructure is therefore more than hardware availability. It determines whether a model can move from experimental validation to sustained operation. A system may demonstrate strong capability in a laboratory setting but fail in practice when deployed in environments with limited bandwidth, insufficient compute, fragmented data systems, or weak monitoring capacity.

To operationalise these ideas, the framework represents an AI system A deployed in context d through five normalised dimensions. The first three preserve the core CCI terminology: **capability** $C(A, d)$, **constraint burden** $K(A, d)$, and **infrastructure adequacy** $I(A, d)$. Deployment readiness is further decomposed into **operational fit** $O(A, d)$ and **societal-regulatory readiness** $S(A, d)$. Operational fit captures latency tolerance, reliability requirements, domain mismatch, workflow compatibility, maintainability, and human-AI coordination. Societal-regulatory readiness captures fairness, auditability, privacy, legal compliance, documentation quality, contestability, human oversight, and user trust. Overall deployment readiness may be expressed as:

$$D(A, d) = O(A, d) \times S(A, d). \quad (1)$$

Taken together, these dimensions provide a structured explanation of why benchmark performance alone is insufficient for evaluating AI maturity. Capability captures the potential system's, constraint burden identifies sources of friction and risk, and infrastructure adequacy defines the conditions under which the system can be supported. Operational fit and societal-regulatory readiness then determine whether the system can be responsibly embedded in a specific environment. The interaction among these dimensions is especially important for neuroscience-inspired and neuro-symbolic AI. Biological intelligence demonstrates high adaptive capability under severe energetic and structural constraints, supported by efficient neural infrastructure. By contrast, many contemporary AI systems increase capability through large-scale computation and data accumulation, often intensifying infrastructure dependence and constraint burden. The CCI framework therefore provides a bridge between engineering evaluation and cognitively informed assessment by asking whether artificial systems are not only powerful, but also efficient, robust, interpretable, and deployable.

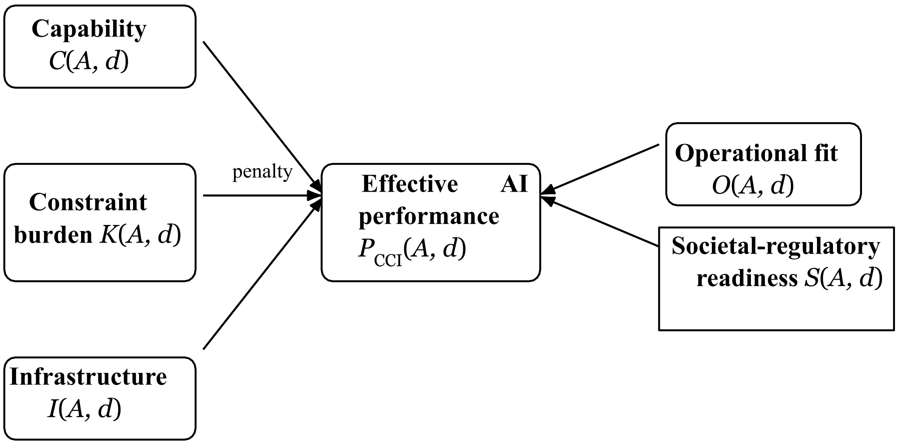

Figure 4. Simplified operational structure of the CCI framework. Capability, infrastructure adequacy, operational fit, and societal-regulatory readiness increase effective AI performance, while constraint burden functions as a penalty term

9.2. Operational Definitions

To make the framework analytically usable, each dimension is represented through normalised indicators that can be adapted to the domain under study. Each component is scored on the interval [0, 1] using weighted indicators extracted from the literature, benchmark reports, deployment documentation, expert assessment, regulatory requirements, or domain audits. For system A in deployment context d, the component scores are defined as follows:

$$C(A,d) = \sum_{r=1}^{m} \alpha_r c_r(A,d), \qquad K(A,d) = \sum_{j=1}^{n} \beta_j k_j(A,d), \quad (2)$$

$$I(A,d) = \sum_{\ell=1}^{p} \gamma_\ell i_\ell(A,d), \qquad O(A,d) = \sum_{u=1}^{q} \delta_u o_u(A,d), \quad (3)$$

$$S(A,d) = \sum_{v=1} \eta_v s_v(A,d). \quad (4)$$

The weights α_r, β_j, γ_ℓ, δ_u, η_v are non-negative and each weight family sums to one. Indicators c_r may represent accuracy, reasoning depth, robustness, transfer capability, calibration, multimodal integration, or adaptation. Indicators k_j may capture energy cost, memory demand, brittleness, security risk, bias, hallucinations, privacy exposure, or explainability deficits. Indicators i_ℓ may represent accelerator availability, storage capacity, network dependence, edge/cloud compatibility, data pipeline reliability, and monitoring infrastructure. Indicators o_u may capture latency, reliability, workflow fit, maintainability, domain validation, and human–AI coordination. Indicators s_v may capture privacy, fairness, auditability, documentation quality, contestability, human oversight, regulatory compliance, and legal accountability. The societal-regulatory component $S(A, d)$ prevents ethical, legal, and social factors from being treated as external commentary; instead, fairness, privacy, auditability, contestability, documentation, accountability, and human oversight become formal determinants of effective AI performance.

Indicators may be normalised through min–max scaling, percentile ranking, benchmark transformation, expert scoring rubrics, or domain-specific thresholds when direct quantitative measurement is not available. For example, latency may be normalised relative to the maximum acceptable latency in a clinical or industrial workflow, while explainability may be scored through a structured audit of interpretability, documentation, uncertainty communication, and human contestability. This allows the CCI framework to combine quantitative performance evidence with qualitative deployment evidence without reducing all dimensions to accuracy.

9.3. Effective AI Performance Index

The normalised CCI variables may be combined into an effective AI performance index. Rather than treating performance as a simple benchmark outcome, the index estimates the degree to which a system combines task capability, infrastructural support, operational fit, and societal-regulatory readiness while remaining limited by its constraint burden. The index is defined as:

$$P_{\mathrm{CCI}}(A,d) = C(A,d)^{\lambda_C} I(A,d)^{\lambda_I} O(A,d)^{\lambda_O} S(A,d)^{\lambda_S} \left[1 - K(A,d)\right]^{\lambda_K} \tag{5}$$

where $\lambda_C, \lambda_I, \lambda_O, \lambda_S, \lambda_K \geq 0$ and $\lambda_C + \lambda_I + \lambda_O + \lambda_S + \lambda_K = 1$. The parameters λ express the relative importance of each dimension in a specific deployment context. For example, medical and legal applications may assign larger weights to societal-regulatory readiness and constraint burden, whereas embedded vision systems may assign larger weights to infrastructure adequacy, latency, energy, and memory constraints.

The term $1 - K(A, d)$ converts constraint burden into constraint acceptability. When the constraint burden is low, $1 - K(A, d)$ is close to one and the penalty is limited. When the constraint burden is high, $1 - K(A, d)$ approaches zero and the overall score decreases substantially. This is appropriate for safety-critical or resource-constrained environments, where severe opacity, instability, energy demand, bias, or regulatory risk can prevent deployment even when benchmark capability is high.

This formulation has four important mathematical properties. First, it is bounded in the interval $[0, 1]$. Second, it is monotonic: increases in capability, infrastructure adequacy, operational fit, and societal-regulatory readiness increase the index, while increases in constraint burden decrease it. Third, it is only partially compensatory, meaning that very poor performance in one dimension cannot be fully hidden by strong performance in another. Fourth, the exponents provide an interpretable weighting mechanism, allowing the framework to adapt to different domains without changing its general structure.

For interior values $C, I, O, S, 1 - K \in (0, 1]$, the partial derivatives are:

$$\frac{\partial P_{\mathrm{CCI}}}{\partial C} = \lambda_C \frac{P_{\mathrm{CCI}}}{C}, \qquad \frac{\partial P_{\mathrm{CCI}}}{\partial I} = \lambda_I \frac{P_{\mathrm{CCI}}}{I}, \tag{6}$$

$$\frac{\partial P_{\mathrm{CCI}}}{\partial O} = \lambda_O \frac{P_{\mathrm{CCI}}}{O}, \qquad \frac{\partial P_{\mathrm{CCI}}}{\partial S} = \lambda_S \frac{P_{\mathrm{CCI}}}{S}, \tag{7}$$

$$\frac{\partial P_{\mathrm{CCI}}}{\partial K} = -\lambda_K \frac{P_{\mathrm{CCI}}}{1-K}. \tag{8}$$

These derivatives show that the framework behaves as expected: stronger capability, infrastructure, operational fit, and societal-regulatory readiness improve effective performance, whereas higher constraint burden reduces it. The model therefore formalises the core claim of the CCI framework: AI maturity depends not only on what a model can do, but also on the cost, risk, infrastructure, and institutional conditions under which it performs.

9.4. Optimisation Perspective

The CCI framework can also be interpreted as a constrained optimisation problem. Let θ denote a candidate AI system configuration, including model architecture, training regime, compression strategy, deployment infrastructure, monitoring procedure, and governance mechanism. For a deployment context *d*, the design problem can be expressed as:

$$\max_{\theta \in \Theta_d} J(\theta, d) = P_{\mathrm{CCI}}(\theta, d), \tag{9}$$

subject to domain-specific feasibility constraints:

$$L(\theta, d) \leq \tau, \qquad M(\theta, d) \leq \mu, \quad E(\theta, d) \leq \varepsilon, \qquad R(\theta, d) \leq \rho. \tag{10}$$

Here, L, M , E, and R represent latency, memory demand, energy consumption, and residual risk. The corresponding thresholds τ , μ, ε, and ρ define the maximum permissible values for the target domain. The feasible set Θ_d contains all configurations that satisfy the minimum technical, operational, and regulatory requirements of deployment context *d*.

This optimisation perspective explains why the same AI model may have different maturity levels across contexts. A large multimodal model may be feasible in a cloud-based enterprise environment but infeasible in an edge-computing or clinical emergency setting because the latency, energy, memory, or risk constraints differ. Similarly, a medical AI system may require stronger constraints on explainability, auditability, privacy, and human oversight than a low-risk consumer recommendation system. The CCI framework therefore evaluates AI systems as context-dependent configurations rather than as isolated algorithms.

9.5. Capability–Constraint Trade-offs

A key insight of the framework is that AI progress should be interpreted through trade-offs rather than through linear narratives of scale and improvement.

1) *Capability versus Computational Burden:* Foundation models expand linguistic, multimodal, and generative capability, but they also intensify hardware demand, energy consumption, and concentration of compute. From a neuroscience perspective, this is precisely where current AI diverges from biological intelligence: brains achieve adaptive behaviour under tight energetic budgets, whereas many state-of-the-art models improve by scaling parameters and data (Friston, 2010; Hassabis et al., 2017). The CCI framework makes this divergence explicit by separating capability from the infrastructure and constraint conditions that make capability sustainable.

2) *Reasoning versus Statistical Pattern Matching:* Neuro-symbolic approaches are especially important in this trade-off space because they attempt to increase reasoning-oriented capability without relying exclusively on brute-force scaling (Bhuyan et al., 2024; Liang et al., 2025). These

systems combine the pattern-recognition strengths of neural networks with symbolic structures that support compositionality, rule-based inference, knowledge representation, and explicit reasoning. Although neuro-symbolic systems remain imperfect, they offer a principled path towards models that are more auditable, sample efficient, and compatible with cognitively informed accounts of intelligence.

3) *Accuracy versus Explainability and Social Readiness:* High benchmark accuracy may coexist with low societal readiness when a model is opaque, difficult to audit, legally uncertain, or difficult for professionals to contest. Treating explainability, privacy, fairness, and accountability as part of *S* prevents the framework from overstating the maturity of systems that are technically impressive but institutionally fragile. This issue is central in medical AI: a black-box imaging model can produce accurate predictions while still failing clinical adoption because physicians cannot inspect the reasoning pathway, validate subgroup behaviour, explain the basis of a recommendation to patients, or document why the model output should influence care (Rudin, 2019; Tiwari et al., 2025; Toto et al., 2025).

9.6. Illustrative Medical Applications

To demonstrate practical use, Table 3 applies the CCI logic to two representative medical AI scenarios. The values are illustrative rather than empirical measurements. Their purpose is to show how the framework separates benchmark capability from deployable maturity and how the same mathematical structure can be adapted to different clinical environments.

Table 3. Illustrative CCI scoring for two medical AI systems using equal weights $\lambda_C = \lambda_I = \lambda_O = \lambda_S = \lambda_K = 0.20$

CCI component	**Medical imaging AI**	**Clinical early-warning AI**	**Interpretation**
Capability *C*	0.82	0.76	Imaging models may achieve high performance on curated datasets; early-warning systems often face noisier temporal and multimodal clinical data.
Constraint burden *K*	0.45	0.52	Both systems face opacity, subgroup-bias risk, validation burden, and liability concerns; early-warning systems may also suffer from alert fatigue and temporal instability.
Infrastructure *I*	0.70	0.62	Imaging AI can rely on hospital imaging systems and local or cloud compute; early-warning AI requires continuous integration with electronic health records and monitoring devices.
Operational fit *O*	0.65	0.58	Imaging AI can support radiology workflows if integrated with reporting systems; early-warning AI must fit urgent clinical routines without increasing false alarms.
Societal-regulatory readiness *S*	0.55	0.50	Both require documentation, audit trails, privacy safeguards, human oversight, and post-deployment monitoring; readiness is lower when accountability pathways remain unclear.
*P*CCI	0.65	0.58	The imaging case shows stronger technical and infrastructural maturity but remains limited by societal-regulatory readiness. The early-warning case is functional but requires improved governance, validation, and workflow integration.

For the medical imaging system, the CCI index is computed as:

$$P_{\mathrm{CCI}} = (0.82)^{0.20}(0.70)^{0.20}(0.65)^{0.20}(0.55)^{0.20}(1-0.45)^{0.20} \approx 0.65. \quad (11)$$

For the clinical early-warning system, the index is:

$$P_{\mathrm{CCI}} = (0.76)^{0.20}(0.62)^{0.20}(0.58)^{0.20}(0.50)^{0.20}(1-0.52)^{0.20} \approx 0.58. \quad (12)$$

These examples illustrate why medical AI cannot be evaluated by accuracy alone. A model can be technically strong while remaining immature if clinical validation, explainability, workflow integration, and compliance mechanisms are incomplete. The medical imaging example has relatively high capability and infrastructure adequacy, but its societal-regulatory readiness remains a limiting factor. The early-warning example shows a different maturity problem: the system may be useful, but alert fatigue, temporal instability, incomplete documentation, and unclear accountability can prevent safe adoption. The framework therefore provides a structured vocabulary for deciding whether improvement should focus on model performance, infrastructure, interpretability, governance, or operational integration.

9.7. CCI Maturity Model

The four maturity levels provide a simple and interpretable progression from laboratory capability to sustained deployment. The four-level structure corresponds to the main transition points identified in the reviewed literature: experimental model development, functional task-specific use, governed deployment readiness, and adaptive post-deployment resilience. The levels are tied to explicit score ranges and minimum conditions in Table 4. The thresholds are proposed as an operational rubric for comparative analysis rather than as a fully validated measurement scale. They are grounded in the qualitative synthesis of the reviewed literature: systems with high capability but high constraint burden remain experimental, systems with adequate task performance but incomplete deployment safeguards are functional, systems with balanced capability, infrastructure, and readiness are deployment-ready, and systems with monitoring, adaptation safeguards, and sustained compliance are adaptive and resilient.

Table 4. Operational CCI maturity levels of AI systems

Level	Name	Indicative threshold logic	Typical example
1	Experimental	$P_{\mathrm{CCI}} < 0.35$ or $K > 0.70$; deployment mainly laboratory-bound	Frontier prototypes or fragile large models
2	Functional	$0.35 \leq P_{\mathrm{CCI}} < 0.55$ and $O \geq 0.50$, or $P_{\mathrm{CCI}} \geq 0.55$ with incomplete societal-regulatory readiness	Domain-specific AI with limited assurance
3	Deployment-ready	$0.55 \leq P_{\mathrm{CCI}} < 0.75$, $I \geq 0.60$, $S \geq 0.60$, and $K \leq 0.50$	Controlled production systems with governance and monitoring
4	Adaptive / resilient	$P_{\mathrm{CCI}} \geq 0.75$ with monitoring, continual-learning safeguards, sustained compliance, and post-deployment feedback loops	Resource-aware, self-monitoring, continually governed AI

The threshold logic is intentionally conjunctive rather than purely score-based. This prevents a system from being classified as deployment-ready solely because it has a high aggregate value. For instance, an AI system with strong capability and infrastructure but weak societal-regulatory readiness should not be considered fully deployment-ready in medicine, law, public administration, or other high-stakes domains. The maturity model therefore treats the CCI score as a summary indicator while preserving minimum safeguards for infrastructure, governance, and risk.

9.8. Scope and Future Empirical Validation

The CCI framework is proposed here as a conceptual and operational evaluation structure derived from the reviewed literature. Its purpose is to organise the relationship among five dimensions: capability, constraint burden, infrastructure adequacy, operational fit, and societal-regulatory readiness. The framework does not claim that the proposed weights, indicators, or maturity thresholds are final universal constants. Rather, it provides a transparent structure that can be calibrated for different domains.

Future work should test the framework across real applications, including medical imaging, clinical early-warning systems, neuro-symbolic decision support, edge-based AI, educational AI, autonomous systems, and legal or administrative decision-support tools. Empirical validation could compare CCI scores with observed deployment outcomes such as clinical adoption, failure rates, maintenance cost, regulatory approval, user trust, post-deployment drift, and adverse-event reporting. Such testing would allow the weighting parameters, thresholds, and indicator definitions to be refined using real-world evidence. In this sense, the present framework should be understood as a formalised foundation for comparative evaluation, while future studies can transform it into an empirically validated measurement instrument.

9.9. Implications for This Review

This operationalisation clarifies the contribution of the manuscript. The earlier sections of the review provide the evidence base; the CCI framework provides the comparative lens. Sections 4 and 6 primarily inform C, Section 5 informs I, Section 7 informs K and S, and Section 8 informs O and domain-specific deployment fit. In this form, the framework is not merely a relabeling of familiar industry ideas; it becomes a structured, review-driven rubric for analysing why some AI systems scale successfully while others fail outside benchmark settings. The framework also aligns AI evaluation with the broader concerns of neuroscience-inspired intelligence: efficient adaptation, robustness under constraints, interpretable reasoning, and sustainable interaction with complex environments.

10. Discussion

10.1. Synthesis of Key Findings

Across the 229-study corpus, the most consistent pattern is not unlimited progress but *asymmetric progress*. Contemporary AI has advanced rapidly in perception, generation, language modelling, multimodal processing, and narrow-task optimisation, yet these gains remain accompanied by persistent weaknesses in reasoning, continual adaptation, interpretability, robustness, and real-world deployment. The CCI formulation makes this asymmetry visible: many systems achieve high capability C in benchmark settings, but their effective performance remains moderated by high constraint burden K, uneven infrastructure adequacy I, incomplete operational fit O, or insufficient societal-regulatory readiness S.

This finding is important because the dominant evaluation culture in AI remains structurally biased towards what is easiest to measure. Benchmarks reward short-horizon task success, leaderboard gains, and average predictive performance. They rarely reward auditability, long-term reliability, energy efficiency, human contestability, regulatory readiness, or resilience under distributional shift. As a result, resource-intensive models can appear more mature when judged by capability alone, even though they may remain difficult to deploy in regulated, privacy-sensitive, safety-critical, or low-resource environments.

The review therefore supports the development of evaluation frameworks that move beyond accuracy-centred assessment. The proposed CCI framework responds to this need by combining five dimensions: capability, constraint burden, infrastructure adequacy, operational fit, and societal-regulatory readiness. These dimensions make explicit the difference between a model that performs well under controlled conditions and a system that can be responsibly embedded in clinical, educational, industrial, administrative, or social environments. In this sense, the framework reframes AI maturity as a multidimensional property rather than a single performance score.

10.2. Why the Gaps Persist Despite Investment

The persistence of major AI limitations despite large-scale investment can be explained by three interacting factors. First, incentive structures in AI research still privilege leaderboard performance, novelty, and scale over lifecycle reliability, interpretability, maintainability, and deployment safety. Second, compute, data, and specialised expertise remain concentrated in a small number of institutions, which narrows the range of systems that can realistically be developed, reproduced, audited, and independently scrutinised. Third, many of the hardest problems in AI are not single-algorithm problems. Causal reasoning, continual learning, mechanistic transparency, social trust, and responsible deployment require integration across algorithms, hardware, human factors, organisational workflow, legal design, and regulatory oversight.

A neuroscience-facing perspective is particularly useful because biological intelligence suggests that efficiency, modularity, memory consolidation, adaptive control, and context-sensitive learning are not secondary features; they are central organising principles. Human cognition does not depend solely on scale. It also depends on efficient representation, selective attention, embodied feedback, hierarchical control, memory stabilisation, and adaptation under severe energetic constraints. From that standpoint, the dominant scale-first trajectory of current AI is productive but incomplete. Neuro-symbolic architectures, cognitive architectures, neuromorphic computing, sparse computation, and energy-aware learning should therefore be understood as serious responses to structural limits in current deep learning, not as peripheral alternatives.

The CCI framework clarifies this point. A system may improve its apparent capability by increasing model size, training data, or computational intensity, but such improvement may simultaneously increase constraint burden and infrastructure dependence. In contrast, a more mature trajectory would increase capability while reducing unnecessary constraints, improving infrastructure compatibility, strengthening operational fit, and satisfying governance expectations. This distinction is especially relevant for medical AI, where deployment depends not only on model accuracy but also on explainability, workflow integration, privacy, auditability, patient safety, and professional accountability.

10.3. Philosophical Implications of the Reasoning and Generalisation Gap

The reasoning and generalisation gap has philosophical significance because it challenges the assumption that predictive success is equivalent to understanding. A system may generate correct answers across many benchmark instances while lacking stable concepts, causal models, grounded representations, or reflective self-monitoring capacities that support reliable transfer to new situations. This distinction echoes a long-standing tension between behavioural performance and explanatory competence: an AI system can appear intelligent in narrow contexts without possessing the flexible, compositional, and adaptive understanding associated with human cognition (Hassabis et al., 2017; Lake et al., 2017).

This issue also shapes how AI progress should be interpreted. If intelligence is defined only by task accuracy, then scaling data, parameters, and computation may appear sufficient. If intelligence includes abstraction, causal reasoning, memory consolidation, uncertainty management, social understanding, and responsible action under changing conditions, then accuracy is only one component of a broader maturity profile. The CCI framework adopts the second interpretation. It treats reasoning, interpretability, infrastructure, operational fit, and societal readiness as interacting conditions for effective intelligence rather than as secondary attributes added after model training.

This interpretation is also particularly important for high-stakes domains. In medicine, for example, a diagnostic system may correctly classify images while still failing to provide clinically useful reasoning, uncertainty estimates, subgroup reliability, or explanations that can be communicated to patients. In law or public administration, a system may predict outcomes while remaining difficult to contest, audit, or justify. These examples show why AI evaluation must distinguish between statistical performance and accountable intelligence. The CCI framework contributes to this distinction by embedding capability within a wider structure of constraints, infrastructure, and governance.

10.4. Research Opportunities

The most promising research opportunities follow directly from this diagnosis. First, future work should develop AI systems that improve reasoning and transfer without assuming that progress must always depend on ever-larger models. This includes neuro-symbolic systems, causal representation learning, modular architectures, retrieval-augmented reasoning, memory-based approaches, and architectures inspired by cognitive and neural efficiency. Second, continual learning must be treated as a deployment requirement, particularly in domains with evolving data streams, changing user behaviour, shifting clinical populations, or dynamic operational environments. Third, interpretability research should move beyond retrospective explanation towards internal, mechanistic understanding, uncertainty communication, and human-contestable decision pathways. Fourth, evaluation should incorporate social and infrastructural realism. Privacy, fairness, auditability, energy consumption, latency, maintainability, workflow compatibility, legal accountability, and human oversight are not externalities; they are determinants of whether AI creates durable value. A model that performs well in isolation may fail in practice if it cannot be monitored, updated, audited, explained, or integrated into existing workflows. Conversely, a model with moderate technical performance may be more deployable if it is transparent, robust, resource-efficient, well-integrated, and supported by clear governance procedures.

Medical AI illustrates this point especially clearly. Clinical systems must satisfy not only technical performance requirements but also patient-safety expectations, documentation standards, human oversight procedures, privacy safeguards, subgroup validation, and post-deployment monitoring, consistent with emerging reporting and protocol standards for AI-based clinical interventions (Liu et al., 2020; Rivera et al., 2020). The explainability-versus-accuracy trade-off is therefore not merely a technical issue. It is a deployment issue, a trust issue, and a governance issue. For high-stakes applications, accuracy gains obtained through opaque architectures may be insufficient if clinicians cannot evaluate model reasoning, understand uncertainty, identify failure modes, or explain decisions to patients and institutions (Rudin, 2019).

10.5. Limitations and Future Validation of the CCI Framework

Although the CCI framework provides a structured way to evaluate AI systems across capability, constraint burden, infrastructure adequacy, operational fit, and societal-regulatory readiness, it should be interpreted as an initial analytical framework rather than as a fully validated measurement instrument. The proposed dimensions, mathematical index, and maturity thresholds are intended to make the evaluation of AI systems more explicit and comparable, but their empirical calibration requires additional domain-specific testing.

This limitation is important because no single set of weights or thresholds can be assumed to apply equally across all AI domains. A clinical early-warning system, a medical imaging model, an autonomous vehicle, a legal decision-support tool, an educational recommender, and an edge-based industrial monitoring system operate under different risk profiles, latency requirements, regulatory expectations, infrastructure conditions, and human oversight needs. Therefore, the CCI score should not be interpreted as a universal absolute measure. It is better understood as a transparent and adaptable structure for comparing AI maturity within a defined deployment context.

Future work should apply the framework to real AI deployments across multiple domains, including medical imaging, clinical early-warning systems, neuro-symbolic decision support, educational AI, autonomous systems, edge-based AI, and legal or administrative decision-support tools. Such studies could compare CCI scores with observed deployment outcomes, including clinical adoption, model drift, maintenance cost, failure rates, regulatory approval, user trust, audit results, energy consumption, post-deployment safety incidents, and long-term sustainability. This would allow the indicator weights, maturity thresholds, and scoring rubrics to be refined using empirical evidence rather than relying only on conceptual synthesis.

This limitation does not weaken the purpose of the framework. Instead, it defines its current scope. The contribution of the present study is to provide a coherent structure for evaluating AI maturity beyond benchmark accuracy. The next step is to test whether the proposed dimensions predict real-world deployment success, sustainability, safety, and governance quality across different application contexts. In this sense, the CCI framework should be viewed as a formalised foundation for comparative evaluation, while future empirical studies can transform it into a calibrated measurement instrument.

11. Conclusion

This comprehensive study of current AI technologies reveals a field characterised by remarkable technical achievements as well as significant limitations and unresolved challenges. Deep learning architectures, foundation models, multimodal systems, and generative AI have achieved substantial performance gains in pattern recognition, language processing, image analysis, decision support, and content generation. These advances have enabled important practical applications across healthcare, finance, manufacturing, education, scientific discovery, and numerous other domains. However, fundamental constraints on reasoning, generalisation, explainability, robustness, energy consumption, infrastructure dependence, and responsible deployment continue to limit the reliability and applicability of current AI systems.

The analysis identified a set of interdependent challenges that must be addressed for AI to realise its full potential: technical capability, constraint burden, infrastructure adequacy, operational fit, and societal-regulatory readiness. Technical advances are still required in efficient reasoning, causal abstraction, continual learning, explainability, robustness, and resource-aware computation. Current approaches that achieve strong performance in narrow domains often fail when confronted with novel situations, require substantial computational resources, or lack the transparency necessary for high-stakes applications. Infrastructure development must also address the tension between centralised cloud computing and distributed edge deployment, balancing capability with accessibility, cost, latency, privacy, and environmental sustainability. Practical deployment challenges, including data quality, bias, workflow integration, regulation, user trust, and workforce development, require not only technical solutions but also organisational, legal, policy, and societal changes.

This review identifies several promising research directions, including neuro-symbolic AI, neuroscience-inspired architectures, foundation models, multimodal systems, human-AI collaboration, emerging hardware technologies, responsible AI governance, and resource-efficient deployment. Realising these directions requires sustained investment in fundamental research, interdisciplinary collaboration, and inclusive stakeholder engagement. The future of AI depends not only on technical advances but also on the development of governance frameworks, regulatory standards, audit mechanisms, and societal norms that ensure AI systems are safe, fair, transparent, contestable, and aligned with human values. In that sense, legal and ethical design are not external to AI maturity; they are part of the conditions that determine whether a system can be responsibly deployed, monitored, audited, and sustained in practice.

This study also introduced the Capability–Constraint–Infrastructure (CCI) framework as an operational analytical model for understanding AI systems beyond benchmark performance alone. By explicitly linking capability, constraint burden, infrastructure adequacy, operational fit, and societal-regulatory readiness, the framework provides a structured way to evaluate whether an AI system is not only technically powerful but also deployable, sustainable, interpretable, and governable. The accompanying maturity model clarifies how AI systems may progress from experimental capability to functional use, deployment readiness, and adaptive resilience.

The CCI framework should be understood as a formalised foundation for comparative evaluation rather than as a final universal scoring system. Its dimensions and mathematical structure provide a transparent basis for analysis, while future empirical work should test and calibrate the proposed indicators, weights, and maturity thresholds across real applications. Particularly important test domains include medical imaging, clinical early-warning systems, neuro-symbolic decision support, autonomous systems, educational AI, edge-based AI, and legal or administrative decision support. Such validation would help determine whether CCI scores correspond to observed outcomes such as reliability, adoption, safety, regulatory acceptance, maintenance cost, energy efficiency, user trust, and post-deployment robustness.

For researchers, this review highlights critical gaps in reasoning, explainability, robustness, efficiency, and deployment-aware evaluation. For practitioners, it provides a realistic assessment of current capabilities and limitations, supporting more informed decisions about adoption, monitoring, integration, and risk management. For policymakers, it identifies key challenges in fairness, accountability, transparency, privacy, environmental impact, and legal responsibility that require regulatory attention. As AI continues to evolve and permeate society, maintaining a comprehensive and critical understanding of its capabilities, constraints, infrastructural dependencies, and societal implications remains essential. Future AI progress should therefore be judged not only by what systems can achieve in controlled benchmarks, but also by whether they can operate safely, efficiently, transparently, and responsibly in the environments where their decisions affect human lives.

References

Abbasi, M., Nishat, R. I., Bond, C., Graham-Knight, J. B., Lasserre, P., Lucet, Y., & Najjaran, H. (2025). A review of AI and machine learning contribution in business process management (process enhancement and process improvement approaches). *Business Process Management Journal, 31*(4), 1414–1452. https://doi.org/10.1108/BPMJ-07-2024-0555

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Ali, A. B. M. S., & Gandomi, A. H. (2023). Deep learning modelling techniques: Current progress, applications, advantages, and challenges. *Artificial Intelligence Review, 56*(11), 13521–13617. https://doi.org/10.1007/s10462-023-10466-8

Bhuyan, B. P., Ramdane-Cherif, A., Tomar, R., & Singh, T. (2024). Neuro-symbolic artificial intelligence: A survey. *Neural Computing and Applications, 36*(21), 12809–12844. https://doi.org/10.1007/s00521-024-09960-z

Bi, J., Wang, Z., Yuan, H., Shi, X., Wang, Z., Zhang, J., Zhou, M., & Buyya, R. (2025). Large AI models and their applications: Classification, limitations, and potential solutions. *Software: Practice and Experience, 55*(6), 1003–1017. https://doi.org/10.1002/spe.3408

Bunea, A. A., Popovici, D.-M., & Dobre, C. M. (2025). EfficientNet-enhanced U-Net for multi-organ thorax segmentation. In *Proceedings of the 2025 25th International Conference on Control Systems and Computer Science (CSCS)* (pp. 290–297). IEEE. https://doi.org/10.1109/CSCS66924.2025.00050

Confalonieri, R., Coba, L., Wagner, B., & Besold, T. R. (2021). A historical perspective of explainable artificial intelligence. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11*(1), e1391. https://doi.org/10.1002/widm.1391

Dulaj, K., Alhammadi, A., Shayea, I., El-Saleh, A. A., & Alnakhli, M. (2025). Harnessing machine learning for intelligent networking in 5G technology and beyond: Advancements, applications and challenges. *IEEE Open Journal of Intelligent Transportation Systems*, 605–633. https://doi.org/10.1109/OJITS.2025.3564361

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138. https://doi.org/10.1038/nrn2787

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron, 95*(2), 245–258. https://doi.org/10.1016/j.neuron.2017.06.011

He, M., Li, Z., Liu, C., Shi, D., & Tan, Z. (2020). Deployment of artificial intelligence in real-world practice: Opportunity and challenge. *Asia-Pacific Journal of Ophthalmology, 9*(4), 299–307. https://doi.org/10.1097/APO.0000000000000301

Heigl, R. (2025). Generative artificial intelligence in creative contexts: A systematic review and future research agenda. *Management Review Quarterly, 76*, 955–992. https://doi.org/10.1007/s11301-025-00494-9

Karanam, S. R., Kumar, A. P., Yandrapati, P. B., Goud, B. P., Vijaykumar, S., & Swamy, I. M. K. (2025). Advancements and applications of artificial intelligence and machine learning: A comprehensive review. *Advancements in Artificial Intelligence and Machine Learning*, 25–47. https://doi.org/10.2174/97898153225831250101

Kota, S. K. (2025). The evolution of AI agents: From rule-based systems to autonomous intelligence—A comprehensive review. *Journal of Artificial Intelligence & Cloud Computing, 4*(2), 1–5. https://doi.org/10.47363/JAICC/2025(4)433

Kotseruba, I., & Tsotsos, J. K. (2020). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review, 53*, 17–94. https://doi.org/10.1007/s10462-018-9646-y

Kotyal, K., Nautiyal, P., Singh, M., Semwal, A., Rai, D., Papnai, G., Nautiyal, C. T., Malathi, G., & Krishnaveni, S. A. (2024). Advancements and challenges in artificial intelligence applications: A comprehensive review. *Journal of Scientific Research and Reports, 30*(10), 375–385. https://doi.org/10.9734/jsrr/2024/v30i102465

Kumar, V., & Roy, K. (2025). Embracing the changes and challenges with modern early drug discovery. *Expert Opinion on Drug Discovery, 20*(4), 419–431. https://doi.org/10.1080/17460441.2025.2481259

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences, 40*, e253. https://doi.org/10.1017/S0140525X16001837

Lappin, S. (2025). The deep learning revolution in AI. *European Review, 33*(4), 416–425. https://doi.org/10.1017/S1062798725100379

Liang, B., Wang, Y., & Tong, C. (2025). AI reasoning in deep learning era: From symbolic AI to neural–symbolic AI. *Mathematics, 13*(11), 1707. https://doi.org/10.3390/math13111707

Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., & Denniston, A. K. (2020). Reporting guidelines

Lu, P., Qiu, L., Yu, W., Welleck, S., & Chang, K.-W. (2023). A survey of deep learning for mathematical reasoning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 14605–14631). https://doi.org/10.18653/v1/2023.acl-long.817

Medsker, L. R. (1994). *Hybrid neural network and expert systems*. Springer. https://doi.org/10.1007/978-1-4615-2726-8

Miya, M., Islam, M., Saleh, M. A., Mia, M. M. H., & Faruk, M. O. (2025). An overview of architectures, difficulties, and uses for artificial intelligence in real-time systems. *Scientia: Technology, Science and Society, 2*(9), 41–50. https://doi.org/10.59324/stss.2025.2(9).05

Mukhamediev, R. I., Popova, Y., Kuchin, Y., Zaitseva, E., Kalimoldayev, A., Symagulov, A., Levashenko, V., Abdoldina, F., Gopejenko, V., Yakunin, K., Muhamedijeva, E., & Yelis, M. (2022). Review of artificial intelligence and machine learning technologies: Classification, restrictions, opportunities and challenges. *Mathematics, 10*(15), 2552. https://doi.org/10.3390/math10152552

Mundlamuri, R., Gunnam, G. R., Mysari, N. K., & Pujuri, J. (2025). The evolution of AI: From classical machine learning to modern large language models. *IEEE Access, 13*, 178302–178341. https://doi.org/10.1109/ACCESS.2025.3621344

Nanjundan, P., Indu, P., & Thomas, L. (2025). Navigating the ethical landscape of artificial intelligence: Challenges, frameworks, and responsible deployment. In *Artificial intelligence technologies for engineering applications* (pp. 1–13). CRC Press.

National Institute of Standards and Technology. (2023). *AI risk management framework (AI RMF 1.0)*. https://doi.org/10.6028/NIST.AI.100-1

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks, 113*, 54–71.

https://doi.org/10.1016/j.neunet.2019.01.012

Radanliev, P. (2025). Artificial intelligence: Reflecting on the past and looking towards the next paradigm shift. *Journal of Experimental & Theoretical Artificial Intelligence, 37*(7), 1045–1062. https://doi.org/10.1080/0952813X.2024.2323042

Rahmaniar, W., Haq, Q. M. ul, Iskandar, M. E., & Ma'arif, A. (2025). Practical applications of artificial intelligence: A sector-wise review of challenges and future directions. *IT Professional, 27*(4), 61–68. https://doi.org/10.1109/MITP.2025.3578694

Rivera, S. C., Liu, X., Chan, A.-W., Denniston, A. K., Calvert, M. J., & SPIRIT-AI and CONSORT-AI Working Group. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. *BMJ, 370*, m3210. https://doi.org/10.1136/bmj.m3210 for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *The Lancet Digital Health, 2*(10), e537–e548. https://doi.org/10.1016/S2589-7500(20)30218-1

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Safitra, M. F., Lubis, M., Kusumasari, T. F., & Putri, D. P. (2024). Advancements in artificial intelligence and data science: Models, applications, and challenges. *Procedia Computer Science, 234*, 381–388. https://doi.org/10.1016/j.procs.2024.03.018

Saghiri, A. M., Vahidipour, S. M., Jabbarpour, M. R., Sookhak, M., & Forestiero, A. (2022). A survey of artificial intelligence challenges: Analyzing the definitions, relationships, and evolutions. *Applied Sciences, 12*(8), 4054. https://doi.org/10.3390/app12084054

Sharma, A., Choudhary, S., Singh, D. K., & Purohit, H. (2025). Artificial intelligence—An overview. In *Convergence of artificial intelligence, machine learning, and the internet of things in Industry 4.0 applications* (pp. 1–26). Springer. https://doi.org/10.1007/978-981-96-5269-3_1

Sharma, L., & Garg, P. K. (2021). *Artificial intelligence: Technologies, applications, and challenges*.

Shankar, V. (2024). Edge AI: A comprehensive survey of technologies, applications, and challenges. In *Proceedings of the 2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)* (pp. 1–6). IEEE. https://doi.org/10.1109/ACET61898.2024.10730112

Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2025). Generative artificial intelligence: A systematic review and applications. *Multimedia Tools and Applications, 84*(21), 23661–23700. https://doi.org/10.1007/s11042-024-20016-1

Tiwari, A., Mishra, S., & Kuo, T.-R. (2025). Current AI technologies in cancer diagnostics and treatment. *Molecular Cancer, 24*(1), 159. https://doi.org/10.1186/s12943-025-02369-9

Toma, A., Ciobanu, R.-I., Dobre, C., & Răducan, G. (2025). Federated learning for sustainable smart mobility: Enhancing privacy and reducing cloud dependency. In *Proceedings of the 2025 25th International Conference on Control Systems and Computer Science (CSCS)* (pp. 460–467). IEEE. https://doi.org/10.1109/CSCS66924.2025.00074

Toto, G. A., Grilli, L., Traetta, L., Villani, R., Petito, A., & Serviddio, G. (2025). Convergence of disciplines: A systematic review of multidisciplinary development approaches in artificial intelligence. *Frontiers in Digital Health, 7*, 1400338. https://doi.org/10.3389/fdgth.2025.1400338

Ye, S. (2025). From Turing models to large language models: Evolution and convergence of symbolic and connectionist approaches in artificial intelligence. *Highlights in Science, Engineering and Technology, 128*, 264–272. https://doi.org/10.54097/m92y8a35

Zuo, W., Zheng, H., He, T., Vishwanath, V., Chan, M. K. Y., Stevens, R. L., Amine, K., & Xu, G.-L. (2025). Large language models for batteries. *Joule, 9*(8).