The Complete Treatise on Artificial Intelligence: A Multidisciplinary Analysis

Soumadeep Ghosh

Kolkata, India

Abstract

This treatise presents a comprehensive examination of artificial intelligence through the lens of multiple academic disciplines. We explore the theoretical foundations, mathematical frameworks, computational architectures, and philosophical implications of AI systems. The analysis encompasses machine learning algorithms, neural networks, cognitive science principles, ethical considerations, and practical applications across various domains. This work synthesizes knowledge from computer science, mathematics, neuroscience, psychology, philosophy, and engineering to provide a complete understanding of artificial intelligence as both a technical discipline and a transformative force in human society.

The treatise ends with "The End"

Contents

1 Introduction

Artificial Intelligence represents one of the most significant intellectual achievements of the modern era, combining insights from mathematics, computer science, cognitive psychology, neuroscience, and philosophy. The field emerged from the confluence of several theoretical breakthroughs: Alan Turing's computational theory, Claude Shannon's information theory, and early cybernetic principles developed by Norbert Wiener.

The fundamental question driving AI research concerns the nature of intelligence itself. Can machines exhibit genuine intelligence, or do they merely simulate intelligent behavior through sophisticated computation? This inquiry necessitates understanding intelligence from multiple perspectives: computational processes, biological mechanisms, cognitive architectures, and philosophical concepts of mind and consciousness.

Contemporary AI systems demonstrate remarkable capabilities across diverse domains, from natural language processing and computer vision to strategic game playing and scientific discovery. These achievements stem from advances in machine learning, particularly deep learning architectures that can extract complex patterns from vast datasets. However, the theoretical foundations underlying these systems draw from centuries of mathematical and philosophical inquiry into the nature of knowledge, learning, and rational decision-making.

2 Mathematical Foundations

2.1 Probability Theory and Bayesian Inference

The mathematical foundation of modern AI rests heavily on probability theory and statistical inference. Bayesian probability provides a principled framework for reasoning under uncertainty, which is fundamental to intelligent behavior in real-world environments.

Given a hypothesis H and evidence E, Bayes' theorem states:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \tag{1}$$

This framework enables AI systems to update beliefs as new information becomes available, forming the basis for probabilistic reasoning in expert systems, Bayesian networks, and modern machine learning algorithms.

Bayesian networks represent complex probabilistic relationships through directed acyclic graphs, where nodes represent random variables and edges encode conditional dependencies. The joint probability distribution factorizes as:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\text{Parents}(X_i)) \tag{2}$$

2.2 Information Theory

Claude Shannon's information theory provides fundamental insights into the quantification and transmission of information. The entropy of a discrete random variable X with probability mass function $p(x)$ is:

$$H(X) = -\sum_{x} p(x) \log_2 p(x) \tag{3}$$

Information theory concepts such as mutual information, cross-entropy, and Kullback-Leibler divergence have become central to machine learning algorithms, particularly in optimizing neural network parameters and measuring model performance.

2.3 Linear Algebra and Optimization

Linear algebra forms the computational backbone of modern AI systems. Vector spaces, matrix operations, eigenvalue decomposition, and singular value decomposition enable efficient representation and manipulation of high-dimensional data.

Optimization theory provides methods for finding optimal parameters in machine learning models. The gradient descent algorithm iteratively updates parameters θ according to:

$$\theta_{t+1} = \theta_t - \alpha \nabla_\theta J(\theta_t) \tag{4}$$

where $J(\theta)$ represents the objective function and α is the learning rate.

3 Computational Architectures

3.1 Neural Networks and Deep Learning

Artificial neural networks, inspired by biological neural systems, represent the dominant paradigm in contemporary AI. A basic perceptron computes:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{5}$$

where w_i are weights, x_i are inputs, b is the bias term, and f is an activation function.

Deep neural networks extend this concept through multiple layers, enabling the learning of hierarchical representations. Backpropagation algorithm computes gradients efficiently through the chain rule:

$$\frac{\partial J}{\partial w_{ij}^{(l)}} = \frac{\partial J}{\partial z_j^{(l)}} \cdot a_i^{(l-1)} \tag{6}$$

Convolutional Neural Networks (CNNs) exploit spatial locality in data through shared parameters and local connectivity. Recurrent Neural Networks (RNNs) process sequential data by maintaining hidden states that capture temporal dependencies.

Transformer architectures have revolutionized natural language processing through self-attention mechanisms that model long-range dependencies:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{7}$$

3.2 Symbolic AI and Knowledge Representation

Symbolic AI approaches represent knowledge through formal logical systems. First-order logic provides a foundation for representing facts and relationships:

$$\forall x \, (P(x) \rightarrow Q(x)) \tag{8}$$

Expert systems encode domain-specific knowledge through production rules and inference engines. Semantic networks and ontologies structure knowledge through relationships between concepts, enabling automated reasoning and knowledge discovery.

3.3 Hybrid Architectures

Modern AI systems increasingly combine symbolic and connectionist approaches. Neuro-symbolic integration attempts to leverage the pattern recognition capabilities of neural networks with the interpretability and reasoning capabilities of symbolic systems.

4 Learning Algorithms and Paradigms

4.1 Supervised Learning

Supervised learning algorithms learn mappings from input features to target outputs using labeled training data. The empirical risk minimization principle seeks to minimize:

$$R_{emp}(f) = \frac{1}{n}\sum_{i=1}^{n} L(y_i, f(x_i)) \tag{9}$$

where L is a loss function measuring prediction error.

Support Vector Machines find optimal hyperplanes by maximizing margin:

$$\min_{w,b} \frac{1}{2}||w||^2 \quad \text{subject to} \quad y_i(w^T x_i + b) \geq 1 \tag{10}$$

Random forests aggregate multiple decision trees to improve generalization performance through ensemble methods.

4.2 Unsupervised Learning

Unsupervised learning discovers hidden patterns in data without explicit labels. Principal Component Analysis (PCA) reduces dimensionality by finding directions of maximum variance:

$$\max_{w} w^T \Sigma w \quad \text{subject to} \quad ||w||^2 = 1 \tag{11}$$

K-means clustering partitions data by minimizing within-cluster sum of squares:

$$\min_{\{S_i\}} \sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2 \tag{12}$$

4.3 Reinforcement Learning

Reinforcement learning addresses sequential decision-making problems through interaction with environments. The Bellman equation for optimal value functions states:

$$V^*(s) = \max_{a} \sum_{s'} P(s'|s,a)[R(s,a,s') + \gamma V^*(s')] \tag{13}$$

Q-learning algorithms learn optimal action-value functions through temporal difference updates:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)] \tag{14}$$

Policy gradient methods directly optimize parameterized policies through gradient ascent on expected rewards.

5 Cognitive Science Perspectives

5.1 Computational Theory of Mind

Cognitive science provides theoretical frameworks for understanding intelligence through computational processes. The computational theory of mind posits that mental states correspond to computational states, and cognitive processes involve symbol manipulation according to syntactic rules.

David Marr's tri-level hypothesis distinguishes between computational, algorithmic, and implementational levels of analysis. This framework guides AI research by emphasizing the importance of understanding what problems intelligence solves, how these problems are solved algorithmically, and how algorithms are implemented in physical systems.

5.2 Memory and Learning Mechanisms

Human memory systems provide insights for developing AI architectures. The distinction between working memory, long-term memory, and procedural memory informs the design of neural architectures with different memory components.

Attention mechanisms in cognitive psychology parallel attention mechanisms in neural networks, suggesting that selective processing of information is fundamental to intelligent behavior. The limited capacity of working memory motivates hierarchical processing and chunking strategies in AI systems.

5.3 Reasoning and Problem Solving

Cognitive research on human reasoning reveals systematic patterns in logical inference, analogical reasoning, and problem-solving strategies. These findings inform the development of AI systems that can reason effectively about novel situations.

Dual-process theories distinguish between fast, automatic (System 1) and slow, deliberative (System 2) reasoning processes. This framework suggests that effective AI systems might benefit from combining rapid pattern recognition with more deliberate analytical reasoning.

6 Neuroscience Foundations

6.1 Biological Neural Networks

Understanding biological neural networks provides inspiration and constraints for artificial neural architectures. Neurons integrate signals through dendritic trees, generate action potentials through threshold mechanisms, and transmit information through synaptic connections.

Synaptic plasticity mechanisms, including long-term potentiation and long-term depression, provide biological foundations for learning algorithms. Hebbian learning principles suggest that connections strengthen when neurons fire together, inspiring unsupervised learning rules in artificial networks.

6.2 Brain Architecture and Function

The hierarchical organization of the visual cortex, from simple cells detecting edges to complex cells responding to objects, influenced the development of convolutional neural networks. The modular organization of the brain suggests that effective AI systems might benefit from specialized components for different cognitive functions.

Neuroscience research on the prefrontal cortex reveals mechanisms for executive control, working memory, and abstract reasoning that inform the development of AI architectures for complex cognitive tasks. The role of the hippocampus in memory consolidation provides insights for developing AI systems with sophisticated memory mechanisms.

6.3 Embodied Cognition

Embodied cognition theories emphasize the role of sensorimotor experience in shaping cognitive processes. This perspective suggests that intelligence emerges from the interaction between agents and their environments, motivating research in robotics and embodied AI systems.

The concept of affordances, describing action possibilities in the environment, provides a framework for understanding how intelligent agents can perceive and act effectively in complex environments.

7 Philosophical Dimensions

7.1 The Nature of Intelligence and Consciousness

Philosophical inquiry into the nature of mind and consciousness raises fundamental questions about AI systems. The hard problem of consciousness, concerning the subjective experience of mental states, challenges reductive approaches to artificial intelligence.

Functionalist theories of mind suggest that mental states are defined by their causal relations rather than their physical implementation, supporting the possibility of machine consciousness. However, critics argue that computational processes cannot capture the qualitative aspects of conscious experience.

7.2 The Chinese Room Argument

John Searle's Chinese Room argument challenges the claim that computational processes can constitute understanding or consciousness. The argument suggests that symbol manipulation, however sophisticated, cannot generate genuine understanding without semantic content.

This philosophical challenge has motivated research into grounded AI systems that connect symbols to perceptual experience and embodied interaction with the environment. The debate continues to influence discussions about the nature of machine understanding and consciousness.

7.3 The Frame Problem

The frame problem, identified by philosophers and AI researchers, concerns the difficulty of representing what remains unchanged in dynamic environments. This problem highlights fundamental challenges in knowledge representation and reasoning about change.

Various solutions have been proposed, including non-monotonic reasoning systems, situation calculus, and event calculus. The frame problem continues to influence research in knowledge representation and automated reasoning.

8 Ethical Considerations and Societal Impact

8.1 AI Ethics and Moral Agency

The development of increasingly sophisticated AI systems raises important ethical questions about moral agency, responsibility, and the impact of autonomous systems on human welfare. The principle of beneficence requires that AI systems be designed to promote human well-being, while the principle of non-maleficence demands that they avoid causing harm.

Issues of fairness and bias in AI systems have received significant attention, particularly regarding discriminatory outcomes in hiring, lending, and criminal justice applications. Algorithmic fairness requires careful consideration of different notions of equity and their mathematical formulations.

The development of autonomous weapons systems raises profound questions about the ethics of delegating life-and-death decisions to machines. The principle of meaningful human control suggests that critical decisions should retain human oversight and responsibility.

8.2 Privacy and Surveillance

AI systems' capabilities for pattern recognition and data analysis raise significant privacy concerns. The ability to infer sensitive information from seemingly innocuous data challenges traditional notions of privacy and anonymity.

Differential privacy provides mathematical frameworks for protecting individual privacy while enabling statistical analysis. Techniques such as federated learning attempt to train AI models without centralizing sensitive data.

8.3 Economic and Social Implications

The automation capabilities of AI systems have significant implications for employment and economic inequality. While AI may create new economic opportunities, it may also displace workers in various sectors, requiring careful consideration of transition policies and social safety nets.

The concentration of AI capabilities in a few large organizations raises concerns about economic power and democratic governance. Ensuring broad access to AI benefits while managing potential risks requires thoughtful policy approaches.

9 Applications and Domains

9.1 Natural Language Processing

Natural language processing has achieved remarkable progress through transformer architectures and large language models. These systems demonstrate sophisticated capabilities in text generation, translation, summarization, and question answering.

However, challenges remain in achieving genuine language understanding, handling factual accuracy, and avoiding harmful or biased outputs. The development of more robust and reliable language models continues to be an active area of research.

9.2 Computer Vision

Computer vision systems have achieved human-level performance on many image recognition tasks through convolutional neural networks and attention mechanisms. Applications range from medical image analysis to autonomous vehicle perception.

Challenges in computer vision include robustness to adversarial examples, generalization across different domains, and understanding of three-dimensional scene structure. Multi-modal approaches that combine vision with language and other modalities show promising directions for future development.

9.3 Robotics and Embodied AI

Robotics represents the integration of AI with physical systems, enabling intelligent behavior in real-world environments. Challenges include perception under uncertainty, manipulation of complex objects, and navigation in dynamic environments.

Recent advances in deep reinforcement learning have enabled robots to learn complex manipulation tasks through trial and error. However, achieving the robustness and generalization capabilities of biological systems remains a significant challenge.

9.4 Scientific Discovery and Research

AI systems are increasingly being applied to accelerate scientific discovery across various disciplines. Applications include drug discovery, materials science, astronomy, and climate modeling.

Machine learning approaches can identify patterns in complex datasets, generate hypotheses, and even conduct autonomous experiments. However, ensuring the reliability and interpretability of AI-generated scientific insights remains crucial for scientific progress.

10 Current Challenges and Limitations

10.1 Generalization and Transfer Learning

One of the most significant challenges in AI is developing systems that can generalize effectively to new situations and domains. Current deep learning systems often perform poorly when tested on data that differs significantly from their training distribution.

Transfer learning and few-shot learning approaches attempt to address these limitations by leveraging knowledge from related tasks or domains. Meta-learning frameworks aim to develop systems that can quickly adapt to new tasks with minimal additional training.

10.2 Interpretability and Explainability

The black-box nature of many AI systems, particularly deep neural networks, poses challenges for understanding their decision-making processes. This lack of interpretability is particularly problematic in high-stakes applications such as healthcare and criminal justice.

Various approaches to explainable AI have been developed, including attention visualization, saliency mapping, and model-agnostic explanation methods. However, achieving truly interpretable AI systems while maintaining high performance remains an ongoing challenge.

10.3 Robustness and Safety

AI systems often exhibit brittle behavior when confronted with inputs that differ from their training data. Adversarial examples demonstrate that small, imperceptible perturbations can cause dramatic failures in neural networks.

Developing robust AI systems requires advances in adversarial training, uncertainty quantification, and formal verification methods. Safety considerations are particularly important for AI systems deployed in critical applications.

11 Future Directions and Emerging Paradigms

11.1 Artificial General Intelligence

The development of artificial general intelligence (AGI) represents a long-term goal of AI research. AGI systems would possess human-level cognitive capabilities across diverse domains, including reasoning, learning, and creative problem-solving.

Current approaches to AGI include cognitive architectures that integrate multiple AI techniques, neurosymbolic systems that combine neural and symbolic reasoning, and large-scale transformer models trained on diverse tasks.

11.2 Quantum Machine Learning

The intersection of quantum computing and machine learning offers potential advantages for certain computational problems. Quantum algorithms might provide exponential speedups for specific machine learning tasks, particularly those involving large-scale optimization or sampling problems.

However, the practical implementation of quantum machine learning faces significant technical challenges, including noise in quantum systems and the limited availability of quantum hardware.

11.3 Neuromorphic Computing

Neuromorphic computing architectures attempt to emulate the energy efficiency and computational principles of biological neural systems. These approaches use spiking neural networks and novel hardware designs to achieve low-power computation.

Neuromorphic systems show promise for edge computing applications where energy efficiency is crucial, but their integration with current AI software frameworks remains challenging.

12 Conclusion

Artificial Intelligence represents a remarkable synthesis of insights from mathematics, computer science, cognitive science, neuroscience, and philosophy. The field has achieved significant practical successes while continuing to grapple with fundamental questions about the nature of intelligence, consciousness, and machine understanding.

The mathematical foundations of AI, rooted in probability theory, optimization, and information theory, provide rigorous frameworks for developing intelligent systems. Computational architectures, from neural networks to symbolic reasoning systems, enable the implementation of these theoretical insights in practical applications.

Cognitive science and neuroscience provide crucial perspectives on the nature of intelligence and learning, informing the development of AI systems that can match or exceed human cognitive capabilities. Philosophical inquiry challenges AI researchers to consider fundamental questions about mind, consciousness, and the nature of understanding.

The ethical and societal implications of AI require careful consideration as these systems become increasingly powerful and pervasive. Issues of fairness, privacy, safety, and economic impact must be addressed through interdisciplinary collaboration between technologists, policymakers, and society at large.

Current challenges in generalization, interpretability, and robustness continue to drive research in new directions. The pursuit of artificial general intelligence, quantum machine learning, and neuromorphic computing represents exciting frontiers for future development.

The complete understanding of artificial intelligence requires integration of knowledge across multiple disciplines. As AI systems become more sophisticated and their impact on society grows, this multidisciplinary perspective becomes increasingly important for ensuring that these technologies serve human welfare and contribute to the advancement of knowledge and understanding.

The journey toward truly intelligent machines continues to challenge our understanding of intelligence itself, pushing the boundaries of what we know about computation, cognition, and consciousness. This ongoing inquiry represents one of the most profound intellectual adventures of our time, with implications that extend far beyond the technical domain into the very nature of mind and reality.

References

[1] A. M. Turing, Computing machinery and intelligence, *Mind*, 1950.

[2] C. E. Shannon, A mathematical theory of communication, *The Bell System Technical Journal*, 1948.

[3] N. Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press, 1948.

[4] W. S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, *The Bulletin of Mathematical Biophysics*, 1943.

[5] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, 1958.

[6] M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press, 1969.

[7] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences*, 1982.

[8] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, *Nature*, 1986.

[9] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, Backpropagation applied to handwritten zip code recognition, *Neural Computation*, 1989.

[10] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation*, 1997.

[11] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[12] L. Breiman, Random forests, *Machine Learning*, 2001.

[13] G. E. Hinton, S. Osindero, and Y.-W. Teh, A fast learning algorithm for deep belief nets, *Neural Computation*, 2006.

[14] Y. Bengio, Learning deep architectures for AI, *Foundations and Trends in Machine Learning*, 2009.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems*, 2012.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets, in *Advances in Neural Information Processing Systems*, 2014.

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, in *Advances in Neural Information Processing Systems*, 2017.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805*, 2018.

[19] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, Language models are few-shot learners, in *Advances in Neural Information Processing Systems*.

[20] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann, 1988.

[21] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Upper Saddle River, NJ: Prentice Hall, 2020.

[22] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[23] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2009.

[24] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA: MIT Press, 2018.

[25] C. J. C. H. Watkins and P. Dayan, Q-learning, *Machine Learning*, 1992.

[26] R. J. Williams, Simple statistical gradient-following algorithms for connectionist reinforcement learning, *Machine Learning*, 1992.

[27] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, Mastering the game of Go with deep neural networks and tree search, *Nature*, 2016.

[28] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W. H. Freeman, 1982.

[29] A. Newell and H. A. Simon, *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, 1972.

[30] J. R. Anderson, *The Architecture of Cognition*. Cambridge, MA: Harvard University Press, 1983.

[31] D. Kahneman, *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.

[32] A. Baddeley, Working memory: Looking back and looking forward, *Nature Reviews Neuroscience*, 2003.

[33] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley, 1949.

[34] D. H. Hubel and T. N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *The Journal of Physiology*, 1962.

[35] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. J. Hudspeth, *Principles of Neural Science*, 5th ed. New York: McGraw-Hill, 2013.

[36] L. R. Squire, Memory systems of the brain: A brief history and current perspective, *Neurobiology of Learning and Memory*, 2003.

[37] J. J. Gibson, *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979.

[38] G. Lakoff and M. Johnson, *The Embodied Mind: The Bodily Basis of Meaning, Imagination, and Reason*. New York: Basic Books, 1999.

[39] J. R. Searle, Minds, brains, and programs, *Behavioral and Brain Sciences*, 1980.

[40] D. C. Dennett, *Consciousness Explained*. Boston: Little, Brown and Company, 1991.

[41] D. J. Chalmers, Facing up to the problem of consciousness, *Journal of Consciousness Studies*, 1995.

[42] H. Putnam, Psychological predicates, in *Art, Mind, and Religion*, W. H. Capitan and D. D. Merrill, Eds. Pittsburgh: University of Pittsburgh Press, 1967.

[43] J. McCarthy and P. J. Hayes, Some philosophical problems from the standpoint of artificial intelligence, *Machine Intelligence*, 1969.

[44] R. Reiter, *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. Cambridge, MA: MIT Press, 2001.

[45] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, and E. Vayena, AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations, *Minds and Machines*, 2018.

[46] S. Barocas and A. D. Selbst, Big data's disparate impact, *California Law Review*, 2016.

[47] C. Dwork, Differential privacy, in *International Colloquium on Automata, Languages, and Programming*. Berlin: Springer, 2006.

[48] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 2017.

[49] E. Brynjolfsson and A. McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton & Company, 2014.

[50] D. Acemoglu and P. Restrepo, The wrong kind of AI? Artificial intelligence and the future of labour demand, *Cambridge Journal of Regions, Economy and Society*, 2020.

[51] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, M. Sun, P. Sundberg, H. Yee, K. Zhang, Y. Zhang, G. Flores, G. E. Duggan, J. Irvine, Q. Le, K. Litsch, A. Mossin, J. Tansuwan, D. Wang, J. Wexler, J. Wilson, D. Ludwig, S. L. Volchenboum, K. Chou, M. Pearson, S. Madabushi, N. H. Shah, A. J. Butte, M. D. Howell, C. Cui, G. S. Corrado, and J. Dean, Scalable and accurate deep learning with electronic health records, *npj Digital Medicine*, 2018.

[52] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, 2014.

[53] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[54] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, Intriguing properties of neural networks, *arXiv preprint arXiv:1312.6199*, 2013.

[55] A. Kurakin, I. Goodfellow, and S. Bengio, Adversarial examples in the physical world, *arXiv preprint arXiv:1607.02533*, 2016.

[56] S. Levine, C. Finn, T. Darrell, and P. Abbeel, End-to-end training of deep visuomotor policies, *Journal of Machine Learning Research*, 2016.

[57] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, Human-level control through deep reinforcement learning, *Nature*, 2015.

[58] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen,

D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature*, 2021.

[59] S. J. Pan and Q. Yang, A survey on transfer learning, *IEEE Transactions on Knowledge and Data Engineering*, 2010.

[60] C. Finn, P. Abbeel, and S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[61] M. T. Ribeiro, S. Singh, and C. Guestrin, Why should I trust you?: Explaining the predictions of any classifier, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[62] S. M. Lundberg and S.-I. Lee, A unified approach to interpreting model predictions, in *Advances in Neural Information Processing Systems*, 2017.

[63] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, Towards deep learning models resistant to adversarial attacks, *arXiv preprint arXiv:1706.06083*, 2017.

[64] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, Concrete problems in AI safety, *arXiv preprint arXiv:1606.06565*, 2016.

[65] B. Goertzel, Artificial general intelligence: Concept, state of the art, and future prospects, *Journal of Artificial General Intelligence*, 2014.

[66] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, Building machines that learn and think like people, *Behavioral and Brain Sciences*, 2017.

[67] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature*, 2017.

[68] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum*, 2018.

[69] G. Indiveri, B. Linares-Barranco, T. J. Hamilton, A. van Schaik, R. Etienne-Cummings, T. Delbruck, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, K. Hynna, F. Folowosele, S. Saighi, T. Serrano-Gotarredona, J. Wijekoon, Y. Wang, and K. Boahen, Neuromorphic silicon neuron circuits, *Frontiers in Neuroscience*, 2011.

[70] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C.-K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y.-H. Weng, A. Wild, Y. Yang, and H. Wang, Loihi: A neuromorphic manycore processor with on-chip learning, *IEEE Micro*, 2018.

The End