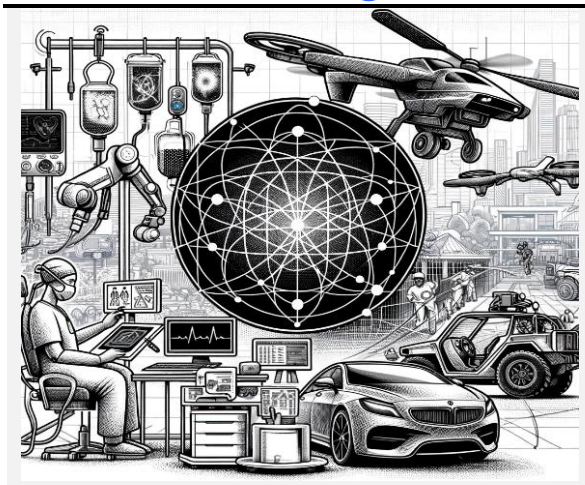
Science & Technology Policy Brief

Artificial Intelligence



The above image was generated by DALL.E, an image generation application, upon the prompt — "generate a composite line art illustration of AI in health, defence, and transport".

Summary

- Currently available AI systems are considered 'intelligent' only within the narrow context for which they are designed.
- AI has use cases in diverse areas such as health, justice, defence, and agriculture.
- Availability of quality datasets and advanced computing is key to the development of AI.
- Certain AI systems use statistical techniques, which may result in errors as well as opaqueness in how they arrive at results.
- Bias in AI may arise from training data as well as design choices, which may perpetuate existing societal prejudices in decision making.

Background

Artificial Intelligence (AI) systems have been in commercial use since at least the 1980s.[1] Recent advancements in AI have been driven by ever-increasing computational power and data availability.[1,2,3,4] Tools like ChatGPT, which can write coherently, code, and answer diverse queries, have raised further curiosity.[5,6] These have fuelled a strong interest in the societal impact of AI.[7,8,9,10,11] Efforts to regulate AI are underway in jurisdictions such as USA, European Union, and China.[12,13,14]

Owing to its transformative potential across diverse sectors, AI is considered to be of strategic importance. In India, the National Strategy for Artificial Intelligence was formulated in 2018.[15] The Strategy highlighted healthcare, agriculture, education, smart cities and infrastructure, and transport as key focus areas for the adoption of AI.[15] In October 2023, the Expert Group on AI constituted by the central government presented its first report.[16] The report provides a blueprint for operationalising the AI ecosystem in India.[16] The proposed Digital India Act is expected to regulate AI systems to some extent.[17,18]

This note explains the technology behind AI systems, its use, and related concerns. We mainly discuss concerns with AI systems that may be considered 'intelligent' only in narrow contexts. The potential effects of Human-like AI (Artificial General Intelligence) or Superintelligence (machine surpassing human intelligence) on society are not covered in this note.[19,20] Presently, these are considered technological frontiers, with varied opinions on their realisation and timeline (see Box 1 on page 2).[7,19,21,22]

Introduction

There is no universally accepted definition of AI.[2] The primary difficulty arises from defining 'Intelligence'.[23,24] A question often debated is that if a machine can do a task that requires human intelligence, does that make it truly intelligent, or is it just mimicking intelligence.[2,24,25,26] Broadly, AI refers to the ability of machines to perceive, understand language, learn, reason, and solve problems. These are akin to intelligence characterised with human cognitive processes.

AI systems can be software or a combination of software and hardware (e.g., robots or autonomous vehicles).[27] AI systems differ from traditional software programs in their approach to problem-solving. Traditional software executes explicit instructions provided by a programmer in a pre-determined order. In contrast, AI systems exhibit the ability to adapt; they can plan and search for optimal solutions or learn from data and improve their performance over time.[2] For instance, AlphaGo Zero is an AI system that could learn the board game of 'Go' purely by playing against itself.[28] It defeated its predecessor, AlphaGo Master (by 89 games to 11), which had beaten the world's number one human player (by three games to zero).[28,29] Go, played on a 19x19 grid, is considered more complex than chess.

Currently available AI systems are considered intelligent only in a narrow sense, i.e., a system which shows intelligent behaviour in one domain, may not be relevant in another. For instance, AlphaGo Zero may not be able to translate a paragraph from Chinese to English, something Ke Jie, the world no. 1 Go player beaten by AlphaGo Master, can be expected to do easily. Thus, these

Saket Surya
saket@prsindia.org

December 12, 2023

systems are referred to as 'Narrow AI'. This sets them apart from humans, who exhibit intelligent behaviour in very diverse areas. Hence, two key frontiers in AI development are considered to be: (i) 'Artificial General Intelligence' (AGI) – intelligence across a wide range of tasks and domains, similar to humans, (ii) 'Superintelligence' – where machine intelligence surpasses human intelligence across all domains and tasks.[19,20]

Box 1: Existential Risks from AI

There has been intense debate on whether AI presents existential risks for humanity.[30,31,32,33,34,35] These are concerns with scenarios where a superintelligent AI may act independently, pursue goals misaligned with human values, and thereby cause harm on a large scale.[19] These stem from the risks of development of self-preservation instinct, capability to self-improve, and loss of human control.[19] In a 2022 survey of AI researchers, the forecast time to a 50% chance of human-level AI was around 37 years, i.e., 2059.[20] A median respondent believed that there is a 5% probability of an extremely negative long-term impact of AI.[20] They acknowledged the need for increased focus on AI safety research.[20] However, past predictions regarding advancements in AI have proven to be overly optimistic.[1] Some leading researchers have argued that current approaches are unlikely to lead to human-level AI.[36,37]

Approaches for building AI Systems

Based on the approach to represent and process information, techniques to build AI systems are broadly classified into two key categories. Many systems may utilise both of these approaches.

Rule-based Systems: These systems use human-encoded knowledge, rules, and logical inferences.[2,19] They mimic the reasoning of a human expert in solving a problem. They can plan and search through encoded knowledge and rules to solve a given problem. 'Deep Blue', which defeated the world chess champion Garry Kasparov in 1997, is an example of such a system.[38] It searched through millions of possible moves encoded in advance, and attempted to find the best move by exploring as many possible moves ahead in the game.[39]

Such an approach works well where there may be relatively clear-cut goals and rules to achieve them.[2] Another advantage may be that the reasoning process may be relatively transparent and interpretable. However, such approaches have typically fallen short in applications likely to encounter very high variations and uncertainty in input. For instance, rule-based approaches could not achieve great results in translating between languages or recognising images.[40] The number of combinations or variety in input may be too large to handcraft clear-cut rules around them.

Machine Learning: These systems use statistical techniques to induce trends from patterns in data.[2,19] They learn and improve from the data they work with, and are able to generalise from the patterns they see in the examples.[41] Generalisation refers to the ability to generate acceptable output for previously unseen input. Translation may be a good example to understand the differences between rule-based approach and machine learning. In a rule-based translation system, a human programmer would encode rules of grammar of two languages, a dictionary of words, and some rules on semantics. On the other hand, a machine learning-based system would start by screening a large sample of already translated texts and inferring patterns from these examples.

The performance and quality of machine learning systems are also reliant on training data.[42,43] The performance may degrade if the input data differs significantly from the training data. Monitoring, re-training, and updates may be required in fast-evolving contexts. Some of the more complex systems are often seen as 'black boxes'; it may be challenging to explain their reasoning.[4]

Recently launched applications such as ChatGPT, Bard, and DALL.E are classified as Generative AI. Generative AI is a sub-field of machine learning, which refers to the capability of generating novel content including text, images, music, and videos.[44] These systems also learn patterns and relationships in vast amounts of already available data.[45] They utilise these learnings to generate new but similar information by predicting the statistically most likely response.[46,47] Their ability to generate content is bounded by their training data.[37,48]

Table 1: Some AI Technologies

Technology	Description	Examples
Computer Vision	Interpreting visual data, recognising objects	Health diagnostics, facial recognition
Natural Language Processing	Understanding and analysing human language	Search engine, machine translation, spam detection
Speech Recognition	Processing spoken language	Speech-to-Text input
Generative AI	Generating new text, images, audio, and other data on prompt	ChatGPT, Bard, DALL.E, Midjourney
Recommendation Systems	Providing personalised suggestions	Content and shopping recommendations
Predictive Analytics	Predicting future outcomes based on historical data	Weather forecast, fraud prevention and detection
Robotics	Machines that move and act autonomously	Industrial robots, drones

Sources: Russell, S., & Norvig, P. (2020); Calvino, F., et al. (2023); PRS.

Use cases of AI

AI has applications across diverse sectors, aiding advanced capabilities and efficiency (Table 2 on the next page). The National Strategy on AI highlighted the following as key challenges for the development of AI in India: (i) availability of quality datasets and advanced computing, (ii) lack of research in foundational technologies, (iii) talent gap, and (iv) unclear regulations on privacy, security, and intellectual property rights.[15]

Table 2: Illustrative list of use cases of AI

Area	Applications
Healthcare[49,50,51]	Diagnosis including interpretation of medical imaging such as cancer detection, prediction of health risks, recommendations on treatment plans, virtual health assistants, drug discovery, robotic surgery
Education[52,53]	Personalised learning, learning content creation, automated grading and feedback
Finance[54,55,56,57]	Fraud prevention and detection, credit scoring, risk modelling for insurance, algorithmic trading and investment
Transport[58,59,60]	Driver assistance, traffic management, navigation, route optimisation, logistics planning, autonomous vehicles
Policing[61,62]	Predicting criminal hotspots and patrolling routes, facial recognition to identify suspects, automated traffic challan, profiling of suspects, crime pattern analysis, social media monitoring to identify threats
Judiciary[63,64,65]	Legal research, summarisation and translation of case laws, case allocation, recommendation on bail, parole and sentencing, prediction of case outcomes and risk of re-offending
Defence[66]	Border patrolling, autonomous combat vehicles, autonomous weapon systems
Media[67,68]	Content recommendation and personalisation, content writing, sentiment analysis, content moderation
Manufacturing[69,70]	Product design, process and task automation, industrial robots, predictive maintenance, demand forecasting
Retail[71,72]	Customer profiling, targeted advertising, purchase recommendations, automated grievance redressal system
Agriculture[73]	Autonomous equipment for field work, crop planning, automated detection of pathogens, yield prediction
Research[74,75,76]	Generation of new research hypotheses, exploratory data analysis, modelling, simulation

Sources: Refer to endnotes marked in the 'Area' column; PRS.

Concerns with Use of AI

While AI has the potential to transform a wide range of sectors significantly, its usage has also raised certain concerns discussed below.

Errors

Where AI systems apply probabilistic approaches, there remains a possibility for error. Errors may be accentuated by problems with training data and gaps in design.[77,78] Consequences of such errors may be severe in many applications such as healthcare and legal systems. For instance, when facial recognition over a video feed gives an incorrect match, it could result in a wrongful arrest.[79,80,81] A series of assessments of facial recognition systems by the National Institute of Standards and Technology of USA has shown a significant reduction in error rates over the years.[82,83,84,85] However, error rates continue to be relatively higher for certain demographic groups such as women, children, elderly, and races such as African and Asian persons.[86,87] Further, accuracy may vary widely across developers.[85,86]

Generative AI systems sometimes show behaviour characterised as 'hallucination'.[88,89] They present information that may be inaccurate, irrelevant, or non-existent, as if they are correct.[88,90,91] Such inaccuracies may be a limitation in their usage in areas such as searching, summarisation, and programming. For instance, errors in information retrieval may result in misinformation.

Bias

Another critical challenge that often intersects with errors is bias. Bias refers to systematic discrimination against certain groups and categories. For instance, a 2016 analysis of COMPAS, an AI system deployed to predict the risk of re-offending in USA, showed that white persons were less likely to be termed high-risk; whereas black persons were more likely to be termed high-risk.[92] Similar racial bias has been found in the predictive systems to guide health decisions.[93] The system in question identified which patients will benefit from high-risk care management programs. Most social media platforms deploy some or other form of automated content moderation to keep harmful content away. Ethnic and gender-based bias has been observed in systems to detect offensive speech online.[94]

Bias may creep in at various levels including training data, algorithm design, and feedback cycles.[95,96,97] As many predictive or recommendation systems are trained on historical data, societal prejudices or biases in the real world also become part of the model through the training data.[98] Similarly, under-representation of certain groups in datasets may adversely impact the quality of results for those groups.[97] By not recognising such biases in data, developers may also perpetuate such biases. Developers may introduce or reinforce biases when they select or assign priorities.[99] Bias may also be observed if the system is used in contexts or with audiences who may not have been accounted for at the design stage.[100]

Explainability

AI systems, especially machine learning systems, use approaches where it may be challenging to determine how a decision was arrived at.[101,102] This limitation may pose a challenge in their adoption in high-risk areas such as justice, healthcare, and finance. Explainability may be necessary for user trust and confidence in the system.[103] Lack of explainability may also be incompatible with existing regulations and standards. For instance, in case of justice, explainability may be considered necessary for adherence to established norms of due process and reasoned orders. Further, it could also act as a safeguard against errors and biases.

Transparency

Another issue is related to a lack of information in the public domain about how these systems exactly work.[104] For instance, the underlying algorithms or datasets may not be publicly accessible.[104] This reduces opportunities for communities to evaluate capabilities as well as risks, which could help in identifying the pitfalls of such systems. Lack of transparency may lead to a lack of evidence to guide their adoption and also erode the trust in the overall system. Arguments also exist against making systems totally transparent, as this may lead to easy replication for misuse, cyber attacks, or revealing of trade secrets.[104,105]

Accountability

Determining responsibility for issues or errors with AI systems is a complex and evolving area of discussion.[23,106] The opaque nature as well as complexity of AI systems may make it difficult to pinpoint individual responsibilities in systems that involve multiple actors and resources.[96] For instance, a question may arise who is responsible for the accident of an autonomous car, resulting in the death of a person crossing the road.

Currently, in areas such as justice, AI systems have not replaced human decision-making; they only aid them.[65] Human oversight is built in as a safety mechanism.[107,108] However, users of such systems may exhibit cognitive bias where they place excessive trust in automated systems (referred to as automation bias), even where it contradicts their own judgement or expertise.[108] This may limit the effectiveness of the human oversight process.

Privacy

Many AI systems are trained on the personal information of individuals. Current privacy protection laws are based on the principle of data minimisation to protect privacy.[109] This means that the least possible data should be collected to meet a purpose, and data should be deleted once the purpose is fulfilled. These principles may be in tension with the nature of AI systems. Machine-learning-based systems need large amounts of data for training and testing algorithms.[36] For instance, the accuracy of facial recognition technology has been improved with large datasets built from publicly available photos over internet.[110] Such systems have faced lawsuits under privacy laws. In Australia, the Privacy Regulator ordered Clearview AI to delete its datasets.[110] AI systems may combine datasets to infer new insights about a person, which may also pose privacy risks.[40]

Intellectual property rights

Questions have emerged on two key grounds: (i) whether original works may be used in training without a license, and (ii) who owns the intellectual property rights for the AI-generated content or invention between the developer and the human operator.[23,111] Generative AI systems may generate content which may not be sufficiently distinct from the copyrighted training data, and can then be prone to copyright challenges.[119,120] Image generation applications such as Stable Diffusion and Midjourney are facing lawsuits for using datasets compiled from scraping the web indiscriminately, which may include copyrighted creations.[121,122] The question arises whether training of AI systems may be covered under fair use.[119,123] Existing copyright laws allow fair use, i.e., use for criticism, comment, reporting, teaching, scholarship, or research.[119,124] Currently, under intellectual property laws worldwide and also in India, rights are given only to human creators.[23,111,119,125] Any AI-generated content or invention is not eligible for copyright or patent, respectively. On the other hand, it may be argued that lack of such rights may be a disincentive for persons who build, own, and use AI.[23]

Box 2: Regulation of AI around the world

European Union: A draft law proposing risk-based regulation is under consideration.[112] Certain applications are sought to be restricted, which include real-time biometric identification systems in public places. Prior impact assessment and regular audits would be mandatory for areas such as health, law, public services, and employment, which intersect with fundamental rights.

USA: In October 2023, the President issued an Executive Order regarding the regulation of AI.[113] The Order provides for the development of standards for testing before the public release of certain AI systems. It requires the government to take steps to protect privacy, address discrimination, and ensure safety. An AI Bill of Rights has been proposed.[114] A local law in New York City regulates the use of AI in employment.[115,116]

China: In China, regulations are in place for certain specific aspects of AI. For instance, separate regulations provide for the use of recommendation algorithms, certain machine learning applications, and Generative AI applications.[117]

India: In 2021, NITI Aayog released a Responsible AI framework which outlines key principles for managing AI.[118]

Employment

While previous technological advances in automation have affected routine or repetitive tasks, AI has the potential to automate non-routine tasks including creative and analytical tasks.[11,126,127] For instance, AI systems can write computer programs and fix bugs in software.[128] Advancements in AI may expose large parts of the workforce to potential disruption. This may happen for both low and high-paid jobs.[127]

Like earlier technological advances, the adoption of AI may also lead to the creation of new types of occupations. For instance, 60% of the employment in 2018 in the USA was in jobs that did not exist in the 1940s.[127] In the case of currently available Generative AI, an ILO study (2023) observed that their adoption is more likely to automate certain parts of jobs than substitute the human worker entirely.[127] However, the impact may be high for certain roles. For instance, clerical work may be more exposed to risks of replacement by AI.[127]

[1] Chapter 1: The Technical Landscape, Artificial Intelligence in Society, OECD, as accessed on November 15, 2023, https://www.oecd-ilibrary.org/sites/eedfee77-en/1/2/1/index.html?itemId=/content/publication/eedfee77-en&_csp_=5c39a73676a331d76fa56f36ff0d4aca&itemIGO=oecd&itemContentType=book.

[2] Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Prentice Hall.

[3] The state of AI in 2022—and a half decade in review, McKinsey, December 6, 2022, https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review.

[4] White Paper on Artificial Intelligence: A European Approach to Excellence and Trust, European Commission, February 19, 2020, https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

[5] Artificial Intelligence Index Report 2023, Stanford University, https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf.

[6] Y. Wang, Y. Pan, M. Yan, Z. Su, and T. H. Luan, "A Survey on ChatGPT: AI-Generated Contents, Challenges, and Solutions" in IEEE Open Journal of the Computer Society, August 16, 2023, doi: 10.1109/OJCS.2023.3300321. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10221755.

[7] The Godfather in Conversation: Why Geoffrey Hinton is worried about the future of AI, University of Toronto, YouTube, June 22, 2023, https://youtu.be/-9cW4Gcn5WY?si=yv1eUeuxiN7QuI1_.

[8] "We analyzed 16,625 papers to figure out where AI is headed next", Karen Hao, MIT Technology Review, January 25, 2019, https://www.technologyreview.com/2019/01/25/1436/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/.

[9] The State of AI in 2023: Generative AI's breakout year, McKinsey, August 1, 2023, https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year.

[10] "Gartner Survey Shows Generative AI Has Become an Emerging Risk for Enterprises", Gartner, August 8, 2023, https://www.gartner.com/en/newsroom/press-releases/2023-08-08-gartner-survey-shows-generative-ai-has-become-an-emerging-risk-for-enterprises.

[11] The Impact of Artificial Intelligence on Work, an Evidence Review prepared for the Royal Society and the British Academy by Frontier Economics, September 2018, https://royalsociety.org/-/media/policy/projects/ai-and-work/frontier-review-the-impact-of-AI-on-work.pdf.

[12] "President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence", White House, October 30, 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.

[13] "EU AI Act: first regulation on artificial intelligence", Website of European Parliament as accessed on November 15, 2023, https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

[14] "China moves to support generative AI, regulate applications", Website of the State Council, The People's Republic of China, July 13, 2023, https://english.www.gov.cn/news/202307/13/content_WS64aff5b3c6d0868f4e8ddc01.html.

[15] The National Strategy on Artificial Intelligence, NITI Aayog, June 2018, https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf.

[16] India AI 2023, First Edition by Expert Group, Ministry of Electronics and Information Technology, October 2023, https://www.meity.gov.in/writereaddata/files/IndiaAI-Expert-Group-Report-First-Edition.pdf.

[17] "MoS Rajeev Chandrasekhar holds Digital India Dialogues in Mumbai on the Principles of the Digital India Act," Press Information Bureau, Ministry of Electronics and Information Technology, May 23, 2023, https://www.pib.gov.in/PressReleseDetailm.aspx?PRID=1926711.

[18] "New vistas. Draft Digital India Act will regulate emerging technologies to protect citizens: Rajeev Chandrasekhar", The Hindu Business Line, June 12, 2023, https://www.thehindubusinessline.com/info-tech/draft-digital-india-act-will-regulate-emerging-technologies-to-protect-citizens-rajeev-chandrasekhar/article66960829.ece.

[19] Bostrom, N. (2016). Superintelligence: Paths, Dangers, Strategies. [Paperback edition]. Oxford University Press.

[20] Katja Grace, Zach Stein-Perlman, Benjamin Weinstein-Raun, and John Salvatier, "2022 Expert Survey on Progress in AI." AI Impacts, 3 Aug, 2022. https://wiki.aiimpacts.org/doku.php?id=ai_timelines:predictions_of_human-level_ai_timelines:ai_timeline_surveys:2022_expert_survey_on_progress_in_ai.

[21] "Artificial General Intelligence Is Not as Imminent as You Might Think", Gary Marcus, Scientific American, July 1, 2022, https://www.scientificamerican.com/article/artificial-general-intelligence-is-not-as-imminent-as-you-might-think1/.

[22] Gary Marcus and Ernest Davis. 2019. Rebooting AI: Building Artificial Intelligence We Can Trust. Pantheon Books, USA.

[23] Abbott, R. (2020). The Reasonable Robot: Artificial Intelligence and the Law. Cambridge: Cambridge University Press. doi:10.1017/9781108631761.

[24] McCarthy (2007). What is Artificial Intelligence? http://jmc.stanford.edu/articles/whatisai/whatisai.pdf.

[25] A. M. Turing (1950). Computing Machinery and Intelligence. Mind. http://lia.deis.unibo.it/corsi/2005-2006/SID-LS-CE/downloads/turing-article.pdf.

[26] Jonas Schuett (2023). Defining the scope of AI regulations, Law, Innovation and Technology, 15:1, 60-82, DOI: 10.1080/17579961.2023.2184135.

[27] The ethics of Artificial Intelligence: Issues and Initiatives, Panel for the Future of Science and Technology, European Parliament Research Service, March 2020, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf.

[28] AlphaGo Zero: Starting from scratch, Google Deepmind, October 18, 2017, https://www.deepmind.com/blog/alphago-zero-starting-from-scratch.

[29] Silver, D., Schrittwieser, J., Simonyan, K. et al. Mastering the game of Go without human knowledge. Nature 550, 354–359 (2017). https://doi.org/10.1038/nature24270.

[30] "A.I. Poses 'Risk of Extinction,' Industry Leaders Warn", New York Times, Kevin Roose, May 30, 2023, https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html.

[31] "What's changed since the "pause AI" letter six months ago?", Melissa Heikkilä, MIT Technology Review, September 26, 2023, https://www.technologyreview.com/2023/09/26/1080299/six-months-on-from-the-pause-letter/.

[32] "AI Is an Existential Threat—Just Not the Way You Think", The Conversation US & Nir Eisikovits, Scientific American, July 12, 2023, https://www.scientificamerican.com/article/ai-is-an-existential-threat-just-not-the-way-you-think/.

[33] "AI will never threaten humans, says top Meta scientist", John Thornhill, Financial Times, October 19, 2023, https://www.ft.com/content/30fa44a1-7623-499f-93b0-81e26e22f2a6.

[34] An Executive Primer on Artificial General Intelligence, McKinsey, April 2020, https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Operations/Our%20Insights/An%20executive%20primer%20on%20artificial%20general%20intelligence/an-executive-primer-on-artificial-general-intelligence.pdf.

[35] The Bletchley Declaration by Countries Attending the AI Safety Summit, November 1-2 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.

[36] Marcus, G. (2018). Deep Learning: A Critical Appraisal. https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf.

[37] Yann LeCun (2022). A Path Towards Autonomous Machine Intelligence Version 0.9.2. https://openreview.net/pdf?id=BZ5a1r-kVsf.

[38] "What the history of AI tells us about its future", Clive Thompson, February 18, 2022, https://www.technologyreview.com/2022/02/18/1044709/ibm-deep-blue-ai-history/.

[39] Murray Campbell. 1999. Knowledge discovery in deep blue. Commun. ACM 42, 11 (Nov. 1999), 65–67. https://doi.org/10.1145/319382.319396 (https://dl.acm.org/doi/pdf/10.1145/319382.319396).

[40] Fry, H. (2018). "Hello World: Being Human in the Age of Algorithms. Doubleday.

[41] Pedro Domingos. 2012. A few useful things to know about machine learning. Commun. ACM 55, 10 (October 2012), 78–87. https://dl.acm.org/doi/pdf/10.1145/2347736.2347755.

[42] Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. **2**, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x.

[43] Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN COMPUT. SCI. **2**, 420 (2021). https://doi.org/10.1007/s42979-021-00815-1.

[44] ACM TechBrief: Generative Artificial Intelligence, ACM Technology Policy Council (Issue 8, Summer 2023), https://dl.acm.org/doi/pdf/10.1145/3626110.

[45] Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. Bus Inf Syst Eng (2023). https://link.springer.com/content/pdf/10.1007/s12599-023-00834-7.pdf.

[46] Emily M. Bender and Alexander Koller. 2020. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 5185–5198, Online. Association for Computational Linguistics. https://aclanthology.org/2020.acl-main.463.pdf.

[47] "Generative AI exists because of the transformer", Visual Storytelling Team and Madhumita Murgia, Financial Times, September 12, 2023, https://ig.ft.com/generative-ai/.

[48] "Introducing ChatGPT", Website of OpenAI, as accessed on November 20, 2023, https://openai.com/blog/chatgpt.

[49] Adam Bohr, Kaveh Memarzadeh, Chapter 2 - The rise of artificial intelligence in healthcare applications, Editor(s): Adam Bohr, Kaveh Memarzadeh, Artificial Intelligence in Healthcare, Academic Press, 2020, Pages 25-60, ISBN 9780128184387, https://doi.org/10.1016/B978-0-12-818438-7.00002-2. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7325854/).

[50] Ethics and Governance of Artificial Intelligence for Health: WHO Guidance, WHO, 2021, https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf?sequence=1.

[51] Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare, ICMR, 2023, https://main.icmr.nic.in/sites/default/files/upload_documents/Ethical_Guidelines_AI_Healthcare_2023.pdf.

[52] AI and Education: Guidance for Policy Makers, UNESCO, 2021, https://unesdoc.unesco.org/ark:/48223/pf0000376709?locale=en.

[53] Guidance for Generative AI in Education and Research, UNESCO, September 2023, https://unesdoc.unesco.org/ark:/48223/pf0000386693.

[54] Artificial Intelligence, Machine Learning, and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers, OECD, 2021, https://www.oecd.org/finance/financial-markets/Artificial-intelligence-machine-learning-big-data-in-finance.pdf.

[55] Generative Artificial Intelligence in Finance: Risk Considerations, International Monetary Fund, August 2023, https://unesdoc.unesco.org/ark:/48223/pf0000386693.

[56] The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector, OECD, 2020, www.oecd.org/finance/Impact-Big-Data-AI-in-the-Insurance-Sector.htm.

[57] Artificial Intelligence Innovation in Financial Services, International Finance Corporation, World Bank, June 2020, https://documents1.worldbank.org/curated/en/839801596184068790/pdf/Artificial-Intelligence-Innovation-in-Financial-Services.pdf.

[58] Artificial intelligence in transport: Current and future developments, opportunities and challenges, European Parliamentary Research Service, March 2019, https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/635609/EPRS_BRI(2019)635609_EN.pdf.

[59] On the road to automated mobility: An EU strategy for mobility of the future, European Commission, May 2018, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0283.

[60] Preparing Infrastructure for Automated Vehicles, International Transport Forum, OECD, 2023, https://www.itf-oecd.org/sites/default/files/docs/preparing-infrastructure-automated-vehicles.pdf.

[61] Christopher Rigano, "Using Artificial Intelligence to Address Criminal Justice Needs", NIJ Journal 280, January 2019, https://www.nij.gov/journals/280/Pages/using-artificialintelligence-to-address-criminal-justice-needs.aspx.

[62] Redden, J., Aagaard, B., Taniguchi, T., & Criminal Justice Testing and Evaluation Consortium. (2020). Artificial Intelligence in Law Enforcement. U.S. Department of Justice, National Institute of Justice, Office of Justice Programs. https://cjtec.org/files/5f5f94aa4c69b.

[63] Global Toolkit on AI and the Rule of Law for the Judiciary, UNESCO, 2023, https://unesdoc.unesco.org/ark:/48223/pf0000387331.

[64] Chen, D.L. Judicial Analytics and the Great Transformation of American Law. Artif Intell Law 27, 15–42 (2019). https://doi.org/10.1007/s10506-018-9237-x.

[65] Risk Assessment Tool Database, Berkman Klein Center for Internet and Society, Harvard University, as accessed on December 1, 2023, https://criminaljustice.tooltrack.org/.

[66] Y. Zhang, Z. Dai, L. Zhang, Z. Wang, L. Chen and Y. Zhou, "Application of Artificial Intelligence in Military: From Projects View," 2020 6th International Conference on Big Data and Information Analytics (BigDIA), Shenzhen, China, 2020, pp. 113-116, doi: 10.1109/BigDIA51454.2020.00026.

[67] "How Generative AI Is Changing Creative Work", Thomas H. Davenport and Nitin Mittal, Harvard Business Review, November 14, 2022, https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work.

[68] Use of AI in Online Content Moderation, 2019 Report produced on behalf of Ofcom, Cambridge Consultants, https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf.

[69] Nolan, A. (2021), "Artificial intelligence, its diffusion and uses in manufacturing", OECD Going Digital Toolkit Notes, No. 12, OECD Publishing, Paris, https://doi.org/10.1787/249e2003-en.

[70] Lane, M., M. Williams and S. Broecke (2023), "The impact of AI on the workplace: Main findings from the OECD AI surveys of employers and workers", OECD Social, Employment and Migration Working Papers, No. 288, OECD Publishing, Paris, https://doi.org/10.1787/ea0a0fe1-en. (https://www.oecd-ilibrary.org/social-issues-migration-health/the-impact-of-ai-on-the-workplace-main-findings-from-the-oecd-ai-surveys-of-employers-and-workers_ea0a0fe1-en).

[71] AI-powered marketing and sales reach new heights with generative AI, McKinsey, May 2023, https://www.mckinsey.com/~/media/mckinsey/business%20functions/marketing%20and%20sales/our%20insights/ai%20powered%20marketing%20and%20sales%20reach%20new%20heights%20with%20generative%20ai/ai-powered-marketing-and-sales-reach-new-heights-with-generative-ai.pdf.

[72] Bawack, R.E., Wamba, S.F., Carillo, K.D.A. et al. Artificial Intelligence in E-Commerce: a bibliometric study and literature review. Electron Markets 32, 297–338 (2022). https://doi.org/10.1007/s12525-022-00537-z.

[73] Artificial Intelligence in the Agri-Food Sector, Scientific Foresight Unit, European Parliament Research Service, March 2023,

https://www.europarl.europa.eu/RegData/etudes/STUD/2023/734711/EPRS_STU(2023)734711_EN.pdf.

[74] Artificial Intelligence in Science: Challenges, Opportunities, and the Future of Research, OECD, June 2023, https://www.oecd.org/publications/artificial-intelligence-in-science-a8d820bd-en.htm.

[75] Wang, H., Fu, T., Du, Y. et al. Scientific discovery in the age of artificial intelligence. Nature 620, 47–60 (2023). https://doi.org/10.1038/s41586-023-06221-2.

[76] AI and science: what 1,600 researchers think, Richard Van Noorden, Jeffrey M. Perkel, September 27, 2023, https://www.nature.com/articles/d41586-023-02980-0.

[77] Northcutt, C. G., Athalye, A., Lin, J. Pervasive Label Errors in ML Benchmark Test Sets, Consequences, and Benefits. In NeurIPS 2020 Workshop on Security and Data Curation Workshop (2020).

[78] The Foundations of AI Are Riddled With Errors, Will Knight, Wired, March 31, 2021, https://www.wired.com/story/foundations-ai-riddled-errors/.

[79] "Wrongfully Accused by an Algorithm", Kashmir Hill, New York Times, June 24, 2020, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

[80] "The new lawsuit that shows facial recognition is officially a civil rights issue", Tate Ryan-Mosley, MIT Technology Review, April 14, 2021, https://www.technologyreview.com/2021/04/14/1022676/robert-williams-facial-recognition-lawsuit-aclu-detroit-police/.

[81] "How Wrongful Arrests Based on AI Derailed 3 Men's Lives", Khari Johnson, Wired, March 7, 2022, https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/.

[82] 'Face Recognition Technology Evaluation (FRTE) 1:1 Verification', National Institute of Standards and Technology, USA, https://pages.nist.gov/frvt/html/frvt11.html.

[83] Patrick Grother, Mei Ngan, and Kayee Hanaoka. Face recognition vendor test (frvt) part 1: Verification. Interagency Report DRAFT, National Institute of Standards and Technology, October 2019.

[84] Patrick Grother, Mei Ngan, and Kayee Hanaoka. Face recognition vendor test (frvt) part 2: Identification. Interagency Report 8271, National Institute of Standards and Technology, September 2019. https://doi.org/10.6028/NIST.IR.8271.

[85] Patrick Grother, Mei Ngan, Kayee Hanaoka, Joyce C. Ang, Austin Home. Face Recognition Technology Evaluation (FRTE) Part I: Verification, National Institute of Standards and Technology, October 2023. https://github.com/usnistgov/frvt/blob/nist-pages/reports/11/frvt_11_report.pdf.

[86] Patrick Grother, Mei Ngan, Kayee Hanaoka. Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects, National Institute of Standards and Technology, December 2019. https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf.

[87] Buolamwini, J; Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, in Proceedings of Machine Learning Research. https://proceedings.mlr.press/v81/buolamwini18a.html.

[88] Yuyan Chen, Qiang Fu, Yichen Yuan, Zhihao Wen, Ge Fan, Dayiheng Liu, Dongmei Zhang, Zhixu Li, and Yanghua Xiao. 2023. Hallucination Detection: Robustly Discerning Reliable Answers in Large Language Models. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23). Association for Computing Machinery, New York, NY, USA, 245–255. https://doi.org/10.1145/3583780.3614905. (https://dl.acm.org/doi/10.1145/3583780.3614905).

[89] Beutel, G., Geerits, E. & Kielstein, J.T. Artificial hallucination: GPT on LSD? Crit Care 27, 148 (2023). https://doi.org/10.1186/s13054-023-04425-6.

[90] Church, K., Yue, R. (2023). Emerging trends: Smooth-talking machines. Natural Language Engineering, 29(5), 1402-1410. doi:10.1017/S1351324923000463.

[91] "Two US Lawyers fined for submitting fake court citations from ChatGPT", Dan Milmo and agency, The Guardian, June 23, 2023, https://www.theguardian.com/technology/2023/jun/23/two-us-lawyers-fined-submitting-fake-court-citations-chatgpt.

[92] Machine Bias, Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, Pro Publica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[93] Ziad Obermeyer et al, dissecting racial bias in an algorithm used to manage the health of populations. Science366,447-453(2019). DOI:10.1126/science. aax2342.

[94] Bias in Algorithms: Artificial Intelligence and Discrimination, European Union Agency for Fundamental Rights, 2022, https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf.

[95] Miron, M., Tolan, S., Gómez, E. et al. Evaluating causes of algorithmic bias in juvenile criminal recidivism. Artif Intell Law 29, 111–147 (2021). https://doi.org/10.1007/s10506-020-09268-y.

[96] Novelli, C., Taddeo, M. & Floridi, L. Accountability in artificial intelligence: what it is and how it works. AI & Soc (2023). https://doi.org/10.1007/s00146-023-01635-y.

[97] Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A. and Hall, P. (2022), Towards a Standard for Identifying and Managing Bias in Artificial Intelligence, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.SP.1270, (https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=934464).

[98] Richardson, Rashida and Schultz, Jason and Crawford, Kate, Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice (February 13, 2019). 94 N.Y.U. L. REV. ONLINE 192 (2019), Available at SSRN: https://ssrn.com/abstract=3333423.

[99] X. Ferrer, T. v. Nuenen, J. M. Such, M. Coté and N. Criado, "Bias and Discrimination in AI: A Cross-Disciplinary Perspective," in IEEE Technology and Society Magazine, vol. 40, no. 2, pp. 72-80, June 2021, doi: 10.1109/MTS.2021.3056293.

[100] David Danks and Alex John London. 2017. Algorithmic bias in autonomous systems. In Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17). AAAI Press, 4691–4697.

[101] Luca Nannini, Agathe Balayn, and Adam Leon Smith. 2023. Explainability in AI Policies: A Critical Review of Communications, Reports, Regulations, and Standards in the EU, US, and UK. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23). Association for Computing Machinery, New York, NY, USA, 1198–1212. https://doi.org/10.1145/3593013.3594074/.

[102] Cecilia Panigutti, Ronan Hamon, Isabelle Hupont, David Fernandez Llorca, Delia Fano Yela, Henrik Junklewitz, Salvatore Scalzo, Gabriele Mazzini, Ignacio Sanchez, Josep Soler Garrido, and Emilia Gomez. 2023. The role of explainable AI in the context of the AI Act. In 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23), June 12--15, 2023, Chicago, IL, USA. ACM, New York, NY, USA 12 Pages. https://doi.org/10.1145/3593013.3594069.

[103] Explainable AI: the basics, The Royal Society, November 2019, https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf.

[104] Diakopoulos, Nicholas, 'Transparency', in Markus D. Dubber, Frank Pasquale, and Sunit Das (eds), The Oxford Handbook of Ethics of AI (2020; online edn, Oxford Academic, 9 July 2020), https://doi.org/10.1093/oxfordhb/9780190067397.013.11.

[105] Weller, A. (2019). Transparency: Motivations and Challenges. In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L., Müller, KR. (eds) Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Lecture Notes in Computer Science(), vol 11700. Springer, Cham. https://doi.org/10.1007/978-3-030-28954-6_2.

[106] Kroll, Joshua A., 'Accountability in Computer Systems', in Markus D. Dubber, Frank Pasquale, and Sunit Das (eds), The Oxford Handbook of Ethics of AI (2020; online edn, Oxford

Academic, 9 July 2020), https://doi.org/10.1093/oxfordhb/9780190067397.013.10.

[107] Ben Green, The flaws of policies requiring human oversight of government algorithms, Computer Law & Security Review, Volume 45, 2022, 105681, ISSN 0267-3649, https://doi.org/10.1016/j.clsr.2022.105681. (https://www.sciencedirect.com/science/article/pii/S0267364922000292).

[108] Laux, J. Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act. AI & Soc (2023). https://doi.org/10.1007/s00146-023-01777-z.

[109] Article 5, GDPR, European Union, https://gdpr-info.eu/art-5-gdpr/.

[110] "Clearview AI breached Australians' privacy", Australian Information Commissioner and Privacy Commissioner, November 3, 2021, https://www.oaic.gov.au/newsroom/clearview-ai-breached-australians-privacy.

[111] Public Views on Artificial Intelligence and Intellectual Property Policy, US Patent Office, October 2020, https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf.

[112] EU AI Act: first regulation on artificial intelligence, Website of European Parliament, as accessed on December 1, 2023, https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence/.

[113] FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, White House, October 30, 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.

[114] Blueprint for an AI Bill of Rights, White House Office of Science and Technology Policy, October 2022, https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

[115] NYC's New AI Bias Law Broadly Impacts Hiring and Requires Audits, Jonathan Kestenbaum, Bloomberg Law, July 5, 2023, https://news.bloomberglaw.com/us-law-week/nycs-new-ai-bias-law-broadly-impacts-hiring-and-requires-audits.

[116] "Automated Employment Decision Tools", Local Law 144 of 2021, New York City Council, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=ID%7CText%7C&Search=.

[117] China's New AI Regulations, Latham and Watkins, August 16, 2023, https://www.lw.com/en/admin/upload/SiteAttachments/Chinas-New-AI-Regulations.pdf.

[118] Responsible AI: Approach Document for India Part-I Principles for Responsible AI, NITI Aayog, February 2021, https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf.

[119] "Generative Artificial Intelligence and Copyright Law", Congressional Research Service, as accessed on November 20, 2023, https://crsreports.congress.gov/product/pdf/LSB/LSB10922.

[120] "Judge pares down artists' AI copyright lawsuit against Midjourney, Stability AI", Blake Brittain, Reuters, October 31, 2023, https://www.reuters.com/legal/litigation/judge-pares-down-artists-ai-copyright-lawsuit-against-midjourney-stability-ai-2023-10-30/.

[121] "Getty Images lawsuit says Stability AI misused photos to train AI", Blake Brittain, Reuters, February 6, 2023, https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/.

[122] "Artists take new shot at Stability, Midjourney in updated copyright lawsuit", Blake Brittain, Reuters, December 1 2023, https://www.reuters.com/legal/litigation/artists-take-new-shot-stability-midjourney-updated-copyright-lawsuit-2023-11-30/.

[123] "Meet the Lawyer Leading the Human Resistance Against AI", Kate Knibbs, Wired, November 22, 2023, https://www.wired.com/story/matthew-butterick-ai-copyright-lawsuits-openai-meta/.

[124] Section 52, Copyright Act, 1957, https://www.indiacode.nic.in/bitstream/123456789/1367/1/A1957-14.pdf.

[125] Section 2 read with Section 6, Patents Act, 1970, https://www.indiacode.nic.in/bitstream/123456789/1392/1/AA1970___39.pdf.

[126] The Impact of Artificial Intelligence on the Future of Workforces in the European Union and the United States of America, European Commission, December 2022, https://digital-strategy.ec.europa.eu/en/library/impact-artificial-intelligence-future-workforces-eu-and-us.

[127] Gmyrek, P., Berg, J., Bescond, D. 2023. Generative AI and jobs: A global analysis of potential effects on job quantity and quality, ILO Working Paper 96 (Geneva, ILO). https://doi.org/10.54394/FHEM8239. (https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_890740/lang--en/index.htm).

[128] Getafix: How Facebook tools learn to fix bugs automatically, Meta, November 6, 2018, https://ai.meta.com/blog/getafix-how-facebook-tools-learn-to-fix-bugs-automatically/.