

Review

Not peer-reviewed version

The Hidden Legal and Ethical Questions Behind Artificial Intelligence

[Alia Hamdy](#)*

Posted Date: 27 November 2025

doi: 10.20944/preprints202511.2052.v1

Keywords: artificial intelligence (AI); AI ethics; algorithmic bias; autonomous systems; legal and policy frameworks; socioeconomic impact; AI in education



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

The Hidden Legal and Ethical Questions Behind Artificial Intelligence

Alia Hamdy

Independent Researcher, Egypt; alia_hamdy2009@hotmail.com

Abstract

Artificial intelligence (AI) has become an integral component of modern society, transforming the way individuals communicate, conduct business, and interact. Drawing on recent case law and statutory developments, the discussion assesses how existing legal frameworks grounded in human accountability fail to account for autonomous decision-making systems. It assesses the challenges posed by data misuse, copyright infringement in generative technologies, and algorithmic bias, which perpetuate social and economic bias. Ethical analysis extends to questions of moral status and emotional manipulation, arguing that the anthropomorphism of AI risks diminishing human values while evaluating the extent to which AI should be involved in decision-making. Through the lens of autonomy and human agency, the study demonstrates how reliance on technologies can risk eroding critical thinking, compromising professional judgement, and jeopardising social isolation. Socioeconomic considerations further reveal AI's disruptive role in labour markets, widening skill gaps, and reshaping employment hierarchies while substituting existing jobs. Within education, AI serves as a transformational tool by offering personalised learning and administrative efficiency. Yet, the absence of emotional intelligence underscores the irreplaceable role of human connection in fostering creativity and safety, alongside concerns about bias, discrimination and algorithmic manipulation further complicate the relationship between innovative technologies and societal well-being, emphasising the need for adaptive governance that balances innovation with ethical integrity to safeguard individuals' safety and integrity in a rapidly evolving environment. This study analyses proposed and existing frameworks, while suggesting alternative frameworks to help mitigate such issues.

Keywords: artificial intelligence (AI); AI ethics; algorithmic bias; autonomous systems; legal and policy frameworks; socioeconomic impact; AI in education

Introduction

Artificial intelligence (AI) is expeditiously revolutionising multiple aspects of modern life, permeating critical sectors such as communication, transportation, and healthcare, while extending its reach into broader social, economic, and legal domains. Its capacity to enhance decision-making, streamline operations, and generate meaningful insight has positioned it as a central force across diverse fields ranging from marketing and education to personalised healthcare, predictive policing, and AI-generated art. Despite these advancements, as artificial intelligence becomes increasingly integrated into our daily lives, it implicitly reshapes human behaviour and social dynamics, contributing to detrimental behavioural patterns in ways we have yet to discern. The ubiquity of reliance on AI risks fostering cognitive offloading, where individuals forget how to perform basic skills, as well as overdependence, where critical thinking abilities, creativity, and problem-solving skills all deteriorate in the long term.

Amidst technological progression and the evolution of societal dynamics, a critical question emerges: "Do human rights policies and artificial intelligence coexist?" The growing autonomy of AI raises significant ethical concerns, particularly when it is entrusted with the authority to make

decisions that have consequential implications. Is it appropriate to let algorithms determine who is entitled to employment, financial aid, or housing? In a period dominated by data-driven systems, are our privacy rights adequately protected, especially when data misuse and breaches continue to pose an imminent threat? These concerns illuminate the enduring dissonance between accelerated technological advancement and the intricate legal and ethical imperatives it confronts, a paradox that continues to animate contemporary discourse on artificial intelligence.

The modern AI movement originated after World War II, when Alan Turing first posed the question of whether machines can think (Groundup.Ai, 2025). The development of early computer science and the progression of machine learning subsequently rose. In the 1980s and 1990s, AI experienced a period of fluctuation known as the "AI winters," where increasing expectations did not match the disappointing progress. Despite this technological recession, the introduction of IBM's Deep Blue computer in 1997, the first to defeat a renowned global chess champion, Garry Kasparov, marked a turning point in history. It emphasised that AI is capable of performing better than humans, enabling increased computational power, data storage, and ultimately global reliance and widespread use of AI, which we see today. In particular, the investment in research and development of AI by nations, such as the US, China, and the EU, is part of a technological arms race to dominate AI innovation.

This paper examines the legal implications of AI, ethical concerns, societal impacts and proposed policy frameworks to address these issues. It also emphasises how rapid technological growth has outpaced governance, thereby contributing to uncertainty and potential harm. AI has become an integral component of modern life and must be developed in ways that comply with the standards of fundamental human rights that govern other technologies. Through case studies, international comparisons, and evaluation, this review provides valuable insights for interdisciplinary collaboration and inclusive frameworks to guide the development of AI toward positive societal outcomes.

Section I: Legal Implications of AI

1.1 Data Protection and Privacy

AI systems rely on comprehensive datasets, many of which include sensitive information such as personal, financial, health, location, and social media data to fuel their intellectual power. AI works like the human brain, relying on memory, which it later acts upon. Nevertheless, memory in this instance is referred to as data. It draws upon the information it obtains for extensive training and decision-making to mirror the human brain. However, it is designed to transcend human limitations by completing the same tasks with more efficiency and in a shorter time. Consequently, burgeoning issues arise regarding data privacy risks and invasive surveillance.

In the financial sector, artificial intelligence has been a transformative tool, facilitating market predictions, strategic decision-making, and efficient anomaly detection, such as unauthorised transactions and illicit activities (*AI In Finance: Applications, Examples & Benefits* | Google Cloud, n.d.). However, this heightened dependence on AI introduces critical financial risks. Systems that collect information such as credit card numbers, income levels, and bank account balances can become prime targets for data breaches and cyberattacks, allowing for the manipulation of financial data, unauthorised access to accounts, and identity theft. One significant insidious threat is 'data poisoning,' wherein hackers infiltrate businesses' confidential information and input incorrect data into the system, corrupt existing data, or delete crucial details (*Data Poisoning: Current Trends and Recommended Defense Strategies*, 2025). The underlying prospect of competitors using this approach to skew analysis, destabilise decisions, and misallocate capital could ultimately contribute to the inevitable failure of the business. In addition, breaches of this nature may provoke potential lawsuits from affected customers and investigations from the Information Commissioner's Office (ICO) or similar agencies, as well as regulatory fines, as the General Data Protection Regulation (GDPR) gives

the right to claim compensation from an organisation in the case of its breach of data protection law, and, finally, inflict long-lasting reputational damage (*ICO, n.d.*).

The central theme is that, despite AI's immense advances, the intensifying risk of data breaches is imminent. The deployment of facial recognition, location tracking, and biometric identifiers, including fingerprints, retinas, and voiceprints, has further intensified public fear over surveillance. George Orwell's *1984* exemplifies these fears by portraying a society void of anonymity and freedom in which every movement or expression is susceptible to recording, analysis, and unauthorised data storage. Privacy laws exist in certain jurisdictions, yet regulation remains incomprehensible and often unenforceable. This issue is particularly evident in the context of mass surveillance practices and the validity of consent obtained from individuals under surveillance. Clearview AI's facial recognition software highlights the regulatory challenges posed by such practices. The company collected billions of images from social media without user consent, leading to lawsuits from 2020 onward in the United States, the European Union, and Canada under the GDPR, Illinois's Biometric Information Privacy Act, and Canadian privacy law. Similarly, Google has faced allegations of continuing to track user activity in Incognito Mode, monitoring locations when location history was turned off, and collecting biometric data without proper consent. These practices have led to claims that the companies violated the Texas Capture or Use of Biometric Identifier Act (CUBI) and the Texas Deceptive Trade Practices Act laws.

This case highlights the threats posed by AI-driven data collection and illustrates how data exploitation infringes upon the fundamental right to privacy and informed consent. In practice, corporate assurances regarding data protection often contradict their actual operations, which expose users to unauthorised data collection and monitoring. To exacerbate this, Google is also involved in a series of other lawsuits, including accusations of using health data tracking for unauthorised personalised advertising and litigation over location history (Top Class Actions, 2025). Google reached a \$391.5 million settlement with a coalition of attorneys general from 40 states over privacy issues, thus heightening the scrutiny of its data practices. Privacy is an integral human right which one is entitled to. It is perceived as personal and therefore must be protected accordingly to safeguard it from unauthorised access.

Artificial intelligence is further cited as a tool within democratic governments to track citizens, suppress dissent, or control public behaviour. Former CIA systems analyst Edward Snowden revealed that the US National Security Agency (NSA) was involved in the collection of tens of millions of Americans' telephone records (BBC News, 2014). Under the surveillance programme known as *PRISM*, the NSA gained access to the servers of nine prominent technology firms, among them Facebook, Google, Microsoft, and Yahoo, to monitor digital communications. Although justified as a method of security, at what point will the line between privacy rights and state safety start to infringe on and blur the individual right to privacy?

Incidents of large-scale data collection and misuse demonstrate how data exploitation compromises both privacy and consent, challenging the very basis of human rights protections. They also highlight the need for legal safeguards such as the GDPR and the California Consumer Privacy Act (CCPA) to function as genuine mechanisms of accountability rather than symbolic frameworks detached from practical enforcement in an increasingly digital economy.

1.II Liability and Accountability

AI can help industrialise workflows and processes and, in particular, work autonomously. Nevertheless, the technology's complex nature has rendered the attribution of novel liability particularly challenging, thereby revealing significant gaps in legal and regulatory structures.

Contract law enables a claimant to file a contractual claim if the AI system is unfit for its intended purpose, is of unsatisfactory quality, or does not match the description (Consumer Rights Act 2015). These claims originate from statutory warranties or implied terms; however, there is legal conflict over whether an AI system qualifies as a "product" under consumer and commercial law. Defendants relying on contractual claims or legal liability face a significant risk, as the enforceability of their

claims is contingent upon the legal test of reasonableness, especially in business-to-business contracts.

AI's autonomous capabilities raise concerns in tort law. To establish negligence in court, the claimant must prove that the defendant is obligated to express a duty of care towards them due to their proximity relationship. This is difficult to establish when the manufacturer has ceded the task to the AI system and no longer manages it, meaning they can no longer prove they owe a duty of care to the claimant.

Furthermore, it is made difficult for the claimant to identify the cause of the harm in a case. To establish liability, the claimant must prove that the defendant's actions, or lack thereof, were the proximate cause of the conflict. However, attribution is difficult to discern due to modern technological systems. It is frequently unclear whether the failure originated from the manufacturer's hardware, a subsequent software update, or the artificial intelligence itself.

This issue is further complicated in autonomous systems that continue to develop after they are implemented. Their ability to gradually change their behavioural patterns causes unpredictability, making it difficult to identify the origin of harm, assess the foreseeability, and, ultimately, determine the defendant's liability.

This dilemma is epitomised in the case of Tesla's Autopilot, which has resulted in multiple fatal crashes and investigations by the National Highway Traffic Safety Administration (NHTSA) (Christensen, 2025). It elucidated the gaps in liability systems where courts and lawmakers struggled to assign responsibility among developers for obscuring Autopilot's capabilities, manufacturers for creating technologies that are ineffective in detecting and responding to various road conditions, or users for misusing the system or failing to adhere to warnings.

Tesla has prevailed in several court cases by arguing that Autopilot is solely intended as an advanced driver-assistance system (ADAS) and should not be used as a substitute for human driving (ESIMauricio, 2025). However, it has still faced significant scrutiny, and this does not diminish the underlying issue that AI is not only unreliable and deleterious to our fundamental human skills, but it has proven to be dangerous.

Moreover, contractual provisions may be scrutinised to ensure that liability is appropriately addressed and allocated within the terms of the agreement. As an initiative originally adopted in 1985, the European Commission has recently modernised the Revised Product Liability Directive to conform to the increasing presence of technology and digital services (*EU Adapts Product Liability Rules to Digital Age and Circular Economy*, 2024). Its scope has been widened by expanding the definition of a 'product' to explicitly encompass digital software, AI systems, and digital manufacturing files, subjecting manufacturers to strict liability for defective AI, regardless of whether they acted accordingly with safety regulations or negligently (*Revised Product Liability Directive | Think Tank | European Parliament*, n.d.). The Directive implies that the term "manufacturer" refers to both individuals and legal entities who develop, manufacture, or produce products. This aligns with current discussions advocating for the application of strict liability to autonomous and AI systems, particularly in situations where responsibility remains uncertain, such as in medical diagnostic tools, autonomous vehicles, and issues related to algorithmic bias. Under a strict liability regime, legal responsibility is assigned to manufacturers of inherently dangerous technologies, irrespective of evidence of fault or negligence. In addition, manufacturers are held liable for adverse repercussions of cybersecurity faults, including cyberattacks.

However, it can be argued that the directive relies on the traditional concept of a 'defect,' yet it undermines the ambiguity of the term within complex AI systems. The directive identifies AI systems as products; nevertheless, "defects" aren't necessarily explicitly identifiable in the way traditional product flaws are, given that, in most cases, defects do not stem from clear flaws but from unpredictable outcomes where the machine deviates from its intended purpose. This restriction can further weaken the users' claims in legal contexts due to the difficulty in proving the exact cause of harm. Despite the significant progression in adapting to modern technologies that the directive presents, it may require further refinement to address necessary legal gaps.

AI technological advances were not evaluated when traditional legal systems were developed. Subsequently, legal gaps are created where victims cannot hold defendants accountable, as AI systems often act unpredictably, functioning as 'black box' systems, making it difficult to establish fault, negligence, or intent. To mitigate these issues, contracts should require the implementation of frequent audit trails. Specifically, these encompass detailed records of transactions, project details, accounting records, and any documentary evidence of actions that have influenced the corresponding procedure (*What Is an Audit Trail? Everything You Need to Know*, n.d.). Focusing on this approach enables lawyers to explicitly delineate each party's involvement and responsibilities within the procedure. Audit trails may then be deployed to support or rebut claims in a product liability case and to verify that a business has not made any modifications beyond its established practices to address potential issues before they arise. Consequently, it helps protect the rights of users while building public trust in new technologies.

I.III Intellectual Property

Generative AI has become a prevalent component in modern industries, capable of creating visuals ranging from videos to static images, such as pencil and watercolour paintings, as well as photographs. The resulting products are compelling, exceeding the quality and capabilities of an average human. Data lakes and question snippets are utilised to process image and text archives, enabling the recognition of patterns and relationships and the generation of prompts. However, AI challenges the notion of Intellectual Property Law (IP) by creating legal ambiguity over ownership and limiting innovation due to the uncertainty regarding the preservation of rights and recognition. It has been increasingly capable of generating original works, as demonstrated in the Mauritshuis event. 'Midjourney' was used to produce an AI-generated variant of "Girl with a Pearl Earring" as a substitute for the original when it was loaned to a Vermeer exhibition. This event raises a fundamental question: "Who is the creator?" Is it the AI itself, the developers of the model, or even the artist who prompted the AI? Alternatively, should the use of AI be considered unacceptable, with the authorship reserved solely for the artist of the original piece? This section argues that the legal system is outdated, with intellectual property laws, particularly copyright and patent laws, considered obsolete when measured against the progression of generative AI.

One significant area is patent laws. In 2022, the US Court of Appeals for the Federal Circuit conceded that the DABUS AI system could not be recognised as an inventor on patent applications across several jurisdictions (*Thaler v. Vidal*, No. 21-2347 (Fed. Cir. 2022)). The central issue before the court jurors was whether non-human systems could qualify as inventors under existing IP law frameworks or if inventorship must remain exclusively human. The appeal was upheld, yet the case highlighted significant ambiguities within patent laws. Denying patent protection may discourage innovators from employing such systems for product development due to the absence of economic incentives or legal recognition. This limitation, therefore, has the potential to slow down progress in technological and scientific fields.

Another cause of concern is copyright issues. The use of generative AI has blurred the lines between inspiration and infringement, challenging the core principles of copyright law. It raises concerns regarding human authorship and automated content, making it difficult to establish who owns the legal rights to the work produced and whether AI-generated works can be protected under the law.

As previously discussed, generative AI are trained on extensive datasets composed of images and text from data archives. In practice, however, many of these datasets include copyrighted content obtained without explicit authorisation. Between 2020 and 2023, several copyright cases were filed against corporations, including OpenAI, Meta, and Stability AI, for allegations of using unauthorised copyrighted work to train their AI systems without clear legal guidance or user consent. In 2023, Getty Images filed a lawsuit against Stability AI, claiming that it was using unauthorised copyrighted images from Getty to train its AI models [2025] EWHC 38 (Ch). What distinguishes this case is Stability AI's rebuttal, which highlights and intensifies existing gaps in the legal system. Although

they acknowledged that their AI was trained using unlicensed images, they stated that the training occurred in the US, which is beyond the scope of UK laws. This case depicts a deeper legal challenge by highlighting the lack of clarity in cross-legal matters. Data sets remain international while enforcement frameworks are national, making it strenuous to regulate AI training and establish accountability for the misuse of unauthorised data, ultimately resulting in more legal confusion. The court has conceded that in the event Getty's claims are proven, the case may be dismissed entirely, as it occurred outside the UK's jurisdiction. This raises a significant concern that current legal protections are insufficient for regulating generative AI, whose impacts and developments are undoubtedly transnational in scope. As such, the case recognises both deficiencies in the UK legal system and a broader notion of governance for AI-generated content.

Additionally, the defendants claimed that the work was not a derivative of Getty Images but constituted a new and transformative design. However, current laws do not clearly define what qualifies as transformative regarding generative AI, particularly when outputs replicate elements from data patterns. Copyright law originated from the Patent and Copyright Clause of the Constitution, followed by the Copyright Act of 1976; it entails that all protected works must be both human and original.

Under the Copyright Act of 1976, originating from the Patent and Copyright Clause of the Constitution, all protected works must be both human and original. This precedent provides a foundation for asserting that works generated by AI are ineligible for legal protection or registration, as they lack human authorship. The U.S. Copyright Office likewise refuses to register works created without a human author (U.S. Copyright Office, 2017). Current legislation, however, remains ambiguous in its treatment of transformative works. It focuses solely on 'pastiche,' which refers to the imitation of the styles and character of another work, while offering no clear guidance on transformative creations that introduce new meaning or expression into existing works. This legal gap creates significant uncertainty regarding ownership, originality, and what constitutes infringement.

Currently, there is no definitive legislation defining whether the use of copyrighted work to train generative AI systems constitutes infringement, qualifies as fair use, or falls somewhere in between. This uncertainty highlights the need to revise existing legal frameworks so that copyright law more accurately reflects the involvement of AI in creative and inventive processes. Without clear guidelines, developing an inclusive and modern legal system becomes an ongoing challenge.

Section II: Ethical Implications of AI

II.1 Bias and Discrimination

AI-driven recruitment systems present a range of ethical risks that call for close inspection. Although these tools are designed to complete tasks efficiently and respond accurately to user input, they lack the emotional intelligence and contextual sensitivity inherent to humans. As a result, it can inadvertently perpetuate bias and discrimination, potentially leading to societal harm.

Bias in AI transpires when systems generate systematically biased outputs as a result of training on unjust data, leading to prejudiced treatment for certain communities. Notwithstanding manufacturers' justifications that their procedure precludes algorithmic bias, Miasato and Silva (2019) conclude that without human oversight and ethical precautions, algorithms alone cannot eradicate discrimination. In addition to COMPAS and Amazon's hiring tool, studies show that facial recognition misidentifies people of colour at significantly higher rates than white individuals (Buolamwini, n.d.). For instance, Robert Williams, a Black man in Michigan, was wrongfully arrested based on an inaccurate match by facial recognition software (Allyn, 2020). Given that the footage quality was grainy, it was already inherently unreliable. Moreover, Williams contended that the footage had no resemblance to him. This signifies the inequality gaps in AI systems, as they endorse the marginalisation and discrimination against minority groups.

The integration of AI technologies has become increasingly evident in various sectors of the digital economy. Employers are essentially interested in evaluating candidates' competitiveness when contemplating recruitment decisions. Consequently, this has led to a growing transition from traditional hiring to automated hiring. While this conversion offers numerous benefits, including increased efficiency and time optimisation, it also gives rise to algorithmic bias, which can result in systematic errors that evince discrimination based on race, gender and social inequality. "Predictive bias" occurs when the results for a certain group in evaluations differ from those of other groups that are otherwise identical on a specific criterion (Chegeveronica, 2023). However, this issue is disregarded under the pretence that findings concurred by AI are 'unbiased' and 'factual.' This highlights the underdevelopment and bias present in AI systems that impede the development in both technological and legal sectors and produce discriminatory outcomes that unjustly target certain minorities.

Furthermore, it hinders progress within the economic sector. For instance, when skilled individuals are excluded from organisations due to unfair bias on religion, ethnicity, or gender, it not only violates ethical practices but also reduces the overall efficiency of the business. It fosters an environment that lacks internal diversity, thereby diminishing innovation and reducing businesses' ability to navigate foreign markets and engage with consumers. Additionally, it challenges corporate social responsibility (CSR) practices, leading to a deteriorated brand reputation and loss of consumers who value equality and diversity. Referring back to predictive bias, this phenomenon can lead to job dissatisfaction, which erodes employee morale and productivity, therefore causing a decline in the operational efficiency of the economic sector. These factors, in turn, reduce the overall economic activity by limiting equal opportunity within the labour market. Heightened scrutiny has also been directed towards the discrimination and infringement of human rights surrounding the use of AI in surveillance contexts. Article 8 of the Human Rights Act 1988 protects the right to respect for private and family life, home, and correspondence, and applies in circumstances where individuals are identifiable in surveillance footage. This legislation is essential to ensure technological innovation aligns with human rights protection. However, how can Article 8 be effectively implemented to establish protection when surveillance tools can misidentify people, particularly minority groups, and lead to unjust arrests or targeting? This is evident in the China Initiative, implemented by the Department of Justice (DOJ) in 2018. It aimed to diminish intellectual property theft and address national security issues; however, it mainly racially profiled Chinese Americans. As a result, many false arrests were made, including UCLA graduate student Guan Lei, University of Tennessee professor Anming Hu, and National Weather Service scientist Sherry Chen (Chin-Rothmann & Lee, 2022).

These outcomes are not solely technical flaws; they illustrate prejudicial issues embedded in the real world and within data. The EU Artificial Intelligence Act is the world's first comprehensive legal framework designed to address the risks posed by artificial intelligence systems, governing the development and deployment of AI in the European Union. AI systems are categorised into levels according to their risk level. Systems that pose significant threats to the safety, rights, and livelihoods of users are classified as unacceptable risks and must therefore be banned. High-risk AI systems may pose a threat to the health, safety, or fundamental rights of users. Such systems must be subject to strict guidelines, including data logging, documentation, transparency, human oversight, and security measures. By contrast, limited risk systems require transparency regarding their operations and how they store information, whereas minimal risk poses no harm to users (Regulation (EU) 2024/1689). These measures serve to protect fundamental rights and promote technological innovation, contributing to greater security across the European Union.

However, these precautions require preliminary checks and developer self-assessments that may be insufficient in detecting or preventing subtle yet substantial biases. In particular, those that remain inconspicuous until the system is applied in real-world circumstances and irreversible damage has materialised. Similar to liability under the Revised Product Liability Directive, proving harm under bias is complex. Discriminatory outcomes may not be rooted in direct faults but formed

through systematic patterns of inequalities established through training via biased data. Subsequently, it is challenging for the claimant to establish the correlation between unjust practices and AI systems, weakening their legal claims despite the factual evidence of injustices. While legal frameworks acknowledge the risk of bias and establish initiatives to address these challenges, they require more robust enforcement mechanisms alongside more explicit guidelines to prevent inadvertent consequences that hinder the effective use of AI in sectors and to ensure legal accountability for developers.

II.II Moral Status and AI Rights

Following recent developments, debates centred around the moral standing of artificial intelligence have grown increasingly significant. Saudi Arabia has notably granted citizenship to a robot (Weller, 2017), while the European Parliament is drafting a form of “electronic personhood” for artificial intelligence. These verdicts spark a growing concern in legal and ethical discourse: although AI lacks consciousness, discussions about governments' attempts to propel AI beyond being merely a tool are gaining prominence. As AI develops greater capabilities, such as making complex decisions and feigning emotions, the question of whether machines should be subject to the same rights and moral status as humans becomes conspicuous. To illustrate this debate, Immanuel Kant's theory is instructive. Kant distinguishes between rational agents who possess rationality and autonomy and are therefore justified in being treated morally and nonrational entities that lack sophisticated cognitive capacities. This section explains the extent to which AI should be recognised within our moral framework and the consequences of the widespread use of emotional artificial intelligence.

Scheessele (2018) presents a pyramid, divided into four regions, to illustrate the ethical relevance of artificial agents. Each area represents the moral consideration an entity is entitled to and the degree of responsibility it should bear for its actions. At the top is “Full Moral Status,” which refers to entities capable of being both moral agents and moral patients. According to Scheessele, a moral patient must be the object of moral concern and interests that are rightfully protected. Additionally, a moral agent must possess the capacity for ethical reasoning and accountability. Below the highest region are three subordinate levels: SF “Significant Full Moral Status,” MS “Minimal-Significant Moral Status,” and NM “Negligible Moral Status.” The threshold rule implies that if an entity qualifies for one level, it warrants all the levels below it.

Building on Scheessele's model, the question of whether AI can occupy the higher tiers remains philosophically contentious. While some philosophers proclaim they are capable of achieving moral standing, Scheessele argues that contemporary systems fail to achieve the necessary provisions.

AI lacks the fundamental prerequisites inherent to moral agency and patience, as it is bereft of the capacity to engage in independent moral deliberation and is without conscious awareness or individual perceptiveness. Rather than acting intentionally, it mimics behaviour by being trained using existing databases, bearing no genuine feelings, and being incapable of suffering, desiring, or possessing any rationality. Its lack of sentience undermines their claim to moral relevance and, by extension, precludes them from the higher tiers of Scheessele's moral status framework. Consequently, if AI cannot obtain Full Status, the threshold implies it cannot meet the standards for Significant Full Moral Status. This illustrates that AI does not possess the ethical obligations required for the prevalent delegation of morally sensitive tasks and its accountability for such.

Notwithstanding its advanced capabilities, it should remain within the negligible to minimum regions. Conferring moral status to artificial technologies blurs the line between humans and technology, leading to the erosion of human rights and legal confusion. If AI is considered a moral agent, who is to blame for the malfunction of a program? The anthropomorphism of machines references the concerns highlighted in Intellectual Property, thereby transcending both ethical and legal frameworks and rendering such treatment unjustified.

Another area drawing concern surrounding AI is its ability to manipulate users by bypassing rational cognition and eliciting emotional responses. In the UK, a study was conducted by using ten qualitative focus groups to assess how 46 people discern emotionally intelligent technology systems

in realistic scenarios (Bakir et al., 2024). The findings indicated a general concern regarding systems that employ emotional profiling to analyse users' psychological and emotional cues. One consequential form of digital manipulation is the creation of deepfakes. These are exceptionally lifelike images or videos generated through artificial intelligence to reproduce the likeness and mannerisms of real individuals with remarkable precision. The increasing presence of such fabrications intensifies concerns about their capacity to amplify misinformation and conspiracy narratives, damage personal and professional reputations, and exploit human emotions. For instance, during the January 2024 election in New Hampshire, voters received calls from what appeared to be President Biden, instructing Democrats to hold their votes (Bond, 2024). It was later revealed that it was a deepfake created with AI. Before remedial information was provided, many voters had made decisions influenced by the emotional manipulation of AI by the time the deepfake was revealed to the public. This incident exemplifies the broader concern that emotionally intelligent systems do not merely respond to human emotions but elicit, exploit, and redirect them, often without human consent.

Another concern raised by the study was regarding "emtoys," children's toys that could interpret and respond to children's vocal and facial expressions. Respondents expressed their concerns focused on how emotional profiling exploited the vulnerability of children's emotional and cognitive immaturities. The extent to which this threat manipulates the children's emotions and develops behavioural patterns remains underlying and raises the question, "Should these interactions be regulated to prevent psychological manipulation?" The findings underline a public demand for civic protections, ethical oversight, and mechanisms to safeguard trust in emotional AI systems.

This section has thus far discussed the degree to which AI can impose harm on humans; however, it has failed to mention how humans can willingly inflict harm on themselves through their ignorance concerning artificial intelligence. A cross-sectional study was conducted with a sample of 872 intellectual individuals above the age of 18 years, 55% of whom preferred AI-based psychotherapy (Aktan, Turhan, & Dolu, 2022). While this preference can be justified with AI's constant accessibility, financial convenience, and lack of judgement, it does not take into account that AI's responses are not genuine. Statistical patterns in language and user behaviour guide how these systems generate responses. This process does not involve empathy, understanding, or compassion, all of which AI systems are incapable of feeling. Users may misinterpret these responses as an indication of emotional intelligence or maturity, yet they fail to acknowledge that they are an outcome of personalised data modelling designed to imitate human reasoning. This false pretence of understanding allows users to accept biased and misinformed data, which can ultimately shape beliefs and be dangerously used in contexts such as mental health.

A psychiatrist at Stanford said that the psychotherapy provided by chatbots is worsening delusions and causing significant harm (*New Study Warns of Risks in AI Mental Health Tools*, n.d.). For instance, hyperdependence on AI emotional systems can gradually pose a substitute for human companions, exacerbating loneliness and leading to social derealisation if individuals feel that face-to-face interactions lack emotion or feel surreal in comparison to AI conversations. This is evident when OpenAI, a widely utilised tool for psychotherapy, announced its decision to revoke its latest update in April 2025 due to concerns that the model exhibited excessive sycophantic behaviours, which induced distress from users (*OpenAI Rolled Back a ChatGPT Update That Made the Bot Excessively Flattering*, 2025). This circumstance highlighted the growing concern that AI emotional interactions can inadvertently distort users' sense of reality, especially in contexts where they may be vulnerable.

Moreover, ChatGPT explicitly states that it can personally identify you and can share all your data in its privacy policy and terms and conditions. However, users remain oblivious to such a reality, as they often overlook the importance of terms and conditions and persist in sharing sensitive information without considering the consequences of their data being stored, analysed, and repurposed.

This accentuates that the danger does not solely lie with AI's ability to perform emotional profiling, but also in humans' willingness to grant it emotional authority due to their unawareness. It further demands regulatory frameworks to prioritise psychological safety in the design of AI systems beyond technical accuracy.

II.III *Autonomy and Human Agency*

A central concern that arises amongst individuals is that the proliferation of AI technologies contributes to an alienated population. As Lisanne Bainbridge states in her research paper *Ironies of Automation*, automation frequently coerces humans into monitoring roles without practice. By assuming that the AI has emotional depth or logical reasoning akin to a human's, a growing reliance on technologies is fostered, which erodes critical human skills and impairs or replaces human judgement. This reliance becomes deleterious in areas such as law or healthcare, where misinformed decisions can have significant adverse consequences for which workers must be held legally liable.

This phenomenon is evident in a study conducted in Germany, where radiologists were asked to interpret BI-RAD scores for mammograms (*AI Bias May Impair Radiologist Accuracy on Mammogram*, n.d.). AI-generated scores were also provided, with some being accurate while others remained deliberately incorrect. When radiologists relied solely on their judgements, their answers aligned with the correct AI scores. However, reliance on AI-generated scores resulted in a diagnostic accuracy of less than 20%. The case highlights a dangerous pattern of both automation bias and diminished human skills. The doctors' accurate judgements were undermined due to their trust in AI systems, disregarding that the systems are trained on real clinical data and can recapitulate errors or biases. Over time, such overdependence may erode expert diagnostic capabilities in scenarios where human intuition is necessary. As a result, patients may experience complications from incorrect treatments, face delays in essential procedures, or undergo unnecessary surgeries due to AI misidentifying symptoms. These failures carry significant ethical and legal implications, including liability lawsuits, which ultimately raise substantial questions regarding accountability.

A key aspect of effective worker-consumer relationships is the preservation of core values such as integrity and transparency. They must ensure workers possess the trust and empathy necessary to make informed decisions aligned with customers' needs while maintaining core values, an aspect often diminished following the pursuit of AI. Moreover, with AI requiring access to personal data, the importance of safeguarding data privacy becomes paramount. AI technologies are meant to enhance decision-making processes, not substitute human judgement. Ethically, users must be entitled to a say in decisions that impact their lives, as this transition may lead to a sense of disempowerment or loss of control.

As the use of artificial intelligence becomes increasingly embedded in daily life, whether through social media, AI-generated content, or emotional detection, it begins to instil potentially detrimental habits in an individual. These habits are formed not solely based on how an individual processes or manages information; they also shape their behaviour, influence behavioural patterns, and affect self-perception.

Constant exposure to AI-generated content perpetuates unrealistic standards of perfection and the need to conform to them, leaving little room for human imperfections. In time, this can distort a person's sense of reality or fracture their identity by internalising unattainable expectations and potentially instigating various forms of insecurities.

Recent developments in AI leave users vulnerable to algorithmic nudging, where AI collects data such as preferences, activity, and user behaviour to deploy personalised systems ranging from social media recommendations to behavioural targeting. In practice, Facebook's algorithm has similarly been shown to prioritise content that elicits strong emotional reactions. However, these harmful recommendations and feedback can reinforce negative beliefs, including anxiety, depression, or low self-esteem, consequently rapidly worsening individuals' mental states. By prioritising engagement over users' emotional safety, it creates a toxic environment that induces fear

or anxiety in users, with internal documents revealing staff within Facebook acknowledging the potential for increased spam, abuse, and clickbait (Deck, n.d.).

These personalised recommendations, or nudges, are designed to subtly manipulate preferences and influence choices while inadvertently exploiting cognitive illusions. This notion refers to systematic errors that alter our judgement, perspectives, and behaviours in ways that feel rational, ultimately distorting the way we act and think. The human brain relies heavily on the information it receives, filtering and rearranging it to establish a formulated outlook on reality (Fitzpatrick, 2024). Prolonged exposure to systematic biases begins to stray our thoughts from the normative principles of logic into a false perception of reality under the pretence that interactions are personal and therefore must be definitively correct; also known as the illusion of objectivity. A common misbelief is that AI systems are unbiased and accurate, as they are trained using fair data and algorithms. In reality, AI adapts its behaviour based on user interaction. These systems identify which content users engage with negatively or positively and continue to endorse the content that appears most favourable. This essentially creates a feedback loop, allowing systems to become more personalised and helpful, which in turn further strengthens users' trust in the data provided. Nevertheless, AI systems can be trained on biased data, subsequently establishing and reinforcing biases in human thinking that are both arduous to identify and unlearn (*Bias in AI*, n.d.).

For instance, recommendation algorithms on social media may continue to perpetuate extreme or harmful content that appears neutral. A 2021 investigation by the Wall Street Journal on TikTok algorithms serves as a prime example. It conceded that once a user interacts with videos on particular subjects, the algorithm begins to recommend content related to those same areas (WSJ Staff, 2021). Although it creates a more enjoyable experience for users when engaging with content, including conspiracy theories, mental health, or specific political topics tailored to their preferences, the underlying issue remains that algorithms are falsely presented as merely objective and can narrow individual perspectives or establish negative behavioural patterns. This issue is made exceptionally worse for unsupervised children and adolescents whose neurological development and behavioural patterns are still being established, leaving them at a significant risk of the psychological effects of AI on human agency.

Section III: Societal Impact and Cross-Cutting Challenges

III.1 Automation and Employment

The swift emergence of artificial intelligence has sparked increasing concern within the workforce, particularly surrounding job stability, economic inequality, and ethical responsibility. As AI continues to advance and integrate across industries, it begins to reshape work dynamics and dictate who has access to employment and under what provisions.

The core of this debate is centred around the complex role of artificial intelligence, which creates opportunities for technological progress yet presents significant challenges to employment and economic stability. Contemporary AI leverages higher or similar human cognitive abilities to perform repetitive tasks and complete work to a standard higher than that of humans (Tyson & Zysman, 2022). This advancement has led to AI machines being employed in lieu of human workers in fields heavily reliant on repetitive functions and manufacturing, consequently increasing unemployment rates and displacing human workers. While high-skilled jobs are rapidly expanding, displaced workers lack the fundamental skills and time to access them.

For instance, following an increase in factory automation, the American Midwest is facing increased redundancies and economic stagnation after being a major manufacturing region (Leonard, 2019). In contrast, urban tech hubs like San Francisco experienced economic booms driven by technological innovation. This employment divide entails how automation is replacing routine manufacturing jobs, leaving many individuals unemployed and making it increasingly difficult to find opportunities in AI-centric industries. In contrast, jobs relying on high-skilled workers and capital to drive innovation and job creation in AI, software, and digital services remain unaffected.

This induces a stark illustration of how automation leads to the polarisation of a fragmented labour market.

A clear analysis of the relationship between innovation and job creation in AI reveals that, bound by certain qualifications, automation can offer job opportunities. The increasing integration of AI software in industries can augment productivity and efficiency by performing tasks traditionally done by humans but executed with enhanced precision, speed, and efficiency. Additionally, it is used to execute repetitive tasks, allowing human workers to concentrate their efforts on more complex responsibilities that require creativity, critical thinking, and sustained effort. While this approach can enhance overall business productivity, it is exceptionally optimal for employers. Automation lowers labour costs and improves efficiency, which accrues primarily to owners who prioritise maximum profit, often at the expense of employee welfare. This perspective has been reinforced by governmental support, including R&D tax incentives, which tend to favour labour-saving approaches rather than those prioritising strengthening employee capabilities and job creation. As a result of tax policies favouring capital over labour, owners are further inclined to adopt automation in the workforce.

This shift in jobs thereby creates a growing demand for skilled workers to oversee the management of AI systems. Cities increasingly depend on intellectual workers to attract investment and secure funding for the development of AI infrastructure and related technologies. As AI becomes more prevalent, a range of new professional roles has emerged. These include data scientists who analyse complex data patterns to make informed decisions, AI trainers responsible for refining system responses to improve their interactions with humans, as well as specialised researchers and machine learning engineers.

However, this transformation presents notable implications, particularly the emergence of a widening skill gap across the labour market. To remain competitive, workers must acquire new expertise to align with the technological demands of AI integration. Those lacking digital literacy or advanced technological skills struggle to keep up with the increasing pace of technological advances. AI-centric companies require intellectual employees familiar with technical expertise to navigate AI systems. In the absence of that, companies are prone to decreased efficiency and increased susceptibility to errors. A constructive course of action would be to focus on training and providing technological support to employees; however, it is time-consuming and costly. This is less optimal for companies prioritising profit maximisation, risking leaving millions behind, as well as the overall efficiency of the business.

III.II AI in Education

Artificial intelligence has introduced notable advancements in education, from personalised learning and administrative efficiency to greater accessibility and inclusivity. By offering tailored instruction, automating routine tasks, and expanding access to quality resources, AI is reshaping how and where learning occurs. However, these innovations also bring complex challenges that warrant critical examination.

Nonetheless, one notable impediment to AI integration in education is the lack of interpersonal interaction between teachers and students. Fostering effective educational environments is rooted in building an emotional connection with students, enabling educators to recognise subtle signals, such as body language and facial expressions. While AI can automate tasks and deliver personalised content, it lacks the emotional intelligence that teachers possess to detect emotions such as distress or fear. In classrooms, teachers do not solely provide information; they act as emotional support systems that encourage and navigate students through challenges. Their human empathy contrasts with the lack thereof in AI systems, which prioritise efficiency and accuracy, limiting their ability to build personal connections and cultivate students' creativity. In the absence of consistent human interaction, students may substitute human interaction for the companionship of AI, ultimately leading to a decline in social interactions.

This decline can be detrimental as it hinders social skills alongside emotional intelligence, factors essential for academic success and students' welfare. Studies insinuate that 35% of students feel isolated in AI-driven learning environments (DigitalDefund, 2025). These factors can decrease their ability to work effectively in teams or group discussions or build connections in friendships, consequently fostering a feeling of loneliness and an eroded sense of belonging. Such challenges are exacerbated for students who already struggle with social interactions, as substantial loneliness brings rise to decreased confidence in social settings and thus leaves them inadequately equipped for the collaborative skills essential for academic and professional contexts. Moreover, this overreliance on technology can affect the quality of students' learning, as it can give rise to mental health issues such as detachment and anxiety. Students experiencing these challenges may find it difficult to contribute, concentrate, or stay motivated. The absence of friendships that students are unable to establish due to their inadequate social skills may result in a loss of interest in attending school, further harming their education.

To reference the data issues associated with AI systems previously mentioned, it is pertinent to note that extensive implementation of artificial intelligence in the educational industry involves risks to the data provided by parents and pupils. It is standard for schools and teachers to be familiar with the sensitive data regarding students, including personally identifiable information (PII) and health and medical information. Nevertheless, when AI systems are introduced into educational contexts, they are inevitably exposed to and analyse sensitive student information. This heightened exposure significantly increases the pertinence of the Family Educational Rights and Privacy Act, 1974 (FERPA), a federal law intended to safeguard students' educational records and allow parents the right to control the disclosure of personally identifiable information. In the instance where a teacher employs third-party AI services to grade students' work that includes their name and other identifying information without parental consent, the teacher has ultimately breached their duty of care by providing personally identifiable information to non-FERPA-compliant services. These misconducts can lead to hacking, cyber harassment, and data breaches, all of which tarnish the institution's reputation, increase its risk of legal repercussions, and compromise students' safety. In compliance with GDPR and FERPA laws, the responsibility centres on educators and academic institutions to maintain equitable practices by obtaining consent from students and parents before using third-party AI services that store personal student information.

Conclusion

This paper investigates the intricate relationship between the legal, ethical, and social dimensions of artificial intelligence, which continue to evolve in scope and significance. Concerns regarding privacy, liability, algorithmic bias, and labour displacement highlight the need for transparent systems of governance. With the rapid pace of technological innovation, the scale of transformation required to regulate AI systems remains a significant task. This challenge becomes even more pronounced when objectives extend beyond economic growth to the pursuit of transparency, accountability, and the equitable application of emerging technologies. As artificial intelligence lacks consciousness or moral awareness, it is crucial to recognise that automated decision-making processes may still produce harmful consequences, whether arising from flawed design or unanticipated effects. Incorporating bioethical reasoning is therefore essential to address the ontological gap of the non-moral nature of artificial systems and their inability to sympathise. Technology cannot replace human oversight and judgement, nor should autonomous systems continue learning while in operation, as unsupervised change may result in harmful or legally ambiguous outcomes. Corporations share a collective responsibility for the responsible development of AI through globally coordinated legislation that embeds principles of justice, accountability, and equity into its design and deployment to mitigate the risk of racial, age-based, or cultural inequities.

The trajectory of artificial intelligence remains contingent on how effectively these measures are implemented in real-world contexts. When grounded in law, ethics, and shared human values, AI can function as a transformative tool. AI has reshaped fields such as healthcare, education, and

industry, demonstrating both its adaptability and reach. Its capacity to enhance efficiency and decision-making underscores its substantial social and economic value. However, without regulation, it risks reinforcing cultural divides or acting as a mechanism of surveillance that undermines personal security. Artificial intelligence is no longer an abstract concept, but a permanent feature of contemporary life. The central goal is to establish a reliable means of integration to protect human welfare. Future research should therefore conduct comparative analyses across sectors and implement robust methods for evaluating AI's prolonged social and economic influence.

Addressing these concerns demands sustained interdisciplinary collaboration among technologists, ethicists, and policymakers. The aim is not merely to pursue innovation but to ensure that the progress of AI aligns with human values and social responsibility. Through deliberate cooperation and reflection, AI can be guided to strengthen the public good rather than undermine it.

References

1. *Introduction to artificial intelligence (AI) - Groundup.ai.* Groundup.ai. <https://groundup.ai/resources/introduction-to-artificial-intelligence-ai/>
2. *AI in Finance: Applications, Examples & Benefits | Google Cloud.* (n.d.). Google Cloud. <https://cloud.google.com/discover/finance-ai>
3. *Data Poisoning: current trends and recommended defence strategies.* (2025, June 24). wiz.io. <https://www.wiz.io/academy/data-poisoning>
4. ICO. (n.d.). *Taking your case to court and claiming compensation.* <https://ico.org.uk/for-the-public/data-protection-and-journalism/taking-your-case-to-court-and-claiming-compensation/>
5. Actions, T. C. (2025, June 30). *Google must face privacy class action over tracking users' health data.* Top Class Actions. <https://topclassactions.com/lawsuit-settlements/lawsuit-news/google-must-face-privacy-class-action-over-tracking-users-health-data/>
6. BBC News. (2014, January 17). *Edward Snowden: Leaks that exposed US spy programmes.* <https://www.bbc.com/news/world-us-canada-23123964>
7. United Kingdom. (2015). *Consumer Rights Act 2015, c. 15.* <https://www.legislation.gov.uk/ukpga/2015/15/contents/enacted>
8. Christensen, D. (2025, May 6). *Tesla Autopilot accidents: legal rights and liability.* Team Justice. <https://teamjustice.com/tesla-autopilot-accidents-legal-rights-liability/>
9. *EU adapts product liability rules to digital age and circular economy.* (2024, December 9). European Commission. https://commission.europa.eu/news-and-media/news/eu-adapts-product-liability-rules-digital-age-and-circular-economy-2024-12-09_en
10. *Revised Product Liability Directive | Think tank | European Parliament.* (n.d.). [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2023\)739341](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)739341)
11. *What is an audit trail? Everything you need to know.* (n.d.). <https://auditboard.com/blog/what-is-an-audit-trail>
12. *Thaler v. Vidal, No. 21-2347 (Fed. Cir. 2022)*
13. [2025] EWHC 38 (Ch)
14. U.S. Copyright Office. (2017). *Compendium of U.S. copyright office practices* (3rd ed.). <https://www.copyright.gov/comp3/chap300/ch300-copyrightable-authorship.pdf>
15. *Masiato, S., & Silva, J. R. (2019). Artificial Intelligence as an Instrument of Discrimination in Workforce Recruitment, Scientific Journal of Sapientia Hungarian University of Transylvania, 8(2), 191-212.*
17. Buolamwini, J. (n.d.). *If you're a darker-skinned woman, this is how often facial-recognition software decides you're a man – MIT Media Lab.* MIT Media Lab. <https://www.media.mit.edu/articles/if-you-re-a-darker-skinned-woman-this-is-how-often-facial-recognition-software-decides-you-re-a-man/#:~:text=A%20new%20study%20from%20MIT,one%2Dthird%20of%20the%20time.>
18. Allyn, B. (2020, June 24). "The computer got it wrong": How facial recognition led to false arrest of Black man. *NPR.* <https://www.npr.org/2020/06/24/882683463/the-computer-got-it-wrong-how-facial-recognition-led-to-a-false-arrest-in-michig>

19. Chegeveronica. (2023, July 6). *Introduction to Predictive bias*. Medium. <https://medium.com/@chegeveronica/introduction-to-predictive-bias-701ce1b995b9>
20. Human Rights Act 1998, s. 8.
21. Chin-Rothmann, C., & Lee, N. T. (2022, April 7). Police surveillance and facial recognition: Why data privacy is imperative for communities of color. *Brookings*. <https://www.brookings.edu/articles/police-surveillance-and-facial-recognition-why-data-privacy-is-an-imperative-for-communities-of-color/>
22. (Regulation (EU) 2024/1689)
23. Weller, C. (2017, October 27). *A robot that once said it would "destroy humans" just became the first robot citizen*. Business Insider. <https://www.businessinsider.com/sophia-robot-citizenship-in-saudi-arabia-the-first-of-its-kind-2017-10>
24. Scheessele, M. R. (2018). *A framework for grounding the moral status of intelligent machines*. PhilArchive.
25. Bakir, V., Laffer, A., McStay, A., Miranda, D., & Urquhart, L. (2024). On manipulation by emotional AI: UK adults' views and governance implications. *Frontiers in Sociology*, 9. <https://doi.org/10.3389/fsoc.2024.1339834>
26. Bond, S. (2024, December 21). How AI deepfakes polluted elections in 2024. *NPR*. <https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections>
27. Aktan, M. E., Turhan, Z., & Dolu, İ. (2022). Attitudes and perspectives towards the preferences for artificial intelligence in psychotherapy. *Computers in Human Behavior*, 133, 107273. <https://doi.org/10.1016/j.chb.2022.107273>
28. *New study warns of risks in AI mental health tools*. (n.d.). Stanford University. <https://news.stanford.edu/stories/2025/06/ai-mental-health-care-tools-dangers-risks>
29. *OpenAI rolled back a ChatGPT update that made the bot excessively flattering*. (2025, April 30). NBC News. <https://www.nbcnews.com/tech/tech-news/openai-rolls-back-chatgpt-after-bot-sycophancy-rcna203782>
30. Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779. [https://doi.org/10.1016/0005-1098\(83\)90046-8](https://doi.org/10.1016/0005-1098(83)90046-8)
31. *AI bias may impair radiologist accuracy on mammogram*. (n.d.). RSNA. <https://www.rsna.org/news/2023/may/ai-bias-may-impair-accuracy>
32. Deck, A. (n.d.). *More internal documents show how Facebook's algorithm prioritized anger and posts that triggered it*. Nieman Lab. <https://www.niemanlab.org/2021/10/more-internal-documents-show-how-facebooks-algorithm-prioritized-anger-and-posts-that-triggered-it/>
33. Fitzpatrick, O. (2024, September 17). *The Cognitive Illusions: How your brain shapes reality*. Owen Fitzpatrick. <https://owenfitzpatrick.com/blog/the-cognitive-illusions-how-your-brain-shapes-reality/>
34. *Bias in AI*. (n.d.). Chapman University. <https://www.chapman.edu/ai/bias-in-ai.aspx>
35. WSJ Staff. (2021, July 21). *Inside TikTok's algorithm: A WSJ video investigation*. *The Wall Street Journal*. <https://www.wsj.com/tech/tiktok-algorithm-video-investigation-11626877477>
36. Laura D. Tyson, John Zysman; *Automation, AI & Work*. *Daedalus* 2022; 151 (2): 256–271. doi: https://doi.org/10.1162/daed_a_01914
37. Leonard, M. (2019, December 2). *Study: Increased use of robotics led to worker displacement in Midwest*. *Supply Chain Dive*. <https://www.supplychaindive.com/news/study-increased-use-of-robotics-led-to-worker-displacement-in-midwest/568281#:~:text=Dive%20Brief,of%20manufacturing%20in%20the%20country.>
38. DigitalDefynd, T. (2025, June 25). *30 Pros and Cons of using AI in Education [Detailed Review] [2025]* - DigitalDefynd. DigitalDefynd. <https://digitaldefynd.com/IQ/ai-in-education-pros-cons/>
39. *Family Educational Rights and Privacy Act*, 20 U.S.C. § 1232g (1974).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.