BEYOND THE ALGORITHM: RECLAIMING HUMANITY IN THE AGE OF ARTIFICIAL INTELLIGENCE

Ayushi Trivedi[1] and Naomi Hannah Cherian[2]

VOLUME 1, ISSUE 2 (JULY- DECEMBER 2025)

ABSTRACT

Artificial intelligence is no longer science fiction. It's making hiring decisions, diagnosing illnesses, and even creating "art." But what happens when this dependence backfires, as the reliance on AI only rises? As AI systems outpace the laws meant to govern them, we're left to decide who is liable for the damage caused by AI.

This paper discusses the paradox of this tool, designed to assist humankind with menial tasks, so that it too often reinforces our worst biases. From racist facial recognition systems to chatbots that spew libel, the failures of AI aren't mere glitches but a reflection of our societal flaws. Hayao Miyazaki, a Japanese artist, called AI-generated art "an insult to life itself, "unfolding a more profound truth within people: when we outsource creativity and judgment to machines, we risk losing what makes us human.

People are already paying the price. Job seekers filtered out by biased algorithms, patients misdiagnosed, and artists whose livelihoods are undercut by synthetic content are just a handful of the ill effects of artificial intelligence. The law, meanwhile, cannot keep up with such technological escalation.

The way forward requires determination through mandatory bias testing, human oversight strictly ingrained into AI systems, and global rules that put people over profit. The use and development of artificial intelligence are inevitable. The goal isn't to eradicate AI but to force it to serve us and prevent the other way around. This isn't a debate about technology, but what kind of future we will strive for.

Keywords: Artificial Intelligence, Human rights, Innovation, Transparency, Liability

[1] Ayushi Trivedi, *School of law, CHRIST (Deemed to be) University, Bangalore.*
[2] Naomi Hannah Cherian, *School of law, CHRIST (Deemed to be) University, Bangalore.*

INTRODUCTION

AI's increasing use raises concerns such as: who is accountable when it causes harm, and how do we prevent it from violating human rights? With the changing times, artificial intelligence has emerged from a science fiction concept and is now mainstream reality. Murat Durmus, in his book Beyond the Algorithm, said, "In every whisper of the algorithm, there is an echo of human thought, blurred and distorted, like a philosopher's dream meandering through the night".[3] His words remind us that AI is never neutral – it reflects human choices, data, and biases, even when disguised in the language of objectivity and code. Even the increasing technology deeply ingrained in our social, legal, and economic fabric is evident from the recent increased popularity of generative AI tools, which range from content makers to predictive policing software's and medical diagnoses. Every innovation brings changes and many significant questions, such as "are these tools being created and applied with sufficient ethical consideration, or are they being used without foresight of societal ramifications. This algorithmic neutrality hides a deep-rooted prejudice and raises concerns about democratic accountability, transparency, and monitoring, which need to be considered. On the recent trend of transforming personal photos into studio Ghibli art using AI tools, experts warn that the trend conceals a darker reality where casual sharing can lead to unforeseen privacy breaches and data misuse.[4]

The founder of the Ghibli Art, Miyazaki Hayao stated that the idea for artificially generated art with strong contempt. "I don't want to associate this with our work; it feels like a huge insult to life. It feels like the end of the world is near. We humans are losing faith in ourselves".[5] His words resonate beyond aesthetics; they reflect a deep unease about the erosion of human creativity, emotion, and purpose in an age where machines increasingly simulate what was once considered uniquely human. When we allow algorithms to replace imagination, decision-making, and empathy, we risk technical malfunction and a profound moral and cultural collapse. This raises concerns about humanity and artificial intelligence. It calls attention to the ethical, legal

[3] Murat Durmus, Beyond the Algorithm: An Attempt to Honor the Human Mind in the Age of Artificial Intelligence (Wittgenstein Reloaded), *Medium* (Jan. 24, 2024), https://murat-durmus.medium.com/beyond-the-algorithm-an-attempt-to-honor-the-human-mind-in-the-age-of-artificial-intelligence-c0b09542dd2b.

[4] PTI, Studio Ghibli AI Art Trend: A Privacy Nightmare in Disguise, Experts Warn, *Econ. Times* (Apr. 6, 2025), https://economictimes.indiatimes.com/tech/artificial-intelligence/studio-ghibli-ai-art-trend-a-privacy-nightmare-in-disguise-experts-warn/articleshow/120035607.cms.

[5] Eugenie Shin, An Insult to Life Itself: Ghibli-Style AI Images Raise Ethical Concerns, *Tokyo Weekender* (Mar. 31, 2025), https://www.tokyoweekender.com/japan-life/news-and-opinion/ghibli-style-ai-images-raise-ethical-concerns/.

and social tensions that arise when innovation outpaces regulation, and when human judgment is sidelined in favour of algorithmic efficiency.

HISTORICAL CONTEXT OF AI

Humanity was never intended to be replaced by artificial intelligence. Its early proponents saw technology as a tool to enhance human potential rather than replace it. Building systems that could use language, create abstract ideas, and solve problems in ways that resemble human intellect, especially for the benefit of humans. This was the goal of the historic 1956 Dartmouth Conference, where John McCarthy first used the term "artificial intelligence."[6] It was not a means to replace human reasoning or creativity, but rather, it freed individuals from monotonous, repetitive work and freed up brain energy for more complicated tasks.

Practical demands served as the foundation for this human-centric design philosophy. Machines like Alan Turing's theoretical 'Turing Machine'[7] and Leonardo Torres y Quevedo's 'El Ajedrecista'[8] were developed in the early and mid-20th century to facilitate specific, limited human pursuits like chess or mathematical computing. These devices weren't self-governing agents; they were innovative tools. Rather than emulating the entire range of human intellect, early AI pioneers were more concerned with resolving specific technological issues. In other words, AI was never meant to replace our mental abilities, but rather to complement them.

This subservient role was even highlighted in literary representations. "The Engine", a primitive idea similar to modern language models, was first presented by Jonathan Swift in his 1726 Gulliver's Travels.[9] It was designed to rearrange words and generate ideas. Because it augmented human creativity rather than completely replaced it, employing robots to control language was intriguing both then and now. Similarly, a picture of robots as workers rather than intellectuals or decision makers was portrayed in the Czech playwright Karel Čapek's 1921 play R.U.R.,[10] where he coined the term "robot." Robota translates to "forced work," suggesting the element of servitude rather than independence. As technological capabilities improved and AI's geopolitical

[6] Dartmouth College, A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (Aug. 31, 1955), http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf.
[7] The Turing Machine, *High-Performance Embedded Computing* (2d ed. 2014), *ScienceDirect*, https://www.sciencedirect.com/topics/computer-science/turing-machine.
[8] Torres Quevedo Invents El Ajendrecista, the First Decision-Making Automation, *History of Information*, https://www.historyofinformation.com/detail.php?id=569.
[9] Chris Garcia, Gulliver's Engine, *Computer History Museum* (Nov. 28, 2012), https://computerhistory.org/blog/gullivers-engine/.
[10] John Jordan, The Czech Play That Gave Us the Word 'Robot', *MIT Press Reader* (Jul. 29, 2019), https://thereader.mitpress.mit.edu/origin-word-robot-rur/.

and economic utility became significant, AI rose in status. A concrete shift occurred in the late 20th century when business and military interests began intertwining with AI development. In addition to being a tool for efficiency, these stakeholders saw AI as a competitive edge in commercial markets and combat.

Consequently, research objectives and funding priorities changed from augmentation to automation. AI could now replace human workers, analysts, and even commanders, making it more than a collaborator. Applications in the actual world demonstrate this change. From customer service chatbots to self-driving delivery drones, businesses in the private sector started investing in AI systems that could foresee consumer behaviour, improve supply chains, and replace human Labour with algorithms. In the meantime, military initiatives invested in AI for threat identification, surveillance, and battle scenario decision-making, including the contentious creation of autonomous weaponry. This tendency was further accelerated by the development of deep learning, which enables machines to "learn" from large datasets without the need for explicit programming[11]. This change had profound implications. AI expanded beyond specific jobs and invaded fields once considered exclusively human, such as judgment, ethics, and interpretation. Artificial Intelligence models, such as OpenAI's GPT-3, can produce human-like writing and conversation. Vision systems are now embedded in facial recognition systems that have real implications for privacy and law enforcement. These developments raise questions about control, agency, and trust by muddying the difference between simulation and decision-making.

While this is almost always true, transparency, accountability, and human oversight are often compromised. This philosophical rupture, between augmentation and substitution, now defines the AI landscape, where AI no longer imitates human behaviour or societal engagement, but intrudes on participants lurking in many aspects of societal decision-making. Bernard Marr notes that modern AI systems are valued for making decisions "faster and more accurately than humans".[12] AI is becoming mundane due to this move toward autonomous decision-making, but there may be unforeseen repercussions. The very institutions created to promote human happiness now risk harming it due to factors like algorithmic bias and Labour displacement.

[11] Deep Learning Explained: How Deep Learning Works in AI, *Shopify* (May 21, 2024), https://www.shopify.com/ng/blog/deep-learning.

[12] Bernard Marr, The Key Definitions of Artificial Intelligence (AI) That Explain Its Importance, *Bernard Marr & Co.*, https://bernardmarr.com/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/.

ETHICAL AND ACCESSIBILITY CHALLENGES IN AI DEPLOYMENT

Ethical principles serve as a framework to guide the responsible development, deployment, and use of AI systems. These principles ensure that AI aligns with societal values, respects human rights while promoting fairness, transparency, and accountability.[13] Artificial Intelligence is becoming increasingly common in critical areas such as healthcare, and policing. Such developments have emerged as primary sources of ethical and human rights concerns. Despite the efficiency and scalability of AI, its presence has also raised accessibility issues and exacerbated systemic biases. Predictive policing and facial recognition tools participating in false arrests is one example, having the highest impact on the particular experience of racial minorities, and the apparent violation of the right to equality and non-discrimination. Many AI hiring tools have reproduced gender-based biases by filtering out women candidates based on skewed data sets used for training the algorithms.[14] Even with the rise of fake AI platforms, the demand for AI-based services has led to counterfeit AI Platforms designed to deceive users and distribute malware, steal sensitive data, or enable financial fraud. Examples include HackerGPT Lite, which at first glance appears to be an AI tool but is suspected to be a phishing website that distributes malware.[15]

The question of accessibility is a vital issue. Many AI tools are developed without consideration for people with disabilities, resulting in digital exclusion. For example, voice assistants may not recognize the speech of individuals with speech impairment, and sometimes, visual interfaces may not be perceived by screen readers for the visually impaired. These accessibility challenges create exclusion that widens the digital divide to marginalize back shaped demographics. There is an urgent need for governance, inclusivity, and collaboration across the knowledge and action sectors. The Black Box Problem in AI is the difficulty in understanding, interpreting, or explaining how complex AI models, intensive learning, and neural networks arrive at their decisions or predictions. These systems often process vast amounts of data and involve layers of internal computation that are not transparent to users, developers, or even the AI's creator. This

[13] Umair Ejaz & Olaoye Godwin, Ethical Considerations in the Deployment and Regulation of Artificial Intelligence, ResearchGate (Feb. 8, 2024), https://www.researchgate.net/publication/378070372.

[14] Jefferey Dastin, Amazon Scrapped 'AI Recruiting Tool' That Showed Bias Against Women, *Reuters* (Oct. 11, 2018), https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/.

[15] Check Point Research, *AI Security Report 2025* (June 8, 2025), https://engage.checkpoint.com/2025-ai-security-report/.

lack of transparency raises questions around accountability, fairness, trust, and ethical responsibility, particularly when AI systems are involved in sensitive applications such as healthcare, finance, or criminal justice.[16] For instance, when an AI system denies an application for a loan or disallows a specific medical intervention, there is often no clarity around the reasons for the decision rendered. They are therefore often referred to as 'Black Box' models because of the appeal of this interpretability.

An ever-growing invasion of Artificial intelligence use in surveillance and data gathering introduces significant ethical concerns, particularly concerning privacy and autonomy. Scholars have cautioned of a potential 'panoptic society' where individuals are constantly being watched. Surveillance is changing the way that we behave. Such an environment takes away our autonomy and our right to privacy. AI surveillance also has profound implications for marginalized and vulnerable communities as they are disproportionally impacted, risking further social inequalities, and giving greater probabilities of discriminatory targeting.[17] While these may undermine an individual's right to privacy, overcriminalization and profiling, it results in exclusion from essential services, thereby furthering embedded systemic bias within society. AI surveillance will normalize surveillance, ultimately undermining fundamental democratic values unless there is a more rigorous, transparent regulation system regarding algorithmic decision-making. The necessity for ethical safeguards and accountability processes remains even more imperative now than before, an ethical issue, if not worse, one of the most emergent of AI use in weaponizing and military activity.

Autonomous Weapons Systems (AWS), often referred to as "killer robots," can select and engage targets without direct human intervention, raising grave ethical and legal questions about accountability, proportionality, and compliance with international humanitarian law. The delegation of life and death decisions to machines risks eroding human dignity and violating the principle of distinction and necessity under the laws of armed conflict.[18] Furthermore, the use of AI in military surveillance, drone strikes, and cyber warfare has increased anxiety around an AI

[16] Sandra Wachter, Brent Mittelstadt & Chris Russell, Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI, 41 *Comput. L. & Security Rev.* (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3547922.

[17] Rishab Debnath, Vaishav Veeraraghavan P & Nikita Hapse, AI and Privacy: Ethical Concerns in Data Collection and Surveillance, 6 *Int'l J. Factual & Multidisciplinary Res.* (Nov.–Dec. 2024), https://www.ijfmr.com/papers/2024/6/32150.pdf.

[18] Noel Sharkey, Saying 'No!' to Lethal Autonomous Targeting, 9 *J. Mil. Ethics* 369, 383 (2010), https://doi.org/10.1080/15027570.2010.537903.

arms race where nations will compete to deploy increasingly autonomous and lethal systems without adequate regulation. Scholars warn that this uncontrolled militarization of AI could destabilize the security of nations and lower the threshold for conflict.[19] Therefore, ethical governance of AI must address commercial and civilian applications and priorities strict international norms to prevent misuse in warfare. The emergence of tools like FraudGPT, for instance, has empowered cybercriminals to forge high-quality phishing campaigns that perfectly replicate banking websites within seconds. AI agents are now used to craft millions of personalized scams tailored to victims' specific digital profiles and psychological vulnerabilities. Deepfakes are also used to defeat biometric verification systems, create fraudulent documents and videos sophisticated enough to bypass Know Your Customer and Anti-money Laundering regulations, and enable social engineering fraud, such as in the case of the fraudulent CFO who instructed an employee to transfer $25 million during a fake virtual call.[20] There are ethical and accessibility concerns in AI which stem from systemic bias, lack of transparency, misuse in surveillance and warfare, and digital exclusion. To address these issues, we need inclusive design practices, regulatory oversight, plus an international consensus on the responsibility of AI within a democratic society where AI serves humans in responsible and equitable ways.

LIABILITY FRAMEWORK CONCERNING AI

The pressing question today in the legal realm, with the accelerated participation of Artificial Intelligence in human activities, is who is responsible when things go wrong. The complexity of assigning liability in AI-related harm, mainly when that harm stems from bias, is no longer theoretical. It is a legal, ethical, and societal imperative.

One of the most contentious issues is the liability involved with personal AI systems. For example, if a big language model, such as OpenAI's ChatGPT, produces defamatory or deceptive text, such as wrongly accusing someone of a crime, who is to blame?[21] Liability in such instances is determined through sophisticated attribution. The culpability may belong to the end user who triggered the prompt, the developer who designed and trained the system, or even the

[19] Anna Nadibaidze, 'Responsible AI' in the Military Domain: Implications for Regulation, *OpinioJuris* (Mar. 31, 2023), https://opiniojuris.org/2023/03/31/responsible-ai-in-the-military-domain-implications-for-regulation/.

[20] Shlomit Wagman, Weaponized AI: A New Era of Threats and How We Can Counter It, *Harvard Kennedy Sch. ASH Ctr.* (Apr. 8, 2025), https://ash.harvard.edu/articles/weaponized-ai-a-new-era-of-threats/.

[21] Skip Tracing and SEO: A Powerful Combination, *ZapGeeks* (Aug. 3, 2022), https://zapgeeks.com/skip-tracing-and-seo-a-powerful-combination.

platform enabling AI access. Each participant plays a different role in creating and channelling any harmful content, but the existing legal framework is unable to establish liability for the respective parties. A 'responsibility gap' is revealed by the legal ambiguity surrounding AI-generated material, especially when harm results from unpredictable system behaviour or autonomous responses.[22] Liability tends to fall on either developers under product liability principles or users under tort law in the absence of a clear attribution of legal personhood to AI systems, neither of which adequately takes into account the autonomous nature of the system.

Because algorithms contain systemic prejudice, the application of AI by corporations, particularly in banking, healthcare, and recruitment, makes liability issues much more complex. The discriminatory feedback loop that may occur when AI is educated on biased historical data was demonstrated by Amazon's now-defunct hiring tool, which routinely penalized resumes that contained the word "women's." These kinds of situations are not unique. They emphasize the necessity of preventative measures like bias audits. Ex ante fairness assessments are to be considered required, not optional, in high-impact AI systems, per research presented at the 2022 International Association for AI and Law (IAAIL).[23] For audits to work effectively, first intervening to reduce harm before deployment, they must be consistent, transparent and legally enforceable.

However, the current regulatory landscape is not meeting that standard. The EU has taken a significant step forward in incorporating reliability through a risk-based regulatory framework with the AI Act. This paradigm assigns corresponding responsibilities to AI systems limited, high, and unacceptable risk tiers. Notably, law enforcement, education, and employment systems and other domains that are especially susceptible to biased results are categorized as "high risk" and held to stringent data quality standards, openness, and human monitoring. The United States' dependence on post-hoc litigation, where AI-related damages are frequently addressed rapidly through civil cases or agency action, like that of the Federal Trade Commission, starkly contrasts with this proactive strategy. This patchwork approach "fails to generate a consistent deterrent

[22] Francesca Lagioia & Giovanni Sartor, AI Systems Under Criminal Law: A Legal Analysis and a Regulatory Perspective, 33 *Philos. Technol.* 433, 465 (2020), https://doi.org/10.1007/s13347-019-00362-x.

[23] Trevor Bench-Capon, Thirty Years of Artificial Intelligence and Law: Editor's Introduction, 30 *Artif. Intell. L.* 475, 479 (2022), https://doi.org/10.1007/s10506-022-09325-8.

effect" and burdens individual claimants, according to the Cambridge Handbook of Responsible Artificial Intelligence.[24]

Moreover, liability frameworks rarely consider biased AI's economic externalities. Beyond harm to individuals, a poorly designed credit scoring algorithm, for example, or hiring practice, can increase systemic inequality and prevent whole groups from accessing housing, employment, or education. These economic and social rights violations are not just legal violations for individual people but signal a need to shift the tenor of blame from individual people to institutional perpetrators of rights violations. Besides legal culpability, economic rights have been compromised due to AI's effects on Labour. Labour markets are rapidly automating across jobs requiring judgment and discretion, extending beyond operational work to change the nature of many human occupations.

Bridging the responsibility gap is critical for AI to realize its promise of benefiting humanity rather than aggravating our darkest fears. When legal safeguards fall behind technology improvements, the same systems that diagnose illnesses or screen job candidates can propagate bias and cause actual harm. Only by identifying who is liable (whether that is the user, developer, or platform), demanding rigorous bias audits, and implementing comprehensive, risk-based laws will we be able to assure that AI augments human creativity and judgment without jeopardizing economic and social rights. Beyond legal accountability, AI's impact on Labour markets jeopardizes economic rights.

AI VS HUMAN JOBS: COMPETITION OR COLLABORATION

Experts predict that the impact of AI and automation on the job market will be significant. According to the WEF, 85 million jobs will be displaced. It also expects the AI revolution to create 97 million new jobs. This raises a significant question concerning human rights. Is AI a boon or a bane? According to The Future of Jobs Report 2025, 92 million jobs are expected globally between 2025 and 2030 due to structural Labour-market transformation, which includes the impact of technologies like AI and automation.[25] However, this transformation is not entirely

[24] *Christiane Wendehorst, Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks, in The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives 187, 209 (2022),* https://www.cambridge.org/core/services/aop-cambridge-core/content/view/12A89C1852919C7DBE9CE982B4DE54B7/9781009207867c12_187-209.pdf/liability-for-artificial-intelligence.pdf.

[25] *The Future of Jobs Report 2025* (Jan. 7, 2025), *World Econ. Forum,* https://www.weforum.org/publications/the-future-of-jobs-report-2025/digest/.

negative. The same report anticipates creating 170 million new jobs, leading to a net growth of 78 million globally by 2030. Many of these emerging roles are in technology-intensive and green transition sectors such as AI and machine learning specialists, prominent data analysts, renewable energy engineers, and cybersecurity experts–reflecting a shift in demand from routine-based jobs to skill-intensive ones[26]. This evolving dynamic between AI and human Labour invites a deeper reflection on the nature of future employment rather than replacing them outright. For instance, AI can handle repetitive tasks, enabling humans to focus on creative, strategic, and emotionally intelligent tasks, aspects of work–areas where human judgement and empathy are irreplaceable.[27]

DEBUNKING THE "AI THREAT" NARRATIVE

The greatest threat to human society is not artificial intelligence but how humans create, implement, and oversee these systems. The notion that artificial intelligence is a sentient being poised to upend humanity ignores a crucial reality: AI lacks autonomy, will, and consciousness. It makes decisions in the context of other factors. It copies the information and instructions provided by its authors.

According to Melanie Mitchell, a professor at the Santa Fe Institute,[28] even the most sophisticated AI systems do not see the world as humans do. They provide results based on statistical patterns in data, not purpose or comprehension. This framing is critical in determining where it went wrong. When AI creates biased or harmful results, it is most often because of the underlying data or human decisions built into the model's architecture. A well-known example is Amazon's recruitment tool, which penalized résumés with the word "women's," reflecting women's historically low representation in technical professions[29]. The algorithm was not designed to be sexist. It duplicated the trends it discovered in prior hiring data, in which male

[26] Prashant V. Singh, 170 Mn New Roles… Future of Jobs Report 2025 by World Economic Forum Reveals Job Disruption Will Equate to 22% of Jobs by 2030, *ET Now* (Jan. 14, 2025), https://www.etnownews.com/economy/future-of-jobs-report-2025-by-world-economic-forum-reveals-job-disruption-will-equate-to-22-of-jobs-by-2030-article-117238134.

[27] Artificial Intelligence vs. Human IQ: Who Wins? *Free IQ Test* (Jan. 10, 2024), https://www.free-iqtest.net/24/artificial-intelligence-vs-human-iq-who-wins/.

[28] Richard Waters, Melanie Mitchell: Seemingly 'Sentient' AI Needs a Human in the Loop, *Fin. Times* (Oct. 21, 2024), https://www.ft.com/content/304b6aa6-7ed7-4f18-8c55-f52ce1510565.

[29] Dastin, supra note 14.

candidates dominated. Similarly, predictive policing techniques and judicial AI systems such as Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) have been criticized for practicing racial bias. It was discovered that COMPAS disproportionately categorized Black defendants as high-risk, even when they did not reoffend,[30] compared to white defendants with identical backgrounds.[31] Such outcomes reveal how historical biases in criminal justice datasets can persist and even worsen through AI systems.

However, if appropriately applied, AI's potential also offers a chance to address and lessen these prejudices. According to research from Tulane University, machine learning techniques can reduce bias in judicial sentencing if they are appropriately developed and closely watched. By rating offenders and suggesting sentencing ranges, artificial intelligence was implemented in Virginia to assist judges in determining the likelihood of recidivism. The study discovered that when paired with human judgment, these tools could combat prejudices rather than strengthen them, although critics are still cautious.[32]

Thus, AI has the potential to help to reveal covert systematic discrimination. IBM, for instance, has created bias-detection algorithms that look for skewed results in datasets and machine learning models. By highlighting disproportionate effects on specific groups, these technologies allow developers to address problems before deployment.[33] This illustrates AI's dual nature: when used responsibly and ethically, it may be both a potential danger and a corrective force.

Ultimately, Personal Artificial Intelligence (PAI) systems are not autonomous beings. They are sophisticated mimics that replicate the data and commands of their human creators. The lack of openness, responsibility, and moral supervision around the robots' use poses the most risks, not the technologies themselves. AI is not the actual threat, but the unbridled human effect on artificial intelligence. By acknowledging this, we shift from science fiction concerns to practical changes that make the real players, governments, businesses, and capitalists, responsible for the creation and application of AI.

---

[30] Nikhil Raghuveera & Hannah Biggs, Pretrial Risk Assessment Tools Must Be Directed Toward an Abolitionist Vision, *Atlantic Council* (Dec. 18, 2020), https://www.atlanticcouncil.org/blogs/geotech-cues/pretrial-risk-assessment-tools-must-be-directed-toward-an-abolitionist-vision/.

[31] Julia Angwin et al., Machine Bias, *ProPublica* (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[32] Allyson Brunette, Humanizing Justice: The Transformational Impact of AI in Courts, from Filing to Sentencing, *Reuters* (Oct. 25, 2024), https://www.thomsonreuters.com/en-us/posts/ai-in-courts/humanizing-justice/.

[33] Jennifer Aue, The Origins of Bias and How AI May Be the Answer to Ending Its Reign, *Medium* (Jan. 13, 2019), https://medium.com/design-ibm/the-origins-of-bias-and-how-ai-might-be-our-answer-to-ending-it-acc3610d6354.

ETHICAL DEPLOYMENT: BIAS MITIGATION AND ACCOUNTABILITY

As AI systems are used increasingly daily, it becomes essential to address the ethical implications around their development and adoption. These implications ensure they are used responsibly, equitably, and respectfully, in line with human rights and societal values. The ethical deployment of Artificial Intelligence concerns includes data responsibility and privacy, fairness, explainability, robustness, transparency, environmental sustainability, inclusion, moral agency, value alignment, accountability, trust, and technology misuse[34].

When discussing ethical deployment, it also raises concerns about AI bias. Bias in AI, also known as machine learning or algorithm bias[35], occurs when the results produced by an AI system are skewed due to human biases present in the training data or the AI algorithm itself[36]. Bias can be systemic (encoded through historical data), emergent (arising from feedback loops), or implicit (through developer or user assumptions).

CASE STUDIES

The *COMPAS* Algorithm (Correctional Offender Management Profiling for Alternative Sanctions), used in the U.S. Courts to predict recidivism risks, was found to disproportionately classify Black defendants as high-risk compared to white defendants with similar profiles. This case underscores how reliance on historical data without critical evaluation can perpetuate systemic bias, particularly when used in high-stakes legal decisions.[37] Even *Amazon's AI Recruitment Tool*, Amazon developed an AI hiring system to automate resume screening. However, the tool began penalizing resumes that included terms like "women's" (e.g., "women's chess club captain") and downgraded graduates of all-women's colleges. The system has learned from ten years of biased hiring data dominated by male applicants. Despite corrective efforts, hidden biases persisted, leading to the eventual discontinuation of the tool. This incident highlights the risks of unexamined training data and the necessity of diversity in development teams and datasets.[38]

BIAS MITIGATING TECHNIQUES

[34] What Is AI Ethics? *IBM* (Sept. 17, 2024), https://www.ibm.com/think/topics/ai-ethics.
[35] Dr. Timothy J. Purn
[36] Brunette, supra note 32.
[37] Jeff Larson, Surya Mattu, Lauren Kirchner & Julia Angwin, How We Analyzed the COMPAS Recidivism Algorithm, *ProPublica* (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.
[38] Artificial Intelligence vs. Human IQ, supra note 27.

To address such issues, bias mitigation techniques are typically categorized into these approaches:

- Diverse and Representative Data: Ensuring datasets reflect a wide range of demographics (age, gender, race, etc.) to minimize skewed outcomes.
- Data Preprocessing: Cleaning, anonymizing, and rebalancing data to reduce inherent bias before model training.
- Bias-Aware Algorithms: Choosing or designing algorithms with built-in fairness constraints or using ensemble methods to offset individual model biases.
- Human Oversight: Implementing human-in-the-loop systems to review and correct AI outputs where needed.
- Transparency: Clearly explaining AI decision-making processes enables scrutiny to build trust.
- Ongoing Monitoring: Regularly auditing AI performance with fairness benchmarks to detect and correct biases over time
- Ethical Framework and Diverse Teams: Adopting global AI ethics standards (e.g., EU guideline, FAT/ML) and promoting team diversity to uncover blind spots and ensure inclusive design.
- Training and Awareness: Helping developers and stakeholders understand bias detection, mitigation, and responsible AI.

Bias mitigation is not a one-time process; it requires continuous attention, a commitment to ethical action, and interdisciplinary collaboration to keep AI technologies fair, inclusive, and accountable.

REGULATORY AND ETHICAL FRAMEWORKS

The EU's Artificial Intelligence Act (AI ACT), proposed in April 2021, is the first major regulatory framework to govern AI systems based on their unacceptable, high, limited, and minimal risk levels. Unacceptable uses, such as government social scoring, are banned. High-risk systems, like those used in migration or employment, must undergo conformity assessment, documentation, and ongoing monitoring. The Act has global implications for any company

offering AI services in the EU, regardless of location. It emphasizes ethical principles such as transparency, accountability, human oversight, and fairness.[39]

India's primary document on AI, the national strategy for Artificial Intelligence, was released by NITI Aayog in 2018. It outlines India's vision for AI, emphasizing the importance of leveraging AI for inclusive growth and focusing on five key sectors: healthcare, agriculture, education, smart cities, and smart mobility. National AI Portal 2020 is a platform for AI resources and collaboration. Responsible AI for Youth 2020, a MeitY initiative to train students in AI. Concerning AI Standards, BIS is developing standards on data privacy, quality, and governance.[40]

Additionally, Principles for Responsible AI (NITI Aayog, 2021) lays down seven core ethical principles, namely- safety, inclusivity, equality, privacy, transparency, accountability, and human values, divided across system and societal considerations. Draft Digital India Act (2023), aims to regulate AI in digital services, propose no-go zones for harmful AI use, and introduce strict penalties to ensure user safety. Draft National Data Governance Framework Policy (2022), seeks to modernize data governance and promote AI innovation through open, anonymized datasets and public-private collaboration. TRAT Recommendation on AI (2023) proposes a unified AI regulatory framework with a risk-based classification system and a dedicated statutory body for oversight. India AI National Program (2023) is a holistic initiative to build an AI ecosystem through multi-stakeholder collaboration across government, academia, and industry. National Cybersecurity Reference Framework 2023 provides structured cybersecurity guidelines for critical sectors, addressing AI-related risks and governance architecture. Global Partnership on AI(GPAI), India membership reflects its commitment to globally aligned, human-centric, and trustworthy AI development.[41]

As per OECD AI Principles, 2023, the government reported over 1000 policy initiatives across more than 70 jurisdictions in the OECD.[42] By these principles, policymakers can guide the development and deployment of AI to maximize its benefits and minimize its risks. OECD

[39] *EU AI Act: First Regulation on Artificial Intelligence*, *European Parliament* (July 8, 2023), https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

[40] Gayathri Haridas, Sonia Kim Sohee & Atharva Brahmecha, The Key Policy Frameworks Governing AI in India, *Access Partnership*, https://accesspartnership.com/opinion/the-key-policy-frameworks-governing-ai-in-india/.

[41] Id.

[42] *About the OECD AI Principles*, *OECD*, https://www.oecd.org/en/topics/ai-principles.html.

Principles promote the use of AI that is innovative and trustworthy and that respects human rights and democratic values.[43] Adopted in May 2019, they set standards for AI that are practical and flexible enough to stand the test of time. Policy recommendation for AI: invest in AI R&D, support open, fair data, an inclusive AI Ecosystem, interoperable Governance, Human Capacity & Jobs, and Global Collaboration.[44]

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems has launched Ethically Aligned Design, a vision for prioritizing human well-being with Autonomous and Intelligent Systems, First Edition (EAD1e).[45] According to Doug Frantz of the Organization for Economic Cooperation and Development (OECD): "What has been missing, until now, is a clear, practical road map to guide the development of AI's benefits and address its potential risks. Ethically Aligned Design fills that vital gap. It provides businesses, policymakers and everyday people with the essential tools to understand the stakes and support global standards to maximize the benefits and mitigate the risks.[46]

CONCLUSION

Artificial intelligence does not function in isolation from human motives. Instead, it reflects the creators' interests, ideals, and constraints. The conversation must shift away from sensationalist anxieties about autonomous machines and toward a more nuanced understanding of AI as a socio-technical system that both magnifies and challenges human agency. Crucially, the primary issue is not the machine's autonomy, but rather the distribution of responsibility among those who build, deploy, and regulate it.

AI mirrors the goals of those who created it. However, it develops by discovering patterns we might not even notice, much less comprehend, in contrast to static tools. This presents both a risk and an opportunity. We regain control over AI's course if we view it as a reflection of human will rather than an outside force. To build with inclusivity in mind, this agency must be exercised by raising difficult ethical questions, and turning down convenience when it compromises

[43] Id.
[44] Id.
[45] Christ Brantley, IEEE Global Initiative Releases Treaties on Ethically Aligned Design of AI Systems, *IEEE Global Policy* (Apr. 1, 2019), https://globalpolicy.ieee.org/ieee-global-initiative-releases-treatise-on-ethically-aligned-design-of-autonomous-and-intelligent-systems.
[46] Id.

accountability. The stakes transcend beyond utility, into the fabric of justice, governance, and human dignity.

The ethical deployment of AI is not merely a technical challenge but a societal imperative. As AI systems become deeply embedded in governance, healthcare, employment, and justice, ensuring fairness, accountability, and transparency becomes non-negotiable. Bias mitigation, robust oversight, and inclusive design must work with enforceable regulatory frameworks to safeguard human rights and democratic values. Without a shared ethical commitment across institutions, industries, and borders, technical solutions alone are insufficient. In this light, the future of AI must be steered by human-centric principles and proactive global collaboration, ensuring technology serves humanity, not the other way around. Ultimately, the legitimacy of AI systems will be determined by our collective commitment to regulate them by our ethical and democratic principles, rather than by their technological sophistication.