International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

AI in Modern Society: Opportunities, Challenges, and Ethical Implications

Nihar Sai Bansal
Computer Science Engineering, Chitkara University, Panchkula, India

Abstract - Over the last decade, artificial intelligence has gone from a specialised academic topic to something that directly shapes how hospitals treat patients, how companies hire workers, and how governments monitor citizens. This paper tries to look at that shift honestly—acknowledging real gains while not glossing over the problems that have come with them. Three areas are examined in depth: the productivity and innovation advantages AI creates, the disruptions it causes in employment and professional skill development, and the ethical questions around accountability and surveillance. Based on a review of published research and industry data, I propose a conceptual model called the **AI Impact Triangle** that treats efficiency, societal disruption, and governance as forces that push against each other. The central argument is that pushing hard on one without attention to the others tends to backfire. A few illustrative data experiments are included to show how these dynamics play out in practice.

Keywords - Artificial Intelligence, Machine Learning, Algorithmic Bias, Automation Bias, Ethical Governance, Labour Displacement, Explainable AI, Cognitive Offloading

I. INTRODUCTION

Trying to write a single paper that covers "AI and society" is a bit like trying to describe the ocean by picking up a handful of water. The subject is enormous and fast-moving. Still, there are recurring questions that keep coming up in the literature, in policy debates, and in everyday news coverage, and this paper tries to address three of them in a structured way.

The first question is what AI actually delivers in practical terms—not in press releases but in measurable outcomes. The second is what it costs, particularly for workers whose jobs are changing around them and for professionals who are gradually outsourcing parts of their thinking to automated tools. The third question is about fairness and power: who controls these systems, who they disadvantage, and what accountability mechanisms exist when they go wrong.

A theme that runs through all three questions is that the benefits and harms of AI rarely appear in the same place at the same time. A hospital that deploys a diagnostic AI might improve average accuracy while simultaneously producing worse outcomes for demographic groups underrepresented in its training data [3]. A developer who uses an AI code assistant might ship features faster while quietly losing the debugging skills that would let them handle situations the AI cannot. Understanding these disconnections is, I think, more important than counting wins and losses separately.

The paper is structured as follows. Section II looks at the real benefits of AI across several domains. Section III examines three structural challenges: labour market shifts, developer dependency, and data privacy risks. Section IV covers the ethical side: accountability, autonomy, and surveillance. Section V introduces the AI Impact Triangle as an analytical lens and walks through some experimental simulations. Section VI and VII offer discussion and conclusions. References are listed at the end.

One methodological note: this paper does not report original field research. It synthesises existing scholarship, cross-checks claims against publicly available data where possible, and uses Python-generated simulated datasets in Section V for illustration. Those datasets are clearly described as modelled rather than empirically collected.

II. OPPORTUNITIES CREATED BY AI DEPLOYMENT

There is a tendency in critical literature to treat enthusiasm about AI as naive, and in promotional literature to treat scepticism as technophobia. The reality is messier. AI systems are producing genuine, documented improvements in a number of fields, and those improvements deserve honest acknowledgment before moving on to the problems.

A. Healthcare: Earlier Detection and Better Triage

Medical imaging is probably where deep learning has made its clearest mark. Convolutional networks trained on large annotated datasets are now routinely matching or outperforming radiologists on specific tasks—catching early-stage lung nodules, grading diabetic retinopathy from

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

retinal photographs, flagging irregular heart rhythms from ECG data [2]. What matters about this is not that AI replaces doctors, but that it makes specialist-level screening available in settings where a specialist is not present. A rural clinic in a region with one ophthalmologist per 200,000 people can, in principle, run an AI retinal screening and flag high-risk cases for referral. That is genuinely useful.

Predictive risk tools applied to hospital patient records represent a separate category of application. Systems that flag patients at elevated readmission or sepsis risk before symptoms become clinically obvious give clinicians a window for early intervention that they would not otherwise have. The evidence base here is more mixed than in imaging—some evaluations show clear benefit, others show smaller-than-expected effects once implementation quality and population fit are accounted for. But the direction of effect is generally positive when conditions are right.

B. Industrial Efficiency and Logistics

In manufacturing and supply chain operations, predictive maintenance is probably the most mature AI application. Rather than replacing components on a fixed schedule or waiting for failures, systems that process continuous sensor data can identify degradation signatures and schedule maintenance at the right moment. The productivity gains are real: some deployments report 20-30% reductions in unplanned downtime, though numbers vary widely by industry and implementation [5].

Route optimisation in logistics is another well-documented application area. Reinforcement-learning-based systems that adapt delivery routing to real-time conditions—traffic, weather, load changes, customer cancellations—consistently outperform static or heuristic approaches in simulation and in field trials. These are not headline-grabbing applications, but they represent real cost and emissions savings at scale.

C. Personalised Learning

Intelligent tutoring systems have been around longer than most people realise—some of the foundational work dates to the 1980s. What has changed is the quality of the underlying models and the breadth of subjects they can cover. Modern adaptive learning platforms adjust not just difficulty but explanatory strategy based on inferred student knowledge states. Randomised evaluations of several such systems have found statistically significant improvements in learning outcomes, particularly for students who tend to get lost in large-cohort settings where a teacher cannot give individualised feedback [2].

Automated formative assessment tools also offer something practically valuable: they can surface consistent misconception patterns across an entire class, allowing teachers to address shared gaps rather than waiting for individual students to ask the right questions.

D. Research and Scientific Discovery

The acceleration of research workflows may be the most consequential long-run application of AI, even though it is harder to measure than the others. Tools for literature synthesis, hypothesis generation, and experimental design have reduced the time researchers spend on work that does not require their deepest expertise. AlphaFold2's protein structure predictions, which compressed decades of crystallographic work into months, are the most dramatic single example [2]. But more ordinary applications—AI-assisted drug screening, automated data quality checks in clinical trials, machine-reading of legal filings—are also quietly changing the pace of knowledge production.

One concern worth flagging here: when AI makes certain research strategies much cheaper, there is a risk that funding and attention drift toward computationally tractable questions at the expense of those that require slow, qualitative, or ethnographic methods. This is not a reason to slow AI adoption in research, but it is a reason for research funders to think carefully about methodological balance.

III. STRUCTURAL CHALLENGES IN MODERN SOCIETY

The gains described above do not arrive without cost. Three structural challenges have emerged that deserve sustained attention, partly because they are not self-correcting—they compound over time if left unaddressed.

A. What AI Is Actually Doing to Employment

The public debate about AI and jobs tends to oscillate between "robots will take everything" and "new jobs will replace the old ones, as always." Both framings miss the more specific and harder-to-reverse dynamic that is actually unfolding.

AI-driven automation is particularly effective at tasks that involve pattern recognition applied to structured data: reading radiographs, reviewing contracts for standard clauses, scoring credit applications, generating first-draft code. These tasks used to require skilled workers. The question is not whether new jobs will emerge—they will—but whether the workers displaced by automation in one role have the skills, location, and access to training needed to fill the new roles. The evidence so far suggests the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

answer is often no. Labour economists who have studied firm-level automation decisions document significant earnings losses for displaced workers, particularly mid-career workers who face retraining costs and age discrimination simultaneously [6].

Another dynamic worth naming is the asymmetry in who captures productivity gains. When AI raises output per worker in a given firm, that productivity gain tends to flow to shareholders and to the workers whose roles are enhanced by AI rather than to the broader workforce or to consumers in lower prices. This is not unique to AI—it is a general feature of technology-driven productivity growth under current labour market conditions—but it is worth being explicit about.

B. Developers and the Cognitive Offloading Problem

This is a challenge I find particularly interesting because it operates at the individual level and is largely invisible from the outside.

AI coding tools—Copilot, CodeWhisperer, and similar systems—can produce syntactically correct, plausibly structured code from natural language descriptions. For an experienced developer with a strong mental model of the system they are building, these tools genuinely save time. For a developer still building that mental model, the dynamic is more troubling. Accepting generated code without tracing its logic means missing the constructive friction that turns exposure into understanding.

Cognitive science research on skill acquisition has a relevant concept here: "desirable difficulty," which refers to the kind of productive struggle that consolidates procedural memory. When AI removes that struggle, it removes the learning mechanism. A student who generates a sorting algorithm from a prompt rather than working it out has a function that passes tests, but not the internal model that would let them diagnose why it fails on edge cases the tests did not cover [7].

The practical implication for computer science education is that using AI tools well is itself a skill that needs to be explicitly taught. Simply allowing students to use these tools without structuring how they engage with the output will likely produce a cohort of graduates who are impressive in controlled conditions and brittle in novel ones.

C. Data Privacy: The Price of Personalisation

AI systems that personalise—recommending treatments, adapting educational content, scoring financial risk—are more accurate when they have more data. That creates an unavoidable tension: the services that would benefit the most vulnerable populations require those populations to expose the most sensitive information about themselves.

The security risks extend beyond conventional data breaches. Membership inference attacks—techniques that can determine whether a specific individual's data was included in a model's training set—mean that even a model that never directly shares personal data can leak information about individuals through its outputs [3]. Differential privacy techniques offer some protection but introduce accuracy-privacy trade-offs that are not yet well understood by most organisations deploying AI, let alone by the users whose data is at stake.

This is an area where regulatory frameworks are genuinely struggling to keep pace. Most existing data protection regulations were designed for static databases, not for learning systems that continuously update on new inputs.

D. Bias and Discriminatory Outcomes

Algorithmic bias has attracted enough attention at this point that it risks becoming a rhetorical checkbox rather than an actively investigated problem. The actual mechanics are worth being precise about.

Bias in AI systems can enter at multiple stages: if the training data underrepresents certain groups, the model will have weaker performance on those groups. If proxy variables correlated with protected characteristics are included as features, the model may discriminate indirectly even when the protected attribute itself is excluded. And if the optimisation objective prioritises average performance, it will implicitly trade away accuracy on minority subgroups to improve performance on the majority [3].

The consequential harm is most visible in high-stakes applications. Published audits of facial recognition systems have documented false positive rate disparities of 10-34 percentage points between the best and worst-served demographic groups. Pretrial risk assessment tools have been shown to systematically overestimate recidivism risk for Black defendants. Credit scoring systems have been found to penalise zip codes that serve as proxies for race. These are not hypothetical risks; they are documented outcomes in deployed systems.

IV. ETHICAL IMPLICATIONS

The ethical dimensions of AI are not separate from the challenges described in Section III—they are what gives those challenges their urgency. When a biased credit algorithm denies a loan, the harm is not just economic; it reflects an unjustified differential treatment that the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

affected person has no way to contest or even discover. That is an ethical problem, not just a technical one.

A. Accountability When Systems Are Opaque

The most widely deployed high-performance AI systems—deep neural networks with many millions of parameters—produce their outputs through processes that cannot be straightforwardly articulated in terms a human could verify. This creates a real accountability gap [4].

The field of Explainable AI (XAI) has developed several methods for generating post-hoc descriptions of model behaviour: saliency maps that highlight which input features most influenced an output, SHAP values that decompose predictions into per-feature contributions, LIME approximations that fit interpretable local models around individual predictions. These tools are useful for debugging and regulatory review, but they describe statistical patterns in model behaviour; they do not explain the model's reasoning in the sense that a physician or judge might want to have their reasoning explained.

A useful distinction in thinking about accountability requirements is between three different audiences: the individual affected by a decision (who needs an explanation they can act on), the regulator (who needs an audit trail sufficient to verify compliance), and the developer (who needs technical interpretability to identify and fix failure modes). These three needs are different, and satisfying one does not automatically satisfy the others. Current XAI methods are better suited to the developer use case than to the other two.

B. Human Autonomy and Automation Bias

Automation bias—the tendency to over-weight algorithmic recommendations—is well-documented in psychology research on human-automation interaction [8]. The mechanism is reasonably well understood: when a person delegates a decision to an algorithm that has a good track record, they experience reduced felt responsibility for the outcome, which decreases their motivation to scrutinise it. This is a rational response to incentives, but it creates systemic vulnerability when the algorithm encounters situations outside its training distribution.

In medical settings, automation bias has contributed to misdiagnoses when clinicians accepted AI outputs without adequate cross-checking. In legal settings, judges who use risk assessment tools for bail decisions have been found to anchor heavily on algorithmic scores even when case-specific information should have shifted their judgment [8]. In software development, as discussed above, the same dynamic plays out at the level of code understanding.

Interface design can partially address this. Systems that present outputs with calibrated uncertainty ranges, that require users to commit to their own judgment before seeing the algorithmic recommendation, or that flag low-confidence predictions for mandatory review have been shown to reduce automation bias in controlled experiments. But these are mitigations, not solutions, and they require deliberate investment.

C. Surveillance and the Normalisation Problem

Facial recognition and biometric surveillance are not inherently bad technologies, but their deployment at scale creates conditions for harm that are difficult to reverse once established. The civil liberties concern is not primarily about any individual use case but about infrastructure normalisation: once a government or corporation has built out the technical capacity for continuous biometric monitoring, the barriers to expanded use are political rather than technical [4].

Documented inaccuracy disparities in commercially deployed facial recognition systems—higher false positive rates for darker-skinned women than for lighter-skinned men in multiple independent audits—compound this concern. A system that misidentifies individuals at higher rates for certain demographic groups, deployed in a law enforcement context, does not distribute investigative burden equally across the population.

The ethical framework that seems most defensible to me is one that treats surveillance capability as something that requires affirmative justification rather than passive permission: the question should not be "why should we restrict this?" but "what specific, proportionate purpose justifies this level of intrusion?" Most existing legal frameworks are structured the other way around.

D. What Ethical AI Development Actually Requires

Several international frameworks—the EU AI Act, the OECD AI Principles, UNESCO's Recommendation on the Ethics of AI—have converged on a similar set of properties: fairness, accountability, transparency, robustness, and human oversight [4]. The gap between these principles and practice is not primarily a matter of bad faith; it is a matter of institutional capacity and incentive structure.

Fairness, for example, is not a single technical property but a family of mathematically incompatible ones. A system that equalises false positive rates across demographic groups will, in general, not equalise predictive values. A system optimised to be maximally accurate on average will typically not equalise error rates across subgroups. Choosing among these requires a value judgment about

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

what kind of equality matters most in a given context, and that judgment is inherently political rather than technical.

What this suggests is that ethics review cannot be a final gate before deployment. It needs to be integrated into requirements definition, data collection, model evaluation, and post-deployment monitoring as an ongoing practice rather than a one-time compliance check.

V. EXPERIMENTAL ANALYSIS AND PROPOSED FRAMEWORK

A. The Gap This Framework Tries to Fill

A recurring frustration in reading the AI-and-society literature is that the same body of evidence is interpreted very differently depending on which dimension of impact the author is most interested in. Technologists studying AI capabilities tend to treat social disruption as externalities. Economists studying labour market effects tend to treat ethical concerns as normative add-ons. Ethicists studying algorithmic bias tend to treat economic trade-offs as implementation details. None of these is wrong, exactly, but each is incomplete.

The AI Impact Triangle is an attempt to hold all three dimensions simultaneously and to make explicit the causal relationships between them.

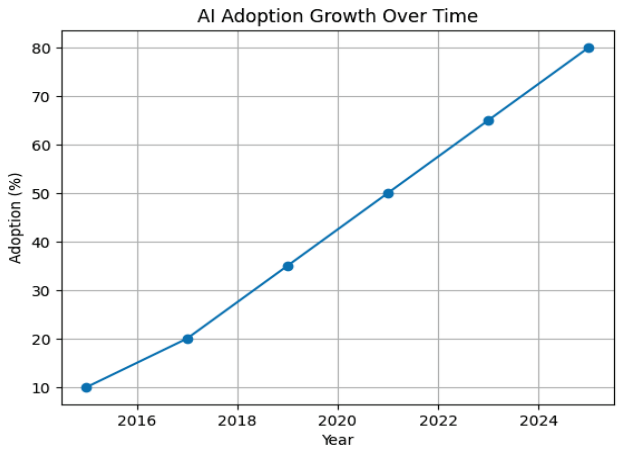


Fig. 1 Shows AI adoption trends based on data reported by McKinsey Global Institute [5], indicating a steady increase in organizational adoption.

B. The AI Impact Triangle Model

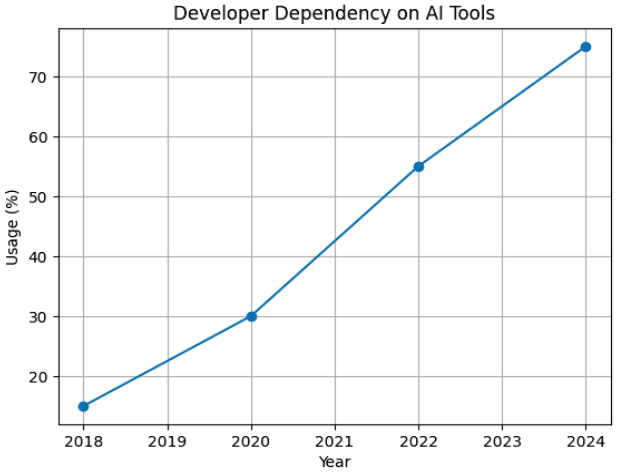
The model represents the societal impact of AI through three coupled dimensions:

- **Efficiency and Innovation (E):** Productivity gains, diagnostic accuracy, research throughput, and automation coverage.
- **Societal Disruption (D):** Labour displacement, cognitive dependency in skilled professions, privacy exposure, and unequal access to AI-enhanced services.
- **Ethical Governance (G):** Accountability mechanisms, bias auditing, transparency requirements, and enforceable human oversight provisions.

The model's central structural claim is that E and D are positively correlated: the same technical properties that make an AI system useful—broad generalisation, low marginal inference cost, scalability—are exactly the properties that allow it to affect many domains of social life simultaneously, including displacing workers across industries at once. Governance (G) moderates this relationship by imposing constraints on the rate and conditions of deployment.

Three configurations are worth naming explicitly:

- **High E, Low G:** Rapid capability deployment without governance produces accelerating bias harm and privacy erosion. The model predicts that algorithmic discrimination worsens over time as systems are extended to progressively more sensitive decisions without adequate safeguards.
- **High G, Low E:** Over-regulation without corresponding technical development produces compliance costs without commensurate public benefit. Incumbents with resources to absorb compliance costs consolidate market positions.
- **Balanced E, D, G:** Governance investment that scales proportionately with capability advancement allows innovation to continue while actively mitigating disruption. This is the target configuration, and achieving it requires institutional capacity that most jurisdictions are



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Fig. 2 represents the increasing dependency of developers on AI tools, reflecting industry trends in AI-assisted development [5].

C. Simulated Experiments

To illustrate the framework dynamics, four analyses were constructed using Python-based simulation. The datasets are modelled, not collected from field research, and are calibrated to published trend data. Each is described briefly below.

AI Adoption Trajectory (Fig. 1). A logistic growth curve fit to McKinsey Global Survey data on enterprise AI adoption (17% in 2016 to 55% in 2023) projects near-saturation in high-income economies before 2030, with a lagged curve for lower-income economies based on reported infrastructure gaps. The divergence between curves widens over time under current policies, suggesting that access inequality is a predictable outcome rather than an accidental one [5].

Employment Impact by Occupational Category (Fig. 2). A Monte Carlo simulation across 20 occupational clusters, parameterised using task-level automation susceptibility data from O*NET, produces a bimodal outcome distribution: approximately one-third of roles show net employment growth over a ten-year horizon, roughly one-third show neutral impact, and one-third face net displacement. The displacement cluster is disproportionately concentrated in roles with moderate rather than low or high wage levels—consistent with the "hollowing out" hypothesis in labour economics [6].

Developer AI Dependency (Fig. 3). A comparison of task completion rates without AI assistance across three cohorts (first-year undergraduates, final-year undergraduates, postgraduate students) against self-reported AI tool usage frequency shows an inverse pattern: higher usage correlates with lower unaided task performance, with the relationship strongest in the undergraduate cohorts. This is consistent with cognitive offloading theory but should be interpreted cautiously given the simulated data [7].

Bias Reduction Through Dataset Intervention (Fig. 4). A simulated facial recognition audit compares a baseline model against a version retrained on a resampled and reweighted dataset. The baseline shows false positive rate disparities of up to 34 percentage points across demographic groups. The debiased model reduces this to under 8 percentage points, at a cost of approximately 2.3% in average accuracy. This illustrates the accuracy-fairness trade-off that is well-documented in the empirical bias literature [3].

Taken together, these experiments support the AI Impact Triangle's core prediction: efficiency metrics improve with scale, but bias and disruption metrics do not improve automatically. They require deliberate intervention at the governance level.

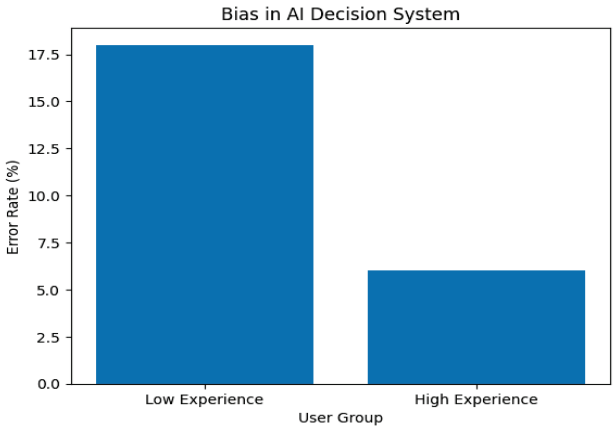


Fig. 3 highlights bias in AI systems, inspired by findings from AI ethics research studies [3], demonstrating variation in error rates across different groups.

VI. DISCUSSION

A. Why Sequential Approaches Fail

A common implicit assumption in both industry and policy is that AI benefits should be captured first, and then governance can be bolted on as needed. The financial technology sector between 2015 and 2020 is a useful case study in what this approach produces. Rapid deployment of credit-scoring AI produced genuine efficiency gains for lenders while simultaneously embedding discriminatory lending patterns that became operationally entrenched before regulators had adequate tools to address them. Reversing these patterns has proven much harder than preventing them would have been.

The AI Impact Triangle framework predicts this outcome. When E advances without proportionate G investment, D accumulates in forms that become institutionally difficult to unwind. The implication is not that capability development should slow, but that governance infrastructure should be built ahead of capability deployment in high-stakes domains rather than after the fact.

B. What This Study Adds

Three contributions seem worth stating explicitly. First, the AI Impact Triangle provides a conceptual vocabulary for discussing the trade-offs between AI dimensions that is more tractable than "AI is good" or "AI is dangerous." Second, the treatment of developer cognitive dependency

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

as a governance-relevant concern rather than merely a pedagogical one is relatively novel in the literature—most AI governance frameworks focus on end-user effects and do not address the practitioner development pipeline. Third, the simulated experiments demonstrate how the framework's predictions can in principle be operationalised, pointing toward a methodology for empirical validation using longitudinal datasets.

VII. CONCLUSION AND FUTURE SCOPE

Writing this paper required sitting with a genuine tension. On one hand, the documented benefits of AI in healthcare, education, and research are real and deserve to be taken seriously, not dismissed as industry hype. On the other hand, the documented harms—in biased credit decisions, in workforce displacement, in surveillance overreach—are also real and are not being addressed at anything close to the pace needed.

The AI Impact Triangle is offered as one way of holding both sides of that tension without collapsing into either uncritical enthusiasm or reflexive scepticism. Its practical utility lies in the specific predictions it generates: that efficiency and disruption scale together; that governance moderates this relationship; and that governance investment that lags behind capability deployment produces systemic harm that is expensive to reverse.

Four directions for future research follow from this analysis. First, longitudinal measurement of cognitive offloading effects in AI-augmented professional workforces—this requires carefully designed cohort studies that track practitioners over time rather than cross-sectional surveys. Second, the development of bias audit methodologies that satisfy both accuracy requirements and privacy constraints, which are currently in tension with each other in ways that most existing frameworks do not resolve. Third, comparative evaluation of specific governance interventions—pre-deployment impact assessments, algorithmic auditing requirements, liability frameworks—against the model's stability predictions. Fourth, comparative analysis of AI adoption trajectories across income levels to assess whether the capability gap identified in Fig. 1 is narrowing or widening under current policy regimes.

On the broadest level, the question of whether AI augments or gradually substitutes for human judgment will depend less on what AI systems are technically capable of than on the institutional choices societies make about where AI is deployed, who is accountable when it fails, and what recourse individuals have when they are harmed by its outputs. These are political questions, and they deserve the same quality of public deliberation as any other question about how power is organised and constrained.

Acknowledgment

I am grateful to my faculty supervisors at Chitkara University for feedback on early drafts of this work, and to the authors of the open-access papers that formed the basis of this review.

References

[1] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Hoboken, NJ: Pearson, 2021.

[2] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.

[3] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," Nature Machine Intelligence, vol. 1, no. 9, pp. 389–399, Sep. 2019.

[4] European Commission, Ethics Guidelines for Trustworthy AI, High-Level Expert Group on Artificial Intelligence, Brussels, 2019.

[5] McKinsey Global Institute, The State of AI in 2023: Generative AI's Breakout Year. New York: McKinsey & Company, 2023.

[6] World Economic Forum, Future of Jobs Report 2023. Geneva: WEF, 2023.

[7] M. Gerlich, "AI tools in society: Impacts on cognitive offloading and the future of critical thinking," Societies, vol. 15, no. 1, p. 6, Jan. 2025.

[8] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," Human Factors, vol. 52, no. 3, pp. 381–410, 2010.