

THE RESEARCH UNIVERSITY PRESS

International
Computational Law Journal

VOLUME 2 / ISSUE 2 / CURATED SOURCE EDITION

Building a Responsible Practice Framework: Navigating the Intersection of Laws, Ethics, and AI

Olga Mack

MIT Computational Law Report · Originally published 2023-08-31

Independent reader edition under CC BY 4.0. This attribution cover was added; original source PDF pages follow unchanged. Original authorship and source identity are retained. MIT, Stanford University, Stanford Law School and CodeX have not sponsored or endorsed this republication.

Source, license and version evidence accompany the online article.

MIT Computational Law Report • Media

Building a Responsible Practice Framework: Navigating the Intersection of Laws, Ethics, and AI

Olga Mack

Published on: Aug 31, 2023

URL: https://law.mit.edu/pub/building-a-responsible-practice-framework-navigating-the-intersection-of-laws-ethics-and-ai

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

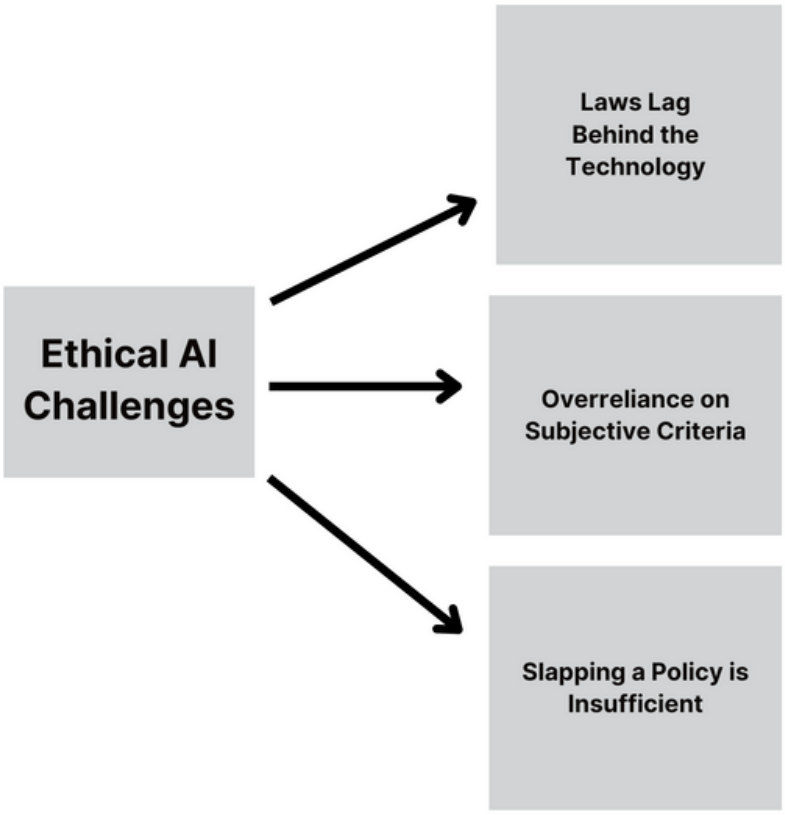
Imagine a groundbreaking AI startup with the potential to transform industries through its innovative products. This startup – let's call it TechNova – has developed a system that analyzes vast amounts of data, identifies patterns, and makes predictions with astonishing accuracy.

TechNova, keen to launch its product, rushes to market without considering the broader implications of what they're releasing.

Unfortunately, the system infringes on privacy rights, as it scrapes data from the internet and isn't adequately programmed to respect privacy laws. Its outputs exhibit bias, reflecting the unconscious prejudices of its creators, leading to several high-profile errors. Eventually, TechNova faces legal actions, incurs irreversible reputational damage, and loses the trust of its clients.

In the end, the TechNova team wishes they had taken the time – before launch – to set up a responsible data usage framework. This approach would have helped them create a product that aligns with societal expectations, legal obligations, and ethical norms. The framework would have been a north star guiding thoughtful and transparent decision-making about the product's data usage and AI application.

Current Challenges at the Intersection of Law, Ethics, and AI



Laws Lag Behind the Technology

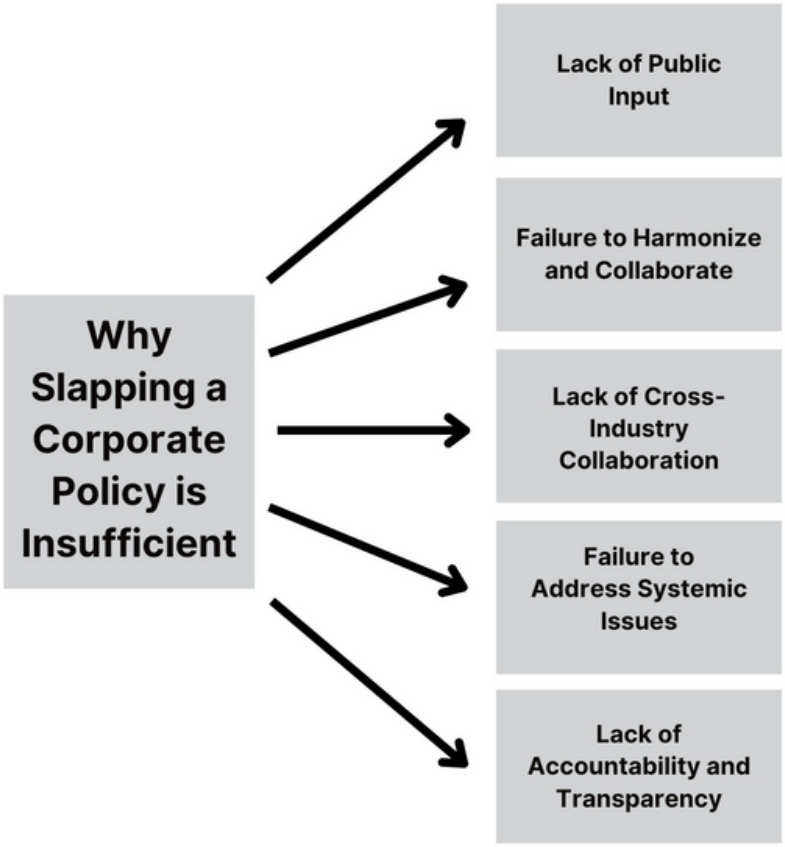
Artificial Intelligence is rapidly reshaping sectors, causing unprecedented dissonance with the legal systems that regulate industries.

Existing legal frameworks, established in an era of more traditional technologies, fail to address the unique, modern challenges AI poses, like who is responsible when AI malfunctions or infringes on privacy laws.

Most existing tech-facing law regulates human behaviors rather than technologies. This traditional approach worked better when technology put humans in the driver's seat, technological evolution was slow, and we could incrementally adjust laws.

Now, technological development has dramatically accelerated with AI, leaving laws trailing behind. Waiting for legislation to provide comprehensive, timely guidance is unrealistic.

In lieu of clear legal direction, the responsibility falls on lawyers, tech companies, and other AI stakeholders to navigate through this uncertain time. Right now, it's up to businesses to define and enforce corporate AI policies that reduce risk and empower responsible decision-making around the ethical, social, and legal implications of their work.



Over-Reliance on Subjective Criteria in Ethical AI Decision-Making

Current efforts to address the ethical implications of AI are falling short because policies are built around subjective criteria.

When companies approach ethical AI, they tend to focus on organizational motives. *Are our intentions good? Will we be embarrassed by the outputs? Will outputs reflect a misunderstanding of the law?*

What these questions really get at is: *how do we mitigate reputational risk?*

Organizations view reputational damage as one of the most significant consequences of AI missteps. As such, their frameworks for managing reputational risk often stand in for a comprehensive ethical framework.

However, this subjective approach can result in an inconsistent and reactive strategy that fails to safeguard against potential damage down the line.

AI managed by corporate policy is also not enough for a larger, more ethical approach to AI. This individualistic framework simply serves as a band-aid to ensure a company's use of AI aligns with current low-hanging ethical and legal responsibilities, but it eventually leads to inconsistencies in AI regulation across sectors and jurisdictions and allows harmful practices to persist until laws finally catch up.

Instead, lawyers, tech companies, and other stakeholders must come together to form a forward-thinking, collaborative, adaptable approach to lawmaking that can maintain pace with AI innovation and focus on data inputs, data processing, and data outputs.

To support a more objective, robust, ethical framework for AI, legal teams must become data stewards, which means taking steps to deeply understand the data that fuels AI models, including data sources, characteristics, biases, and potential impacts.

By adopting this proactive and informed approach, companies will be better positioned to identify and address issues before they grow into major ethical or legal problems. They will be better equipped to build policies that protect the business. Achieving readiness, however, calls for a significant company-wide shift in mindsets and skill sets. It requires building a deep understanding of AI and data, fostering a culture of ethical vigilance, and developing robust systems for monitoring and accountability.

Slapping In a Policy is Insufficient for Doing the Right Thing Systematically

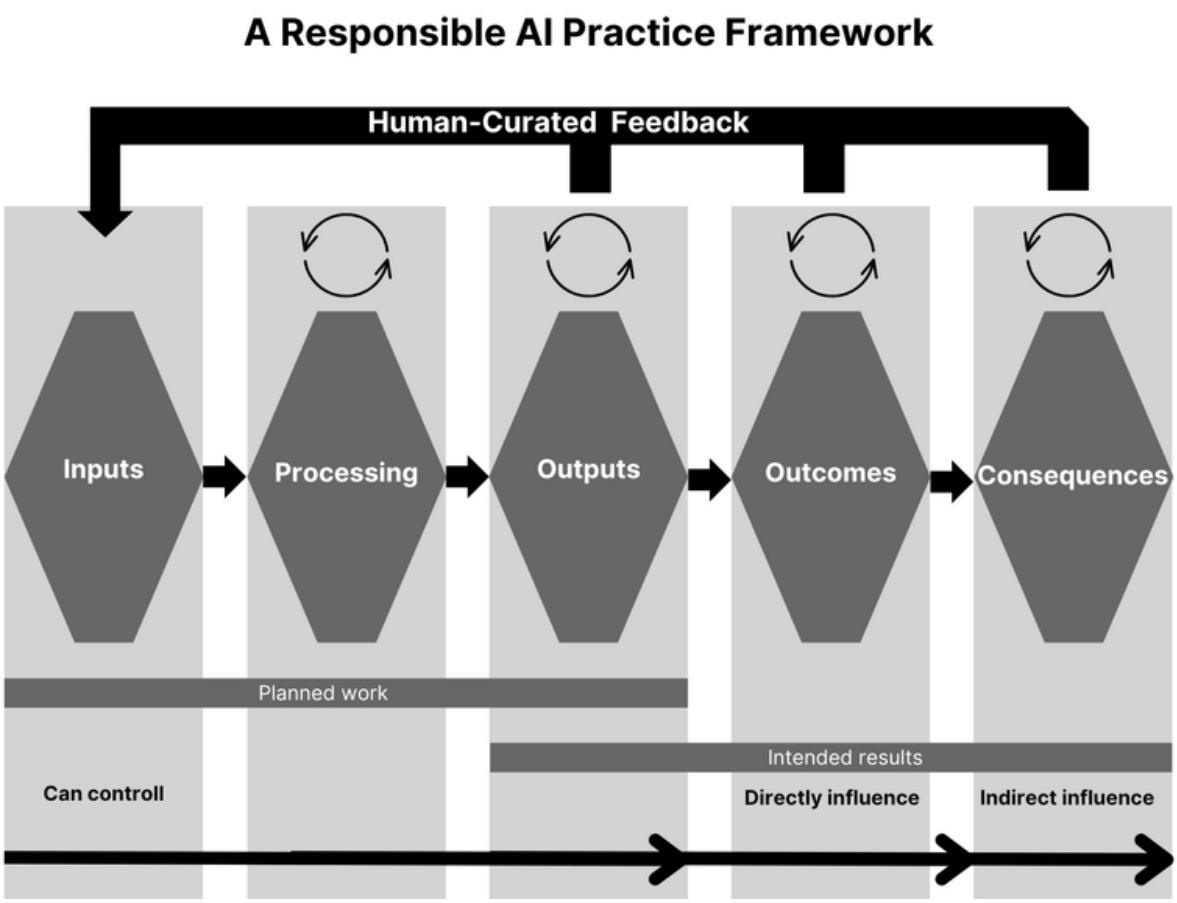
AI brings new, complex challenges in data management—such as the fact that some large language AI models may be built on extensive data scraping, leaving companies liable for the misuse of copyrighted data. Successfully navigating data privacy in the age of AI demands meticulously filtering data into the model—a practice that must be enforced on a larger scale.

Generative AI tools also tend to produce biased outputs, reflective of the unfair world (and biased data) on which they are built. Practitioners must bring equal representation to AI models, especially for populations that tend to be underrepresented in data sources.

Plus, AI systems are iterative and dynamic in nature, continually learning, adapting, and generating "new" data. That "created" data generates increased privacy and security risk for companies. Managing these risks requires a heightened level of stewardship and further complicates data governance.

Finally, AI has a "black box" problem. Users are generally kept in the dark about the mechanics behind outputs. With limited visibility into the data informing the model, its origin, and its subsequent processing, it is nearly impossible for companies to govern data effectively, ethically, and lawfully. AI users need to cultivate and uphold a culture of transparency around AI to achieve more ethical and lawful results.

A Responsible AI Practice Framework

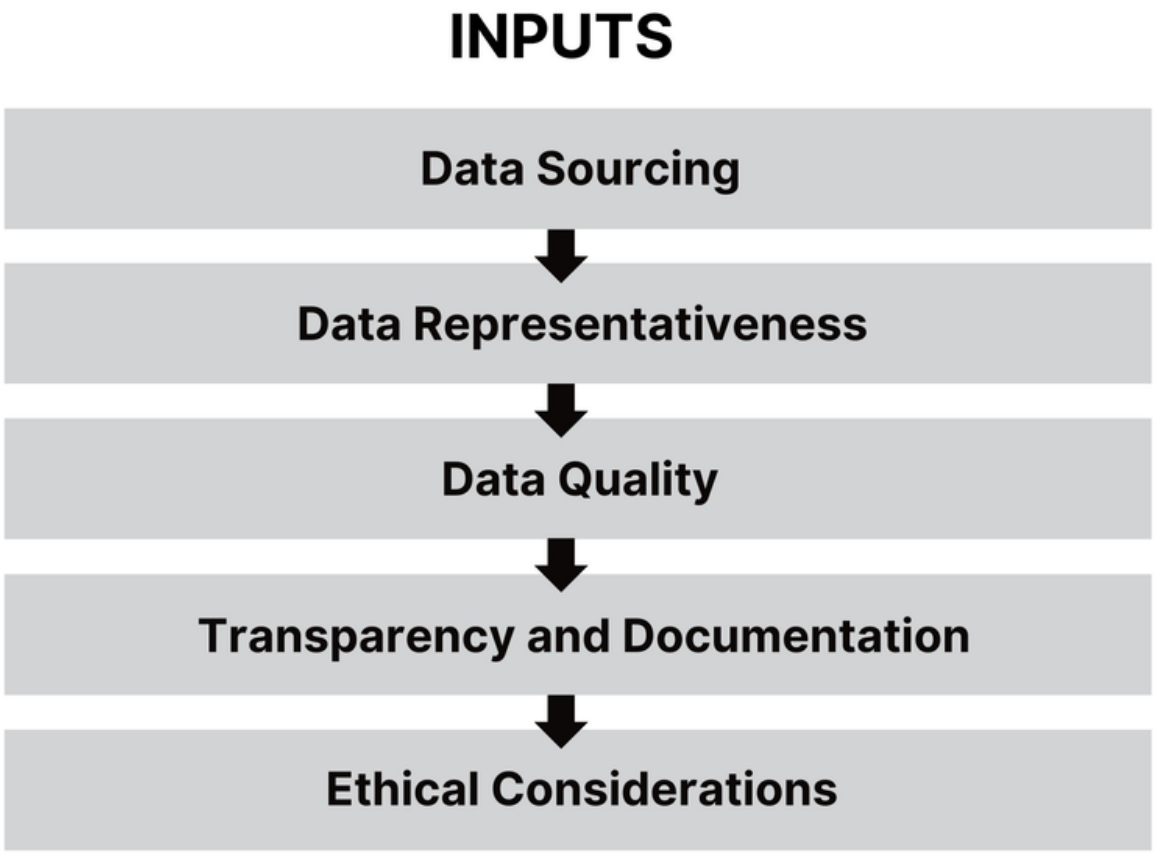


Laws provide a foundational baseline, or a floor, for applying AI in various industries. Ethical considerations extend far beyond the purview of existing laws, encompassing social issues such as fairness, bias, privacy, transparency, and more. By reaching for the ceiling, rather than the bare legal minimum, stakeholders are empowered to design AI applications that uphold collective values, promote trust, and ultimately contribute to a more equitable and just society.

To set up a responsible practice framework that addresses ethical concerns and the dynamics of changing laws, companies must adopt a more conscientious and deliberate approach to decision-making around data and bring a heightened level of transparency about those decisions to every distinct stage of the process (data input, data processing, data output, outcome, and consequences). Isolating and analyzing each stage helps systematically confirm that data management is compliant with applicable laws and ethical principles.

Inputs

A responsible AI practice framework begins with meticulous scrutiny at the input stage. The goal is to ensure data entering the system is in accordance with ethical principles and applicable laws.



Data Sourcing: Origins of input data must be well-understood. This step involves confirming the data is obtained legally, with all necessary permissions, consents, and respect for privacy.

Data Representativeness: Data must accurately represent the intended demographic or population to ensure the model does not reinforce existing biases or create new ones.

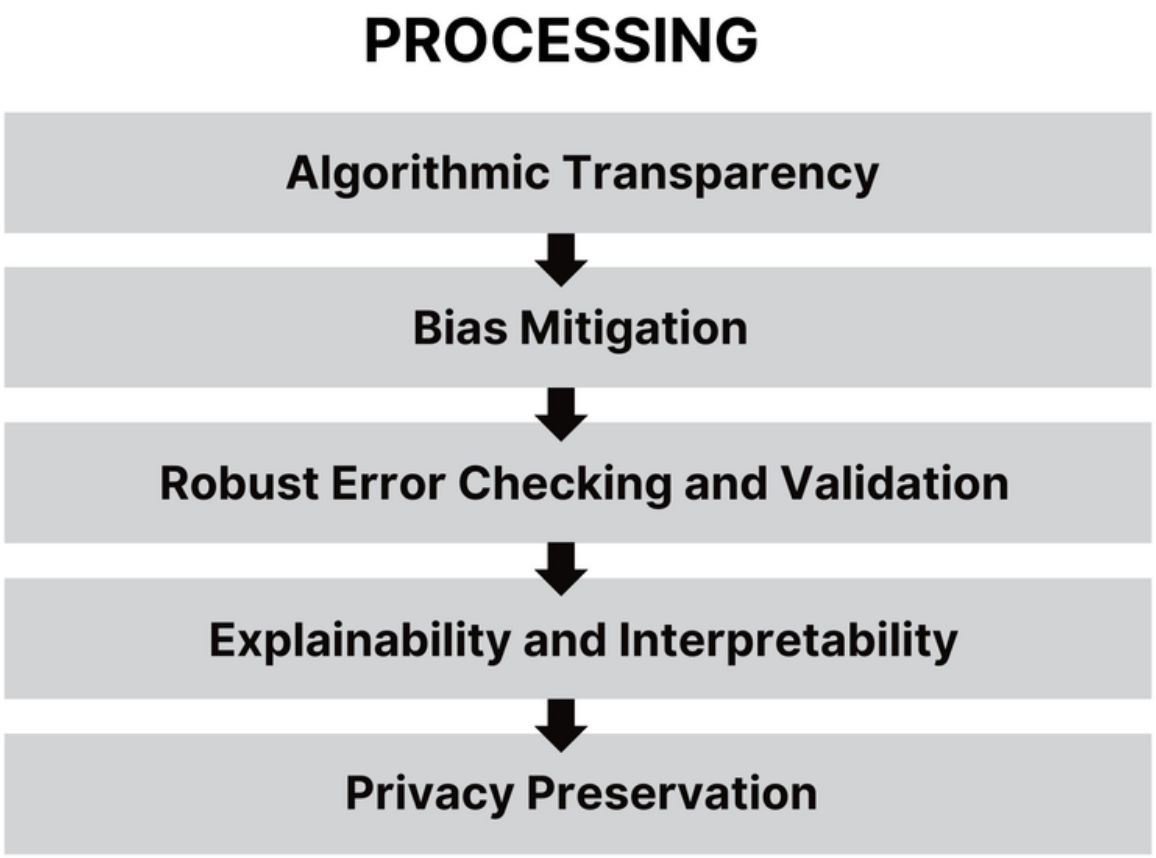
Data Quality: In a reliable and effective AI model, data must be accurate, complete, and relevant. Conduct regular data audits to maintain quality control and assess potential biases.

Transparency and Documentation: Maintain complete transparency regarding the data source, type, quality, and representativeness. Keep and regularly update robust documentation that details every decision around data selection.

Ethical Considerations: Apply higher-level thinking about the ethical implications of using certain data at every step of the input stage, as a checkpoint to ensure it won't harm any individuals or communities or threaten data privacy.

Processing

During the data processing stage, the model analyzes and interprets prepared data inputs, and the AI “learns” to make future predictions.



Algorithmic Transparency: AI models, particularly deep learning models, have been criticized for being "black boxes” — opaque in nature about their inner workings and how the algorithm transforms inputs into outputs. AI practitioners must push for algorithmic transparency, or interpretability (at the very least), to understand how models process data and make decisions.

Bias Mitigation: Monitor and adjust for bias because AI models can unintentionally learn and perpetuate biases present in the training data. Use techniques such as fairness metrics, adversarial debiasing, and reweighting to identify and reduce bias in AI models.

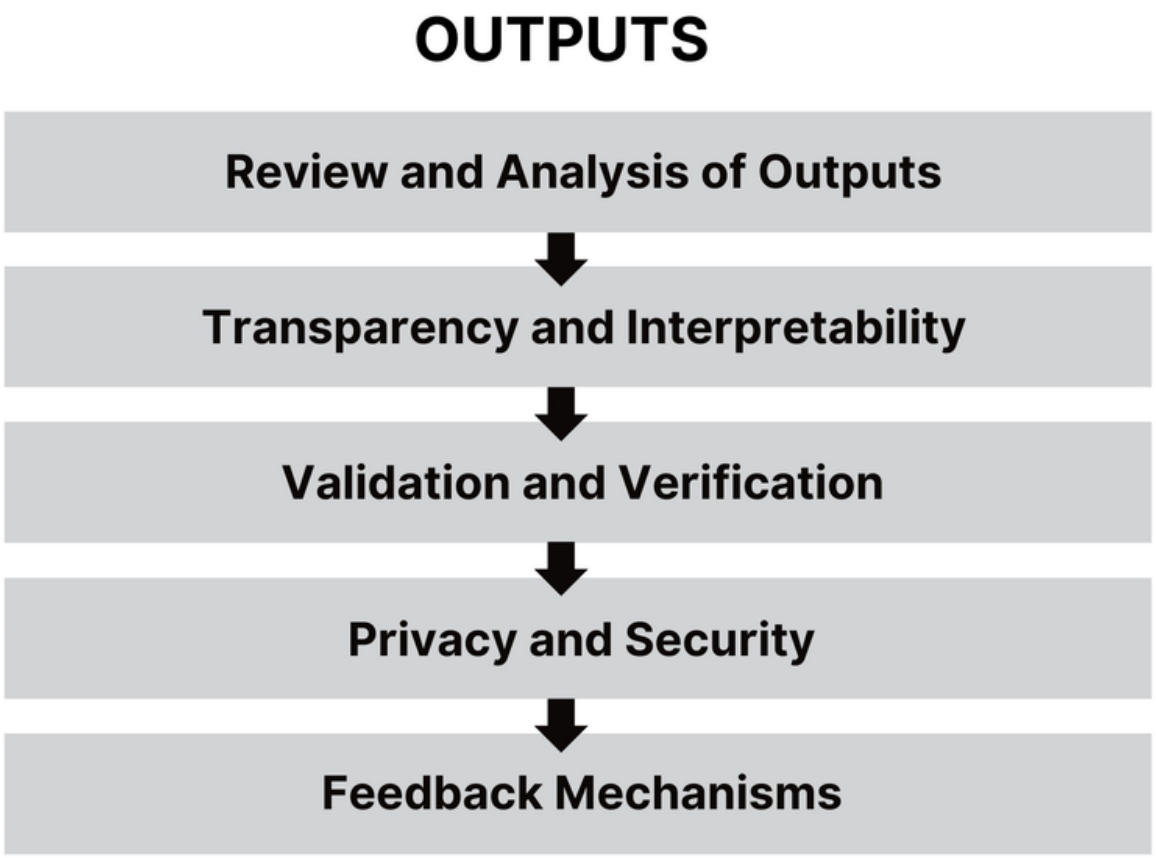
Robust Error Checking and Validation: Regularly test and validate AI models to ensure they process data accurately and reliably. This work includes checking for errors in outputs and making the necessary adjustments to the model’s parameters to fix them.

Explainability and Interpretability: Built-in explainability or interpretability could solve the AI black box problem by providing clear reasons for outputs, ultimately making it easier to identify when and why mistakes occur and allowing for more accountability in AI systems.

Privacy Preservation: Addressing privacy during the data processing stage enhances the fairness, legality, and ethicality of AI models and allows AI to be more transparent, understandable, and accountable and fostering trust in the technology. Thankfully, techniques such as differential privacy protect sensitive information during data processing.

Outputs

During the output stage, AI systems produce results, making predictions or decisions based on processed data.



Review and Analysis of Outputs: Thoroughly review and analyze results for unexpected outputs, biased outcomes, or incorrect predictions. Address issues promptly by revisiting the data input and/or processing stages.

Transparency and Interpretability: In an ideal world, the AI system would provide understandable and detailed explanations that clarify how it arrived at a decision or prediction. Transparency and interpretability at the output stage help flag potential issues or biases in the results.

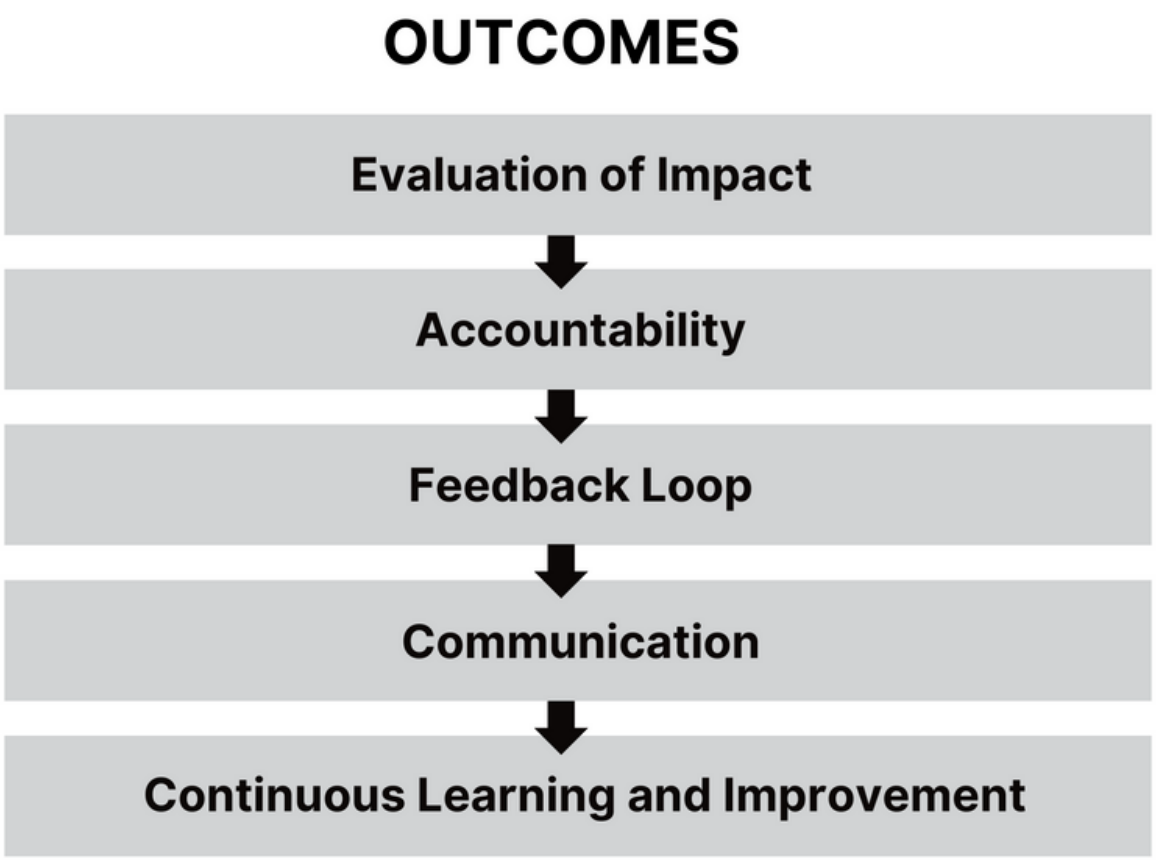
Validation and Verification: To ensure the system's outputs are reliable and accurate, put validation and verification procedures in place at the output stage, such as comparing the system's results with human-made decisions or known outcomes.

Privacy and Security: Outputs may include personal or sensitive data, making privacy and security measures crucial at this stage. Securely store and handle outputs and anonymize any sensitive or personal data appropriately.

Feedback Mechanisms: To continually improve the performance and fairness of an AI model, implement a feedback mechanism at the output stage so the system can take note of and learn from errors and improve future outputs.

Outcomes

The outcome stage refers to the real-world implications of AI-generated outputs, or how generated data could affect individuals, groups, or systems.



Evaluation of Impact: Game out the intended and unintended consequences of using the generated data. If a negative impact is identified, review previous stages (input, processing, and output) to identify and rectify the source of the issue.

Accountability: If the system produces negative or harmful outcomes, there should be clear lines of accountability for addressing these issues, which might involve regulatory oversight or internal review processes.

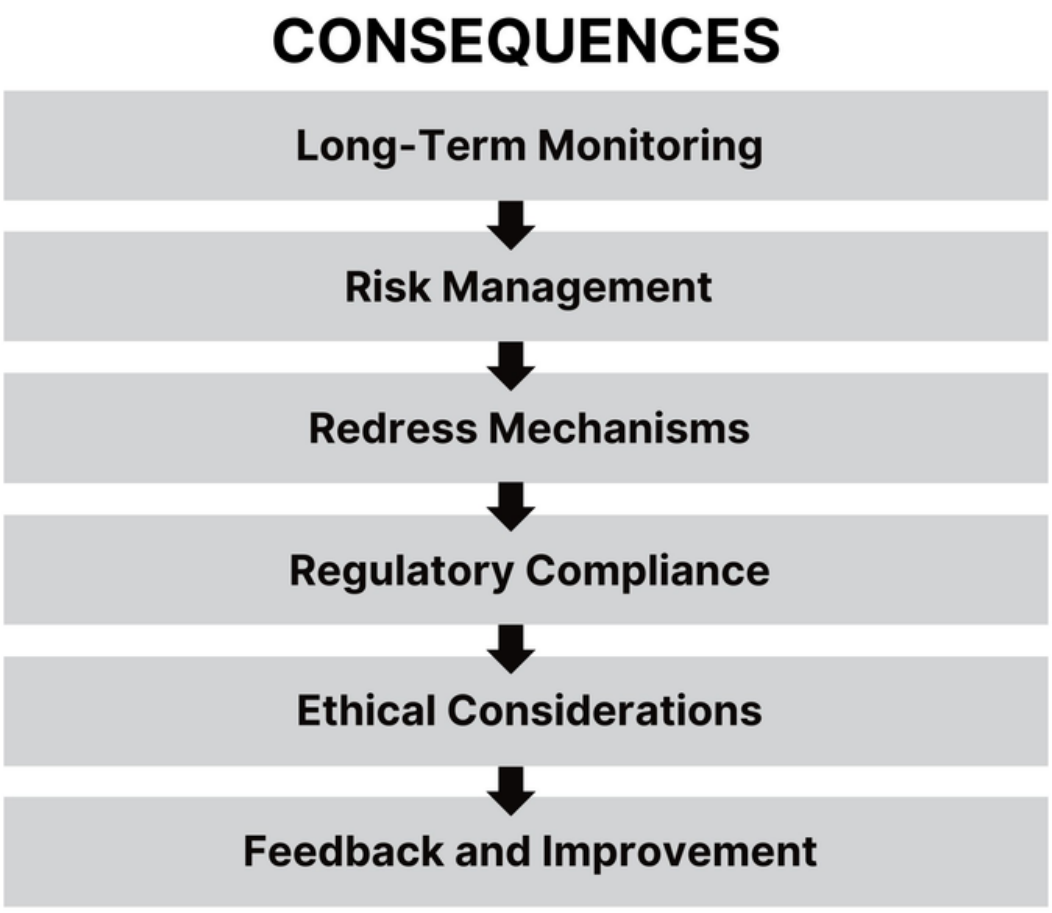
Feedback Loop: The outcome stage, like the output stage, should feed back into the entire AI process. Feedback at this stage helps fine-tune the system, informing future input, processing, and output stages with learnings from real-world outcomes.

Communication: Clearly communicate results and impacts, both positive and negative, to all stakeholders, including those directly affected by the outcomes and those involved in the design and operation of the AI system. Communication and transparency help maintain trust and enables more informed decision-making.

Continuous Learning and Improvement: AI systems are living, breathing, continually evolving organisms. The outcome stage serves as a platform for learning and improving the overall system—an opportunity to refine the AI model, the data it uses, and the processing methods applied.

Consequences

The consequences stage looks at the long-term effects of AI systems on individuals, groups, systems, or even society as a whole, including both direct consequences of the system's outputs and indirect, emergent effects of the system's widespread use or misuse.



Long-Term Monitoring: Outcomes of AI systems may take time to accumulate and become apparent. Continuous, long-term monitoring helps to identify and address negative effects that emerge over time.

Risk Management: Implement risk management strategies, such as scenario planning, impact assessments, or the development of contingency plans to anticipate, mitigate, and manage potential negative consequences.

Redress Mechanisms: Establish redress mechanisms, such as compensation, system adjustments, or even system shutdowns (in extreme cases) in case of negative consequences. Those affected by AI system

consequences should have a clear, accessible route to seek redress.

Regulatory Compliance: As the AI legal landscape continues to evolve, it's imperative to ensure that AI systems and their long-term effects comply with all relevant laws and regulations.

Ethical Considerations: Carefully consider and address the ethical implications of the AI system's consequences, which might involve balancing different interests, assessing the fairness of outcomes, or considering the system's impact on privacy, autonomy, and social norms.

Feedback and Improvement: Like in the output and outcome stage, the consequences stage should feed back into the entire AI system, informing future development and refinement, as the lessons learned from long-term consequences can help improve ethical and legal compliance.

Conclusion

The story of TechNova serves as a cautionary tale that highlights the need for a responsible framework for the development and deployment of AI. Rushing to market without adequate consideration of ethical and legal implications can have major consequences, not only for companies but also for individuals, communities, and society at large. The challenges at the intersection of law, ethics, and AI demand a proactive and comprehensive approach that goes beyond mere compliance with existing regulations.

By implementing a responsible framework that encompasses all stages of the AI process, from data input to processing, outcome, and consequences, companies can ensure that their AI systems align with societal expectations, legal obligations, and ethical norms. That framework should prioritize data stewardship, transparency, fairness, and ongoing monitoring of the system's impact. A culture of continuous learning, improvement, and accountability is also crucial for addressing potential biases, mitigating risks, and addressing any negative consequences that may arise over time.

Only through such a responsible approach can we harness the transformative power of AI while safeguarding privacy, promoting fairness, and maintaining trust in groundbreaking technologies.