

## ***Embedding Ethics and Equity in Artificial Intelligence and Machine Learning Infrastructure***

(*Big Data*, volume 11, supplement 1, 2023)

---

**Question:** *Engineering infrastructure today is both structural and digital. Take artificial intelligence and machine learning (AI/ML) infrastructure, for example, ranging in use from automated robotic surgeries to advanced data analytics in health research. What comes to mind in terms of applying ethical practices across these examples, based on your subject matter expertise?*

### **ChatGPT**

When it comes to applying ethical practices across examples of AI/ML infrastructure, several considerations are important. Here are some key points to keep in mind:

1. **Transparency and Explainability:** Ensuring transparency and explainability in AI/ML systems is crucial. It's important to develop models and algorithms that can provide understandable explanations for their decisions or recommendations. This helps build trust and enables users to understand how and why certain outcomes are produced.
2. **Privacy and Data Protection:** AI/ML infrastructure often relies on large amounts of data. It's essential to implement robust privacy and data protection measures to safeguard individuals' sensitive information. Organizations should adhere to relevant regulations and ensure that data is collected, stored, and processed securely, with user consent obtained where necessary.
3. **Fairness and Bias Mitigation:** Bias in AI/ML algorithms can lead to unfair outcomes and discrimination. Ethical practices involve actively identifying and mitigating biases in data and algorithms. This includes regular auditing of models to ensure they do not disproportionately favor or discriminate against specific groups based on attributes such as race, gender, or socioeconomic status.
4. **Safety and Reliability:** In applications like automated robotic surgeries, safety and reliability are paramount. Ethical practices involve rigorous testing, validation, and monitoring of AI systems to minimize the risk of errors or malfunctions that could harm patients or users. Regular maintenance and updates should also be conducted to ensure the systems remain safe and effective.
5. **Human Oversight and Accountability:** While AI/ML infrastructure can automate tasks, it's important to maintain human oversight and accountability. Establishing clear lines of responsibility and accountability is crucial to address any issues that arise and ensure that humans have the final decision-making authority in critical scenarios.
6. **Collaboration and Interdisciplinary Approach:** Ethical practices in AI/ML infrastructure should involve collaboration among engineers, data scientists, ethicists, domain experts, and other stakeholders. This interdisciplinary approach helps identify potential ethical challenges, encourages diverse perspectives, and promotes responsible development and deployment of AI technologies.
7. **Continuous Monitoring and Improvement:** Ethical practices are not one-time efforts but require ongoing monitoring and improvement. Regularly reviewing the performance and impact of AI/ML infrastructure, collecting feedback from users, and incorporating lessons learned into future iterations helps to address emerging ethical concerns and maintain alignment with evolving societal values.

**Question:** *These are general principles, and specific ethical considerations may vary depending on the context and application of AI/ML infrastructure. It's essential for organizations and practitioners to be proactive in considering and addressing ethical concerns throughout the development and deployment lifecycle.*

*What kind of evidence or use cases might support the development of ethical principles or standards to guide AI/ML infrastructure development and implementation in health research settings?*

The development of ethical principles or standards to guide AI/ML infrastructure in health research settings should be informed by evidence and use cases that highlight the benefits and potential risks associated with these technologies. Here are some examples of evidence and use cases that can support the development of ethical principles:

1. **Case studies on Bias and Fairness:** Demonstrating instances where AI/ML algorithms have exhibited bias or resulted in unfair outcomes can highlight the importance of addressing these issues. Use cases that illustrate how biases can affect underrepresented populations or lead to disparities in healthcare outcomes can underscore the need for fair and unbiased AI/ML systems.
2. **Privacy Breaches and Data Protection:** Documenting instances of privacy breaches or data mishandling in health research involving AI/ML infrastructure can highlight the risks associated with inadequate privacy and data protection measures. These cases can emphasize the need for robust data security protocols and adherence to privacy regulations.
3. **Patient Outcomes and Safety:** Use cases that showcase positive patient outcomes achieved through AI/ML infrastructure in health research can illustrate the potential benefits of these technologies. For example, successful applications of AI/ML in diagnosing diseases, predicting treatment responses, or optimizing healthcare workflows can demonstrate the value of ethical and well-implemented AI/ML systems.
4. **Ethical Dilemmas and Decision-Making:** Real-world scenarios that present ethical dilemmas in health research involving AI/ML infrastructure can help shape ethical principles. These cases may involve situations where AI systems are required to make critical decisions or when conflicts arise between algorithmic recommendations and human judgment. Analyzing these cases can inform guidelines on human oversight, accountability, and the boundaries of AI decision-making.
5. **Stakeholder Perspectives and User Feedback:** Gathering input from various stakeholders, including patients, healthcare providers, researchers, and ethicists, through surveys, interviews, or focus groups can provide valuable insights into their concerns, expectations, and experiences related to AI/ML infrastructure. Understanding the perspectives of those directly involved can help shape ethical principles that address their needs and ensure inclusivity.
6. **Regulatory and Industry Guidelines:** Existing regulations, guidelines, and frameworks developed by regulatory bodies or industry organizations can provide a foundation for the development of ethical principles. Examining these documents and identifying gaps or areas of improvement can help refine and expand the ethical considerations specific to AI/ML infrastructure in health research.

***Question:** These evidence and use cases can be leveraged to inform discussions, policy development, and the establishment of ethical principles and standards that guide the responsible development and implementation of AI/ML infrastructure in health research settings. It's important to ensure a comprehensive understanding of the benefits, risks, and potential ethical challenges associated with these technologies to promote their ethical and responsible use.*

*The engineering field has, as a whole, shifted from ethical awareness of the individual engineer to systems-level ethical awareness. How might this awareness extend into the practice of AI/ML engineering and systems design?*

The shift towards systems-level ethical awareness in the engineering field can and should extend into the practice of AI/ML engineering and systems design. Here are some ways in which this awareness can be incorporated:

1. **Ethical Considerations in Design:** AI/ML engineers and system designers should proactively consider ethical implications during the design phase. This includes identifying potential biases, privacy concerns, fairness issues, and other ethical challenges specific to AI/ML systems. By integrating ethical considerations from the outset, engineers can create systems that align with ethical principles and minimize potential harms.
2. **Interdisciplinary Collaboration:** AI/ML engineering and systems design should involve collaboration with experts from diverse disciplines, such as ethics, social sciences, law, and policy. This interdisciplinary approach helps to incorporate a range of perspectives and ensures that ethical considerations are adequately addressed. Collaboration can foster a more comprehensive understanding of the potential societal impacts of AI/ML systems.
3. **Ethical Guidelines and Standards:** Developing and adhering to ethical guidelines and standards specific to AI/ML engineering and systems design is crucial. These guidelines can provide a framework for ethical decision-making, outlining best practices and principles that engineers should follow. They can cover areas such as transparency, fairness, privacy, and accountability, serving as a reference for responsible AI/ML development.
4. **Ethical Impact Assessments:** Conducting ethical impact assessments can help AI/ML engineers and system designers identify and evaluate the potential ethical consequences of their work. These assessments involve considering the social, cultural, and ethical implications of deploying AI/ML systems. By systematically examining the broader impact of their designs, engineers can proactively address ethical concerns.
5. **Continuous Monitoring and Evaluation:** Once AI/ML systems are deployed, ongoing monitoring and evaluation are essential to assess their ethical impact and ensure alignment with ethical standards. Engineers should regularly assess the performance, fairness, and ethical implications of their systems, and make necessary adjustments or improvements based on feedback and emerging ethical concerns.
6. **Ethical Decision-Making Frameworks:** AI/ML engineers should be equipped with ethical decision-making frameworks that help guide their choices in complex situations. These frameworks can provide a structured approach to consider the ethical implications of different design choices, ensuring that ethical considerations are consistently taken into account.
7. **Professional Education and Training:** Incorporating ethics education and training into AI/ML engineering programs and professional development initiatives is crucial. By fostering ethical awareness and providing engineers with the necessary knowledge and skills to address ethical challenges, the field can cultivate a culture of responsible AI/ML development.

*Question: By extending systems-level ethical awareness into the practice of AI/ML engineering and systems design, engineers can actively consider and mitigate the ethical implications of their work. This approach helps ensure that AI/ML systems are developed and deployed in a manner that is responsible, fair, transparent, and aligned with societal values.*

*Prior work shows there are five ethical themes present in current engineering systems design literature: - Integrating ethics and equity-centered perspectives into design - Recognizing system boundaries - Developing augmented system design criteria - Managing trade-offs and conflicting values - Educating systems designers - Applicability to engineering systems of health. In your perspective, how might these themes apply, or not, to the scope and practice of AI/ML engineering and systems design?*

## ChatGPT

The five ethical themes identified in current engineering systems design literature can be relevant to the scope and practice of AI/ML engineering and systems design. Let's explore how these themes apply:

1. **Integrating ethics and equity-centered perspectives into design:** This theme is highly applicable to AI/ML engineering and systems design. It emphasizes the importance of considering ethical and equity considerations throughout the design process. In the context of AI/ML, this includes addressing biases,

fairness, and potential impacts on marginalized or vulnerable populations. Designers should actively seek diverse perspectives and engage with stakeholders to ensure that AI/ML systems are developed in an inclusive and equitable manner.

2. **Recognizing system boundaries:** Understanding system boundaries is crucial in AI/ML engineering and systems design. AI/ML systems often interact with complex socio-technical ecosystems, and it's essential to consider the broader context in which these systems operate. Recognizing system boundaries involves understanding the interdependencies, potential impacts, and ethical implications of AI/ML systems within their larger environment, including legal, social, and cultural factors.
3. **Developing augmented system design criteria:** Augmented system design criteria are important in AI/ML engineering to ensure that systems go beyond technical performance and also account for ethical dimensions. In addition to traditional design criteria, AI/ML engineers should consider factors such as transparency, explainability, interpretability, and robustness. These criteria can help address ethical concerns, improve system accountability, and foster trust in AI/ML technologies.
4. **Managing trade-offs and conflicting values:** Trade-offs and conflicting values are inherent in AI/ML engineering and systems design. Engineers often face dilemmas when balancing competing priorities, such as accuracy versus interpretability or privacy versus utility. Managing these trade-offs requires ethical deliberation, clear decision-making frameworks, and engagement with stakeholders to navigate and resolve conflicting values in a responsible manner.
5. **Educating systems designers:** Education plays a critical role in AI/ML engineering and systems design. It is important to provide education and training that not only focuses on technical aspects but also incorporates ethical considerations. AI/ML engineers should be equipped with the knowledge and skills to identify and address ethical challenges, understand the societal impact of their work, and navigate the complex ethical landscape associated with AI/ML technologies.

***Question:** Regarding the applicability of these themes to the engineering systems of health, many of these themes align with the specific considerations and challenges encountered in AI/ML engineering for healthcare systems. The integration of ethics and equity-centered perspectives is crucial for addressing healthcare disparities and ensuring fair access to AI-driven healthcare technologies. Recognizing system boundaries is essential when designing AI/ML systems that interact with sensitive medical data and comply with regulatory frameworks. Developing augmented system design criteria helps address ethical concerns related to patient safety, privacy, and clinical effectiveness. Managing trade-offs and conflicting values is particularly relevant in healthcare, where decisions impact patient outcomes and well-being. Finally, educating systems designers in the context of healthcare ensures that ethical considerations and patient-centered perspectives are central to the development and implementation of AI/ML technologies in health settings.*

*In summary, these ethical themes provide a valuable framework for guiding the scope and practice of AI/ML engineering and systems design, including their application in engineering systems of health. By incorporating these themes, engineers can develop and deploy AI/ML systems that are not only technically robust but also ethically sound and aligned with societal needs.*

*Ethics, fairness, and trustworthiness are ethical imperatives for critical infrastructure within systems of health (medical ethical principles of beneficence, nonmaleficence, respect, justice). In your perspective, what are ethical imperatives to which AI/ML infrastructure engineers and system designers should adhere?*

## **ChatGPT**

AI/ML infrastructure engineers and system designers should adhere to several ethical imperatives to ensure the responsible development and deployment of AI/ML systems. These imperatives include:

1. **Beneficence:** AI/ML engineers should prioritize the well-being and benefit of individuals and society. They should strive to develop AI/ML systems that improve healthcare outcomes, enhance patient safety, and contribute to the overall betterment of human health. Engineers should consider the potential

positive impacts of their systems and actively work towards maximizing benefits while minimizing harm.

2. **Nonmaleficence:** Engineers have an ethical obligation to ensure that AI/ML systems do not cause harm to individuals or society. They should be vigilant in identifying and mitigating potential risks, such as biases, errors, or unintended consequences. Implementing thorough testing, validation, and monitoring processes helps mitigate the potential for harm and ensures the safety and reliability of AI/ML systems.
3. **Respect for Autonomy:** AI/ML infrastructure engineers should respect individuals' autonomy and privacy. Systems should be designed to uphold confidentiality, protect personal information, and ensure informed consent is obtained when necessary. Engineers should also consider user preferences, allowing individuals to exercise control and have meaningful input in decisions that affect them.
4. **Fairness and Justice:** AI/ML engineers should strive to ensure fairness and justice in the development and deployment of systems. They should actively identify and mitigate biases, both in data and algorithms, that may lead to unfair outcomes or perpetuate existing social inequalities. Engineers should work towards developing AI/ML systems that are unbiased, equitable, and considerate of diverse perspectives and needs.
5. **Transparency and Explainability:** Engineers have an ethical imperative to design AI/ML systems that are transparent and explainable. Users and stakeholders should be able to understand how the system works, the factors influencing its outputs, and the reasoning behind its decisions. Transparent systems help build trust, enable effective oversight, and facilitate accountability.
6. **Accountability and Responsibility:** AI/ML engineers should embrace accountability and take responsibility for the impact of their systems. This involves designing AI/ML systems with clear lines of responsibility, ensuring appropriate human oversight, and addressing any issues that arise promptly and transparently. Engineers should be mindful of potential consequences and be prepared to rectify problems or errors that may occur.
7. **Continuous Learning and Improvement:** Ethical imperatives for AI/ML infrastructure engineers include a commitment to continuous learning and improvement. The field of AI/ML is rapidly evolving, and engineers should stay informed about emerging ethical concerns, best practices, and advancements in the field. Engaging in ongoing professional development and actively seeking feedback from users and stakeholders helps ensure that AI/ML systems remain ethically sound and responsive to societal needs.

***Question:** Adhering to these ethical imperatives helps AI/ML infrastructure engineers and system designers promote the ethical principles of beneficence, nonmaleficence, respect, and justice within the context of AI/ML systems in healthcare. By integrating these imperatives into their work, engineers can contribute to the development of AI technologies that are trustworthy, accountable, and aligned with societal values.*

*Prior work shows that persons with lived experiences in inequity are often excluded from and/or not effectively involved in data generation (including selecting and acquiring data sources), sharing, and use processes that involve their data. What comes to mind as far as strategies to more meaningfully engage these persons in AI/ML infrastructure development and use?*

## ChatGPT

To more meaningfully engage persons with lived experiences in inequity in AI/ML infrastructure development and use, several strategies can be employed:

1. **Inclusive and Participatory Design:** Involve individuals with lived experiences in the design process of AI/ML infrastructure. Seek their input, perspectives, and feedback to ensure that the technology addresses their needs, concerns, and values. Adopt participatory design approaches that empower individuals to actively contribute to decision-making and influence the development of AI/ML systems.
2. **Collaborative Partnerships:** Forge partnerships with community organizations, advocacy groups, and relevant stakeholders representing the interests of individuals with lived experiences in inequity. These

partnerships can provide valuable insights, facilitate community engagement, and ensure that AI/ML initiatives are aligned with the priorities and aspirations of the affected communities.

3. **Co-creation and Co-development:** Engage individuals with lived experiences as active collaborators and co-developers of AI/ML infrastructure. Foster environments where their expertise and knowledge are valued, respected, and integrated into the development process. This can involve establishing co-creation spaces, workshops, or hackathons where diverse stakeholders work together to shape AI/ML solutions.
4. **Ethical Data Governance:** Establish transparent and inclusive data governance frameworks that involve individuals with lived experiences in inequity. Ensure that they have a voice in decisions regarding data generation, acquisition, sharing, and use. Develop mechanisms for obtaining informed consent and enabling individuals to have control over how their data is collected, shared, and utilized.
5. **Capacity Building and Education:** Provide opportunities for individuals with lived experiences to enhance their understanding of AI/ML technologies, their implications, and their potential benefits and risks. Offer training programs, workshops, and educational resources that empower them to actively participate in discussions about AI/ML infrastructure development and use.
6. **Addressing Power Imbalances:** Recognize and address power imbalances that may exist between AI/ML practitioners and individuals with lived experiences in inequity. Foster a culture that values diverse perspectives and creates spaces for open and respectful dialogue. Facilitate opportunities for mutual learning and shared decision-making to ensure that the voices of marginalized individuals are amplified and integrated into AI/ML development processes.
7. **Advocacy and Policy Engagement:** Encourage and support individuals with lived experiences to engage in advocacy efforts and policy discussions related to AI/ML infrastructure. Empower them to raise concerns, contribute to policy development, and advocate for equitable and accountable AI/ML practices.

These strategies aim to shift the narrative from "doing for" marginalized communities to "working with" them, recognizing their expertise, agency, and rights. By actively involving individuals with lived experiences in inequity in AI/ML infrastructure development and use, the resulting technologies can better address their needs, mitigate biases, and promote more equitable and inclusive outcomes.