

Ethical AI Leadership

Diana Wolfe, Mintesinot Misgano, and Justin Wang
Industrial Organizational Psychology

Abstract

The present theoretical framework evaluated literature on artificial intelligence (AI) ethical frameworks and assembled leadership guidelines for facilitating ethical AI practices. The recommended leadership behaviors are in alignment with the ethical pillars of transparency, non-maleficence, justice, and privacy. Leaders foster ethical AI utilization in organizations by sharing appropriate information, protecting others from harm, creating fairness, and respecting confidentiality. Given the increased implementation of AI in organizations, more research is needed to understand the impacts of leader behaviors on ethical AI practices.

Transparency



Share Information

Be open with stakeholders

How decisions were made
Why these were the best choices
Who is responsible for the choices made

During team meetings

Establish consensus for what is appropriate to share with stakeholders

Remind your team

Transparency builds trust
Correcting errors is part of the process

Do No Harm



Protect Others

Be prepared

Inform stakeholders of limitations and potential adverse impacts
Check for harm at design, launch, and application

Consult multiple teams, departments, and professions about areas of risk

During team meetings

Recognize identification and prevention of discrimination/bias/harm

Remind your team

Harm comes from discrimination, bias, and violation of privacy

Justice



Create Fairness

Be supportive of employees

Create an anonymous ethical violation reporting system
Reward behaviors team member integrity and honesty

During team meetings

Have team members identify fairness outcomes across groups
Acknowledge and disclose known areas of bias

Remind your team

Strive for objective AI use

Privacy



Respect Confidentiality

Be cautious with data

Implement mechanisms that limit privacy risks
Comply with legal regulations
Train employees on the appropriate handling of sensitive data

During team meetings

Provide tools to enable employees to evaluate the risk to privacy

Remind your team

Importance of notifying stakeholders of data breaches.

Introduction

- Leaders in organizations must immediately prioritize ethical considerations in their business model.
- Leaders are essential in the success of OD initiatives for two reasons : (a) they make tough decisions which tests their commitment , and (b) they build consensus for change.
- Leaders play a key role in realizing ethical AI codes in organizations.
- In the literature review on global ethical frameworks for AI, four common themes arose: transparency, non-maleficence, justice, and privacy.

Methods

Ethical Principles Identified in Existing AI Guidelines

| Ethical principle | Number of documents | Included codes |
|--------------------|---------------------|---|
| Transparency | 73/84 | Transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing |
| Justice & fairness | 66/84 | Justice, fairness, consistency, inclusion, equality, equity, non-bias, (non-discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution |
| Non-maleficence | 60/84 | Non-maleficence, security, safety, harm, protection, precaution, prevention, integrity (bodily or mental), non-subversion |
| Responsibility | 60/84 | Responsibility, accountability, liability, acting with integrity |
| Privacy | 47/84 | Privacy, personal or private information |
| Beneficence | 41/84 | Benefits, beneficence, well-being, peace, social good, common good |
| Freedom & autonomy | 34/84 | Freedom, autonomy, consent, choice, self-determination, liberty, empowerment |
| Trust | 28/84 | Trust |
| Sustainability | 14/84 | Sustainability, environment (nature), energy, resources (energy) |
| Dignity | 13/84 | Dignity |
| Solidarity | 6/84 | Solidarity, social security, cohesion |

Jobin et al. (2019)

The described conceptual framework was generated through a process of literature review from both academic and grey literature. Five comprehensive literature reviews on the subject of AI ethics were compared for common themes. Among these identified trends were the ethical pillars of transparency, non-maleficence, justice, and privacy. Research on the role of leaders in organizational change was conducted and a connection between the leader commitment to organizational change and the success of the adoption was described. Due to the nascent nature of this area of research, leader behaviors for ethical AI policies were partially sourced from grey literature including federal and corporate policies.

Discussion

- Leaders seeking to promote transparency can: (a) enable a culture of support for appropriate transparency, (b) train others on standards for data methods, (c) educate stakeholders, (d) ensure data is equitably available for research, (e) employs a system for sharing analytic methods and models with stakeholders, and (f) encourage the correction of errors and explain changes.
- Leaders that want to advance non-maleficence can: (a) establish bias mitigation procedures, (b) adhere to ethical guidelines that honor the privacy of others, (c) increase accountability for outcomes, (d) report limitations and adverse effects, and (g) seek guidance from other scientific communities.
- An inexhaustive list of common practices to ensure justice within AI includes : (a) develop a culture of honesty and integrity; (b) understand and disclose any known deficits or biases; (c) provide mechanisms for employees to anonymously report ethical violations; and (d) use techniques to limit bias and strive for objective AI practices.
- Some strategies leaders can undertake to address privacy in AI include : (a) provide tools for employees to evaluate the risk of exposing privacy; (b) establish protocols for notifying stakeholders if there is a privacy breach; (c) implement mechanisms that limit privacy risks and comply with relevant legal regulations
- Leaders play a pivotal role in the adoption of ethical practices in AI. Unethical AI practices can lead to the creation or perpetuation of harm. Ethical frameworks can limit the threats from unregulated AI. Leaders can take action to promote transparency, non-maleficence , justice, and privacy with their teams . As AI becomes increasingly pervasive in our lives, leaders will need to advance directives which honor human rights and steward social well-being .