Public Perceptions of the Social and Ethical Implications of AI

Prabu David
Professor and Dean

College of Communication Arts and Sciences

Arun Ross
Martin J. Vanderploeg Endowed Professor

College of Engineering Department of Computer Science and Engineering
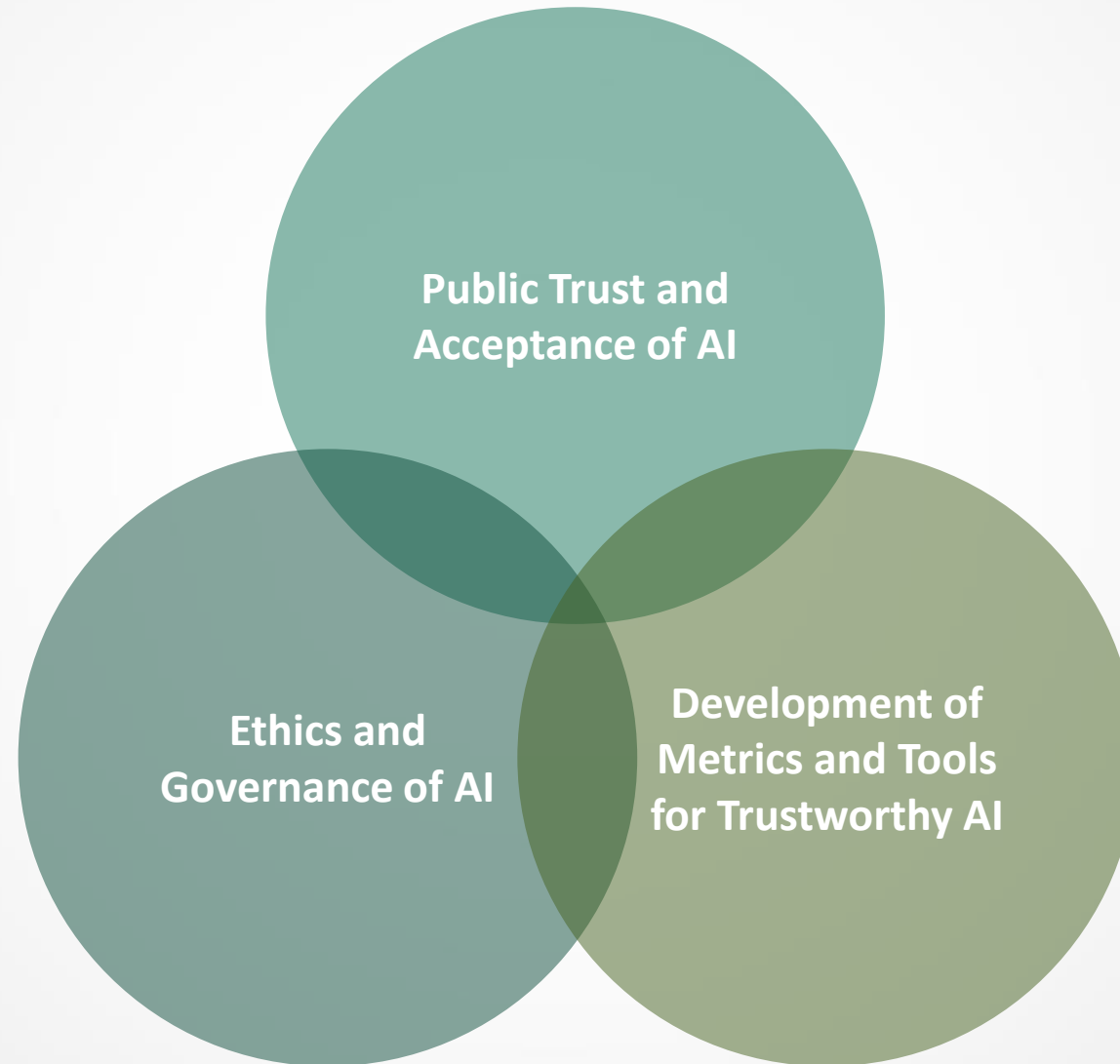
Hyesun Choung
Postdoctoral Research Associate

College of Communication Arts and Sciences

2022 Ethics of Big Data and Artificial Intelligence Conference
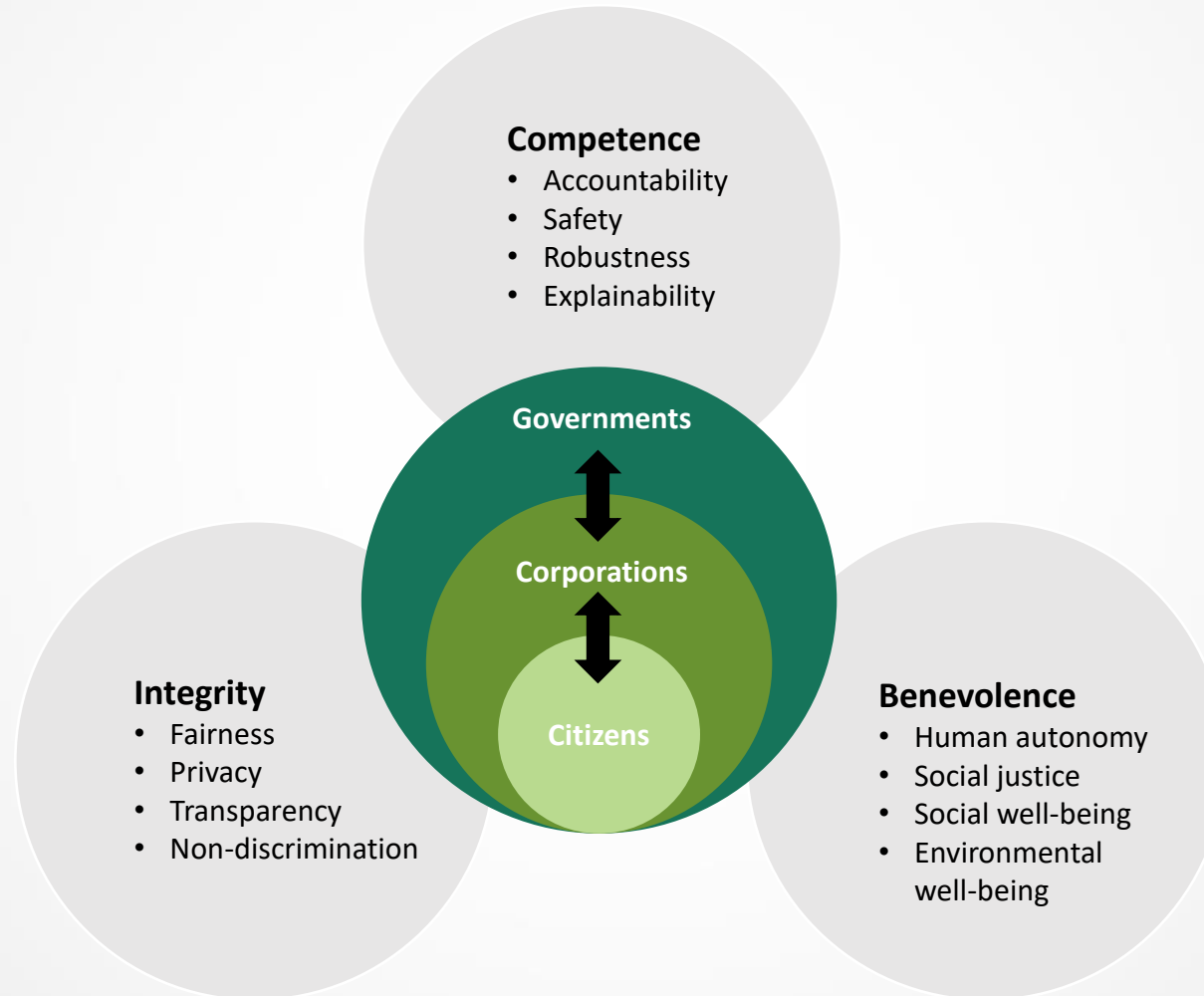November 3, 2022

Ethical Requirements & Governance Framework

Ethics Guidelines for Trustworthy AI
(European Commission, 2019)

Ethics requirements for trustworthy AI	Description
1. **Accountability**	AI mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes.
2. **Transparency**	AI systems and business models should be transparent, and the AI systems' decisions should be explainable to the stakeholders. People need to be informed about the systems' capabilities and limitations.
3. **Privacy and data governance**	Ensure privacy and data protection, which requires adequate data governance framework.
4. **Diversity, non-discrimination, and fairness**	AI systems should be accessible to all, and unfair biases should be avoided. Minimizing algorithmic bias is also important.
5. **Technical robustness and safety**	AI systems should be safe, reliable, and reproducible to minimize unintended harm.
6. **Human agency and oversight**	AI systems should allow people to make informed decisions. There should be a human oversight mechanism through a "human-in-the-loop" approach.
7. **Societal and environmental well-being**	AI systems should benefit human beings and they should take into account the social impact and environmental consequences.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines.

Ethical Requirements & Governance Framework

The Framework of Multilevel Governance and Trust in AI

Competence
- Accountability
- Safety
- Robustness
- Explainability

Governments

Corporations

Citizens

Integrity
- Fairness
- Privacy
- Transparency
- Non-discrimination

Benevolence
- Human autonomy
- Social justice
- Social well-being
- Environmental well-being

Choung, H., David, P., & Ross, A. (2022). Trust and ethics in AI.
Choung, H., David, P., & Ross, A. (2022). Trust in AI and its role in the acceptance of AI technologies.

From Principled Approach to Virtue-based Approach

Principled Approach	**Virtue-based Approach**
Based on deontology – focuses on principles	Treat characters as fundamental to ethics
Focus on "doing" (act-centered)	Focus on "being" (agent-centered)
Development of action-guiding principles	Development of specific positive dispositions
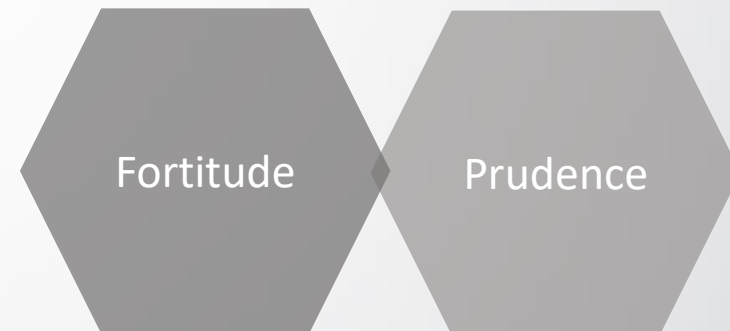
Hagendorff, T. (2022). A virtue-based framework to support putting AI ethics into practice.

Basic AI Virtues

Honesty

Responsibility

Care

Justice

Second-order AI Virtues

Fortitude

Prudence

COM ART SCI

Thank you!

Prabu David	Arun Ross	Hyesun Choung
pdavid@msu.edu	rossarun@cse.msu.edu	choungh@msu.edu