



Trustworthy AI: Ethics

Destiny Pounds

M.S. Student, Biomedical Data Science

Advised by Vibhuti Gupta, Ph.D.

Assistant Professor, Computer Science and Data Science

School of Applied Computational Sciences

Meharry Medical College

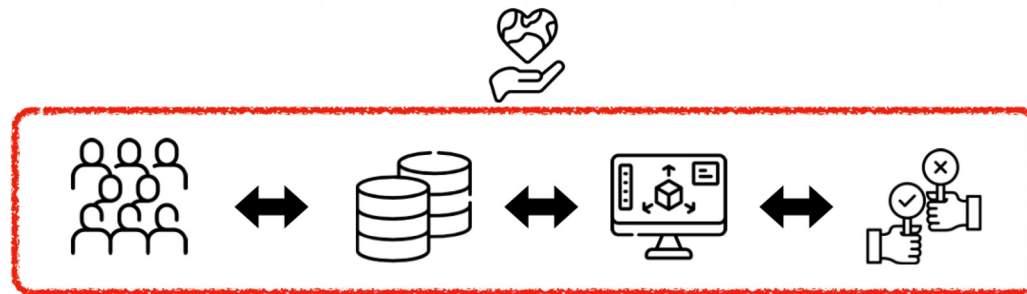


Overview

- What is Ethics?
- Establishing Guidelines
- Ethical Principles
 - Transparency
 - Justice & Fairness
 - Non-maleficence
 - Responsibility
 - Privacy

Ethics

AI ethics is a set of guidelines that advise on the design and outcomes of artificial intelligence.





Ethical guidelines vary from country to country and across different industries.

Table 2 | Ethics guidelines for AI by country of issuer (USA, international, EU and N/A)

Name of document/website	Issuer	Country of issuer
Unified Ethical Frame for Big Data Analysis. IAF Big Data Ethics Initiative, Part A	The Information Accountability Foundation	USA
The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term	AI Now Institute	USA
Statement on Algorithmic Transparency and Accountability	Association for Computing Machinery (ACM)	USA
AI Principles	Future of Life Institute	USA
AI—Our Approach	Microsoft	USA
Artificial Intelligence. The Public Policy Opportunity	Intel Corporation	USA
IBM's Principles for Trust and Transparency	IBM	USA
OpenAI Charter	OpenAI	USA
Our Principles	Google	USA
Policy Recommendations on Augmented Intelligence in Health Care H-480.940	American Medical Association (AMA)	USA
Everyday Ethics for Artificial Intelligence. A Practical Guide for Designers and Developers	IBM	USA
Governing Artificial Intelligence. Upholding Human Rights & Dignity	Data & Society	USA
Intel's AI Privacy Policy White Paper. Protecting Individuals' Privacy and Data in the Artificial Intelligence World	Intel Corporation	USA
Introducing Unity's Guiding Principles for Ethical AI—Unity Blog	Unity Technologies	USA
Digital Decisions	Center for Democracy & Technology	USA
Science, Law and Society (SLS) Initiative	The Future Society	USA
AI Now 2018 Report	AI Now Institute	USA
Responsible Bots: 10 Guidelines for Developers of Conversational AI	Microsoft	USA
Preparing for the Future of Artificial Intelligence	Executive Office of the President; National Science and Technology Council; Committee on Technology	USA
The National Artificial Intelligence Research and Development Strategic Plan	National Science and Technology Council; Networking and Information Technology Research and Development Subcommittee	USA
AI Now 2017 Report	AI Now Institute	USA

Overarching Ethical Principles



Transparency



Justice and Fairness



Non-maleficence



Responsibility



Privacy



Transparency

AI should be interpretable for users.

Explainability describes the extent to which human users can comprehend AI systems and trust the outcomes.



Justice and Fairness

AI should avoid unwanted bias and discrimination.

Algorithmic discrimination indicates the presence of biases.
Fairness refers to the attempts used to correct algorithmic biases.



Non - maleficence

AI should never cause harm.

AI safety identifies causes of unintended behavior and potential harm and develops tools to ensure systems work safely.



Responsibility

AI should make reliable decisions that are accountable.

Accountability refers to the ability to ascertain whether an AI system is behaving properly.



Privacy

The data obtained by AI should be secure and protected.

Privacy-preserving machine learning utilizes techniques to safeguard data to help prevent the exposure of sensitive data.



Thank You

Please send us your questions at:

vgupta@mmc.edu and

dpounds24@email.mmc.edu