



A Very Brief Introduction to Ethics in AI



Artificial Intelligence is the study of “rational agents”: any system that perceives its environment and takes actions that maximize its chance of achieving its goals.

- Russell & Norvig (2003)



...
distinguish
ethics. noun
moral prin
conductin
... and

Dec 2020 EO 13960: Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government

Lawful & respectful of Nation's values	Purposeful & performance- driven	Accurate, reliable, & effective
Safe, secure, & resilient	Understandable	Responsible & traceable
Regularly monitored	Transparent	Accountable

Oct 2022 OSTP: Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People



Safe and
Effective Systems



Algorithmic
Discrimination Protections



Data Privacy



Notice and
Explanation



Human Alternatives,
Consideration, and Fallback

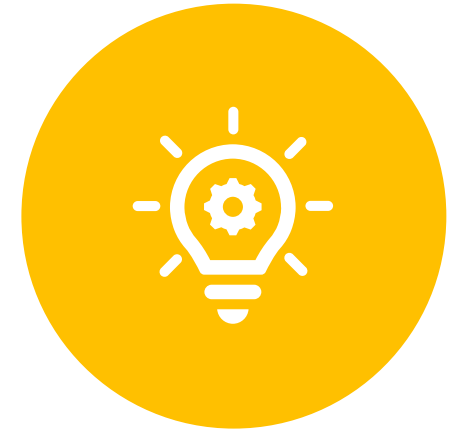
Oct 2023 EO: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.



Provides guidance to specific governmental agencies about roles and responsibilities.

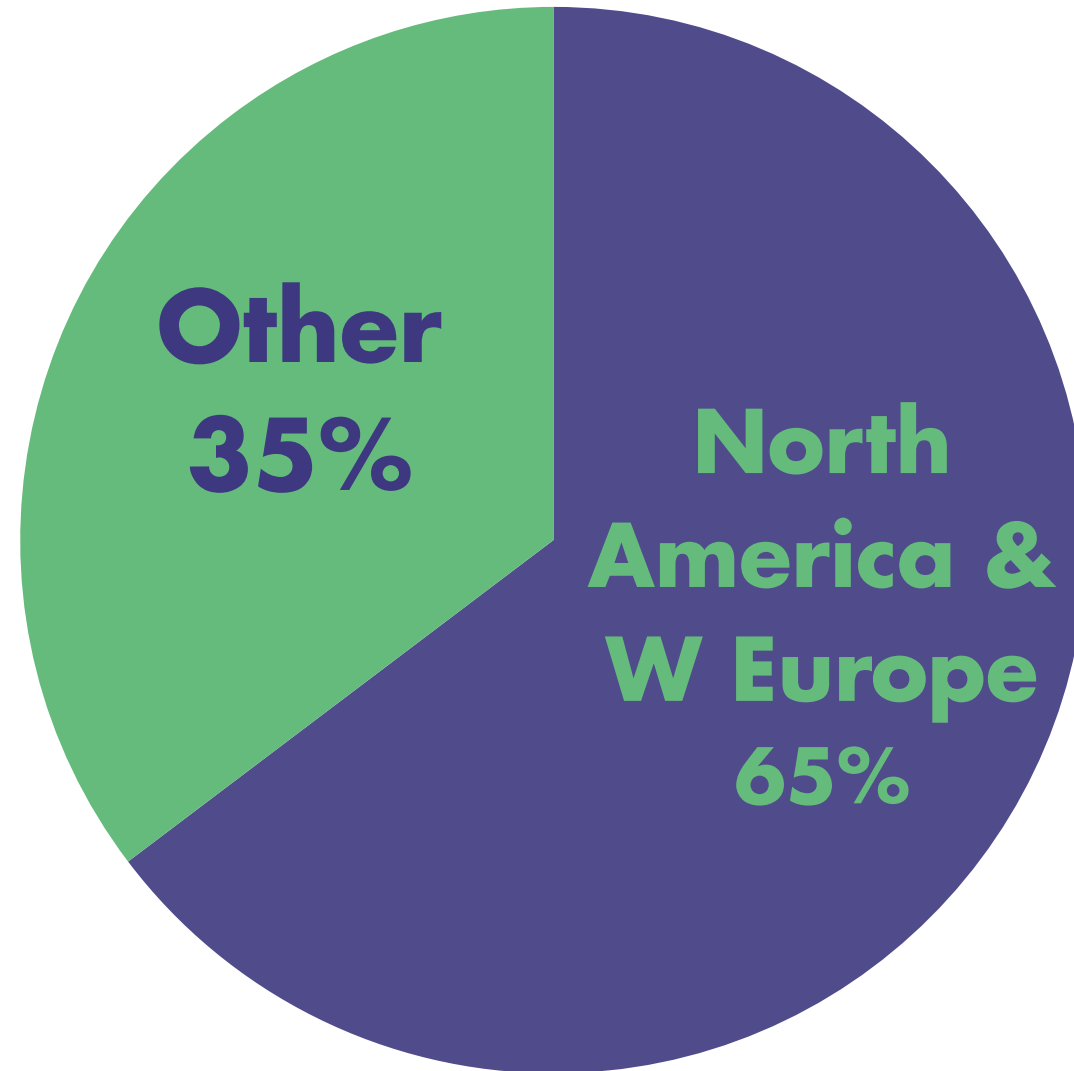


Sets timelines for the creation of governing bodies and reports on AI technology progress and safety measures.



Reaffirms that the U.S. should remain a leader in both innovating and utilizing AI technologies.

ImageNet Country of Origin





Ground truth: Soap

Nepal, 288 \$/month

Azure: food, cheese, bread, cake, sandwich

Clarifai: food, wood, cooking, delicious, healthy

Google: food, dish, cuisine, comfort food, spam

Amazon: food, confectionary, sweets, burger

Watson: food, food product, turmeric, seasoning

Tencent: food, dish, matter, fast food, nutriment



A person is seen from the side, looking at a laptop screen. The screen displays a video call with a smiling woman and a code editor. The text 'TOWARDS INCODING' is in white and 'MAKING CHANGE' is in red.

TOWARDS INCODING MAKING CHANGE

- Identifying Bias
- Curating Inclusively
- Developing Conscientiously



THE SIX PRINCIPLES



.1 Transparency

AI systems must be understandable to all_



.2 Inclusion

These systems must not discriminate
against anyone because every human being
has equal dignity_



.3 Accountability

there must always be someone who takes
responsibility for what a machine does_



.4 Impartiality

AI systems must not follow or create biases_



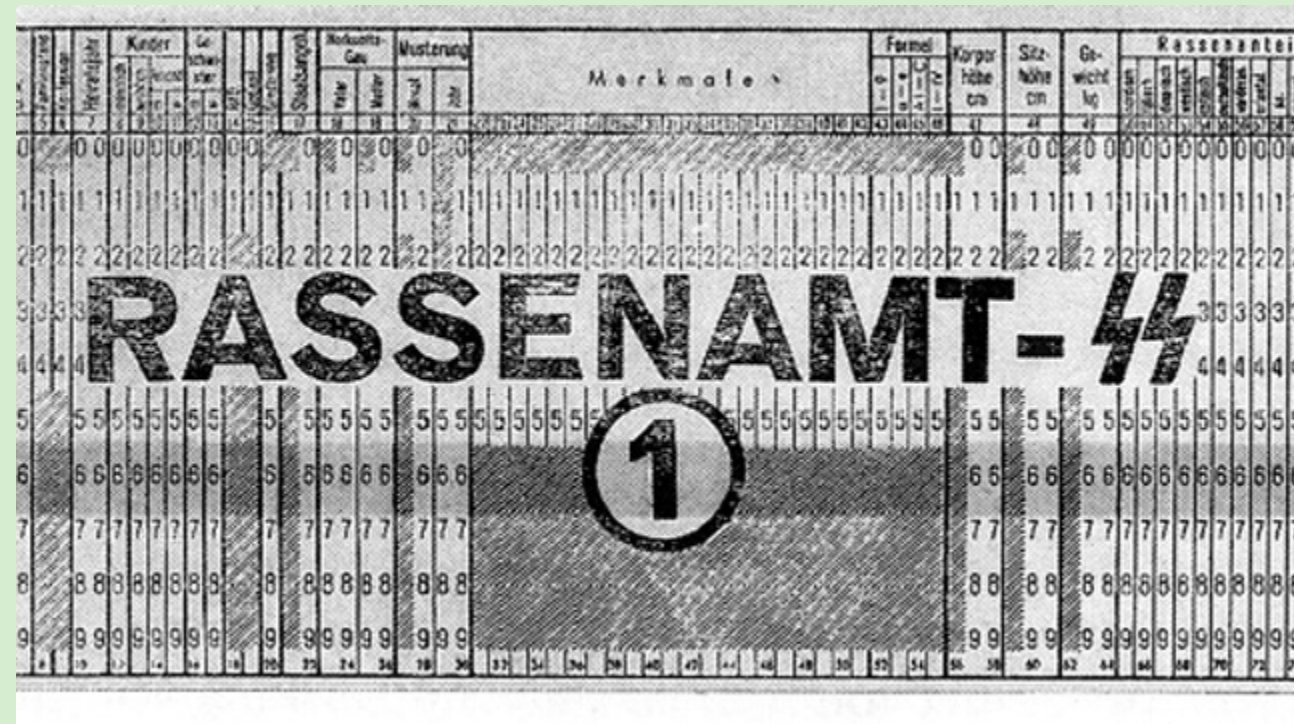
.5 Reliability

AI must be reliable_



.6 Security and Privacy

These systems must be secure and respect
the privacy of users_







The ethical dilemma of self-driving cars - Patrick Lin



Watch later



THE ETHICAL DILEMMA OF SELF-DRIVING CARS



