ARTIFICIAL INTELLIGENCE.
GSR-18 LEADERSHIP DEBATE.

DR. ANASTASSIA LAUTERBACH
GENEVA, JULY 10TH 2018
FIVE OF TOP 10 FORTUNE 500 COMPANIES ARE ‘AI-FIRST’.

Source: Fortune 500, Research 1AU-Ventures
MACHINE LEARNING WILL GET EMBEDDED INTO EVERY APPLICATION, PROCESS AND BUSINESS.

"The business plans of the next 10,000 startups are easy to forecast: Take X and add AI"

Kevin Kelly, Founder ‘Wired’
THE DEFINITIONS, OBJECTIVES AND METHODS OF “ARTIFICIAL INTELLIGENCE” CONTINUE TO EVOLVE.

- Mimic human behavior
- Mimic thinking
- Project human intent
- Advise humans
- Behave intelligently
- Behave rationally
- Behave empathetically

... explicitly model the human brain

Network model of ‘target’ and ‘source’ nodes via anatomical tracing

... or loosely approximate the mental models

Source: Research 1AU-Ventures
# KNOWN TECHNOLOGY-BASED RISKS OF AI.

## Design Mistakes
- Bias in technology reflecting its creators, lack of diversity in gender, age, ethnicities and lack of interdisciplinarities
- Potential rise in inequality in Legal, Social Services, Education

## Malicious Intent
- Unethical behaviors of system makers to exploit users for financial gain or cover up for bigger issues of their technology
- Impact on reputational risk
- New forms of Cyber-hacking

## No Human in the Loop
- Systems becoming capable to reuse data and experiences from other tasks when trying solving new tasks

- Google Translate attributes ‘he’ to soldiers, doctors and entrepreneurs, and ‘she’ to nurses and teachers while translating from Turkish (language with a single gender-neutral pronoun ‘o’);
- Alexa misunderstanding accents;
- FAIR Lab matching Obama image to ‘basketball’

- New ethical standards, e.g. BS 8611 Guide to the Ethical Design and Application of Robots and Robotic Systems, or IEEE P700X ‘human standards’ to support OEMs and ODMs in the ethical application of robots
- Adversarial attacks on data models, e.g. hacking into a fleet of self-driving cars, faking video images

- Future coding will be dominated by automated systems without significant human involvement.

Source: Research 1AU-Ventures
MAJOR THEMES IN A.I. ETHICS FRAMEWORK.

Goal Alignment (between smart & autonomous agents and humans)
Incentive Design for Autonomous Systems

Governance on A.I. Safety
Decision-Making Guidelines for Autonomous Systems

A.I. Ethics

© Anastassia Lauterbach 2018. All rights reserved
ETHICAL NORMS VS. CONTEXT.

Contemporary ethical norms:
- Moral norms
- Human rights
- Equality
- Privacy
- Transparency and accountability

Context:
- Geo-Political environment
- Gender and minorities
- Social mobility

© Anastassia Lauterbach 2018. All rights reserved
SOCIAL GOVERNANCE ON AI.

- AI as a part of sustainability frameworks
- Companies
- Full stack AI companies
- Ethic Advisory Boards
- Municipalities
- Schools
- Transparency over online resources to learn about AI
- Open hacktons with access to data and participance of local companies, startups and individual innovators

Source: Research 1AU-Ventures
"People can't understand new ideas if their livelihood depends on the old ones."

Upton Sinclair