ARTIFICIAL INTELLIGENCE.
GSR-18 LEADERSHIP DEBATE.

DR. ANASTASSIA LAUTERBACH
GENEVA, JULY 10TH 2018

FIVE OF TOP 10 FORTUNE 500 COMPANIES ARE 'AI-FIRST'.

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
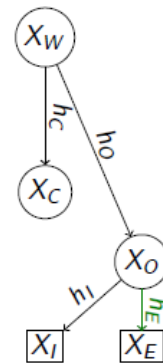- Behave empathetically

… explicitly model the human brain
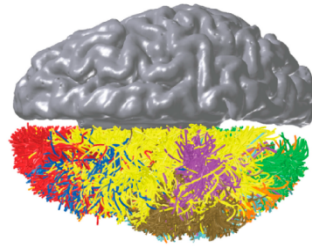
Architecture of the "projectome"
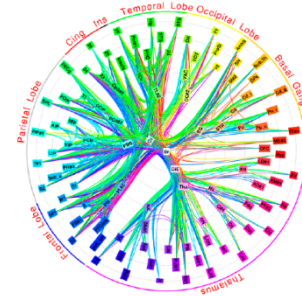
Network model of 'target' and 'source' *nodes* via anatomical tracing

… or loosely approximate the mental models

Multistate Markovian Survival Model

Bayes' Network of perfume characteristics

Source: Research 1AU-Ventures

1AU VENTURES

KNOWN TECHNOLOGY-BASED RISKS OF AI.

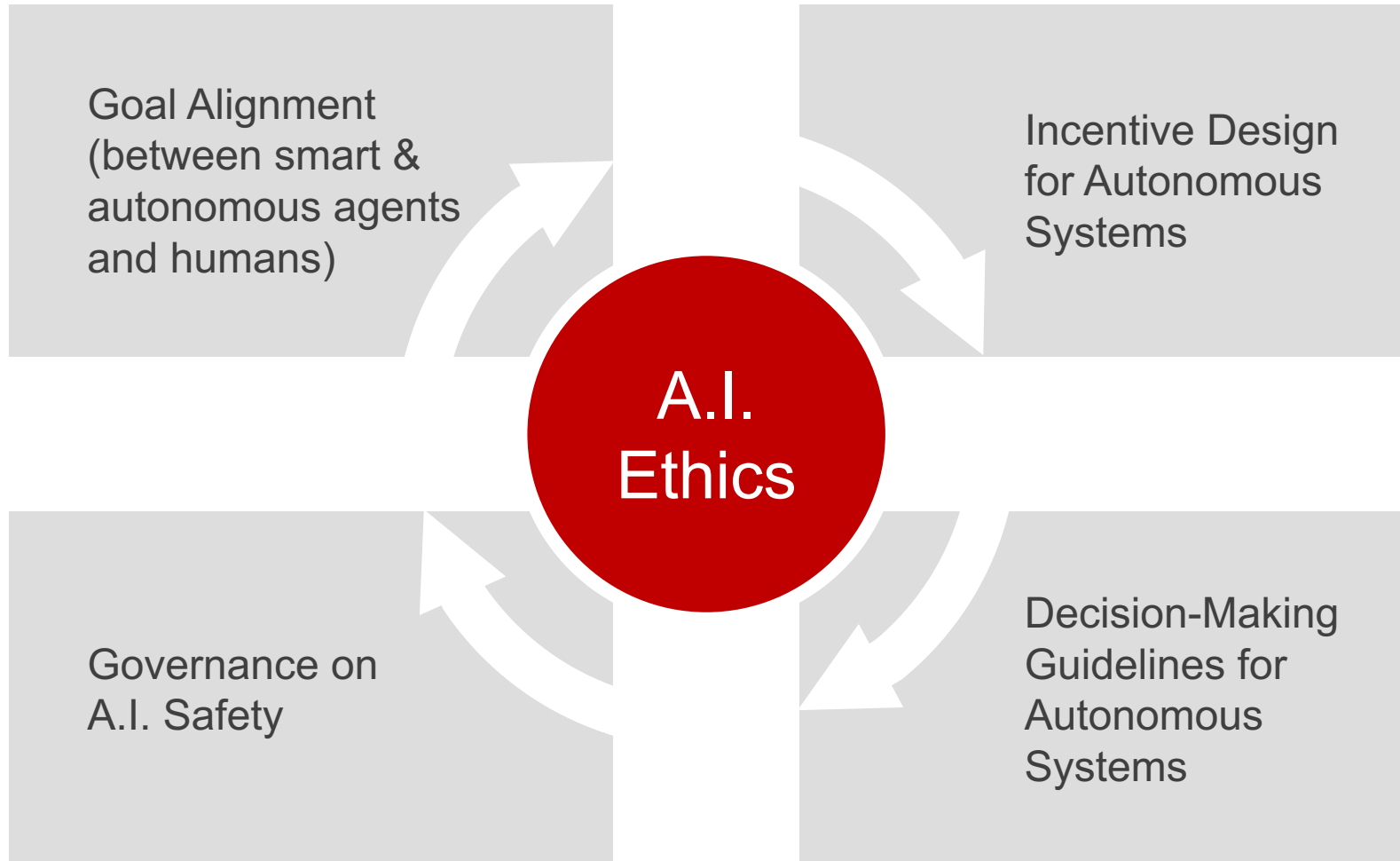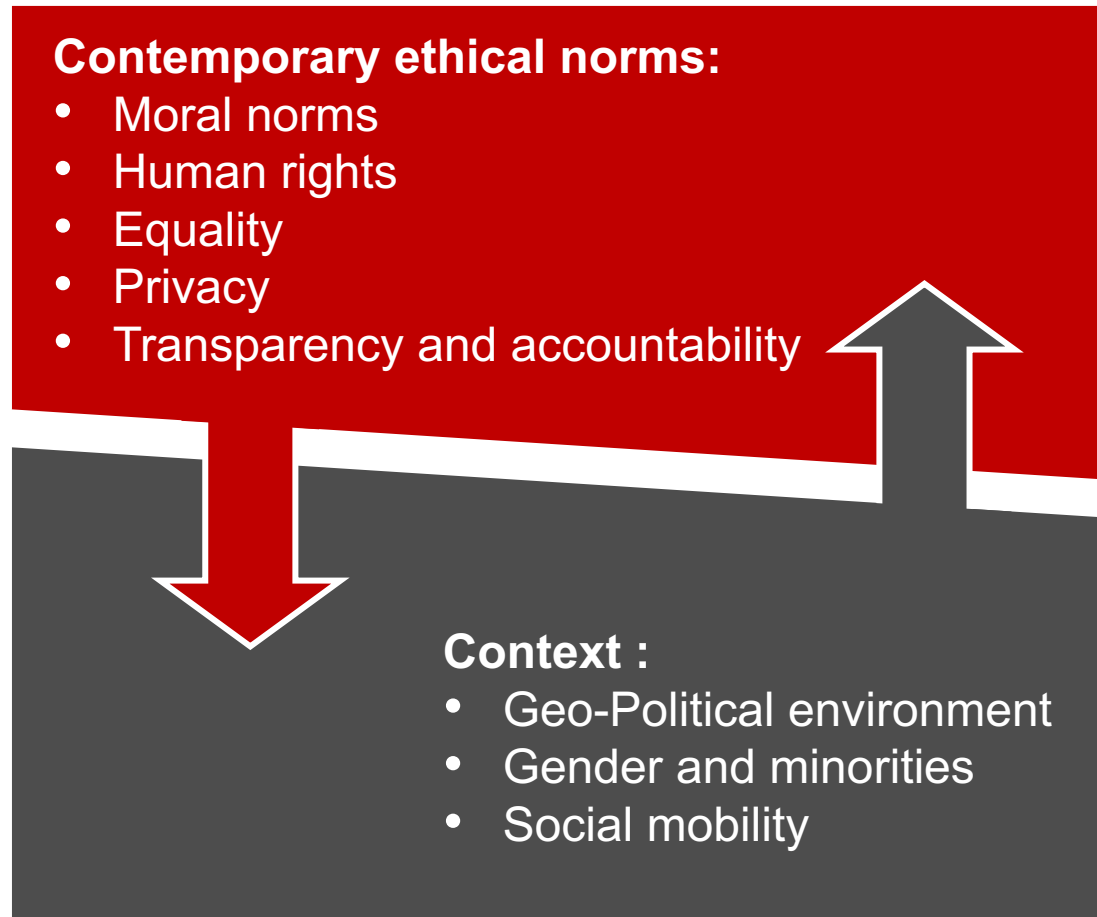| Design Mistakes	Malicious Intent	No Human in the Loop
• Bias in technology reflecting its creators, lack of diversity in gender, age, ethnics and lack of interdisciplinarities • Potential rise in inequality in Legal, Social Services, Education	• Unethical behaviors of system makers to exploit users for financial gain or cover up for bigger issues of their technology • Impact on reputational risk • New forms of Cyber-hacking	• Systems becoming capable to reuse data and experiences from other tasks when trying solving new tasks
• Google Translate attributes 'he' to soldiers, doctors and enterpreneurs, 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.

1AU VENTURES

MAJOR THEMES IN A.I. ETHICS FRAMEWORK.

Goal Alignment (between smart & autonomous agents and humans)

Incentive Design for Autonomous Systems

A.I. Ethics

Governance on A.I. Safety

Decision-Making Guidelines for Autonomous Systems

Source: Research 1AU-Ventures

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

SOCIAL GOVERNANCE ON AI.

Ethic Advisory Boards

Full stack AI companies

AI as a part of sustainability frameworks

Companies

Schools

Transparency over online resources to learn about AI

Municipalities

Open hacktons with access to data and participance of local companies, startups and individual innovators

1AU VENTURES
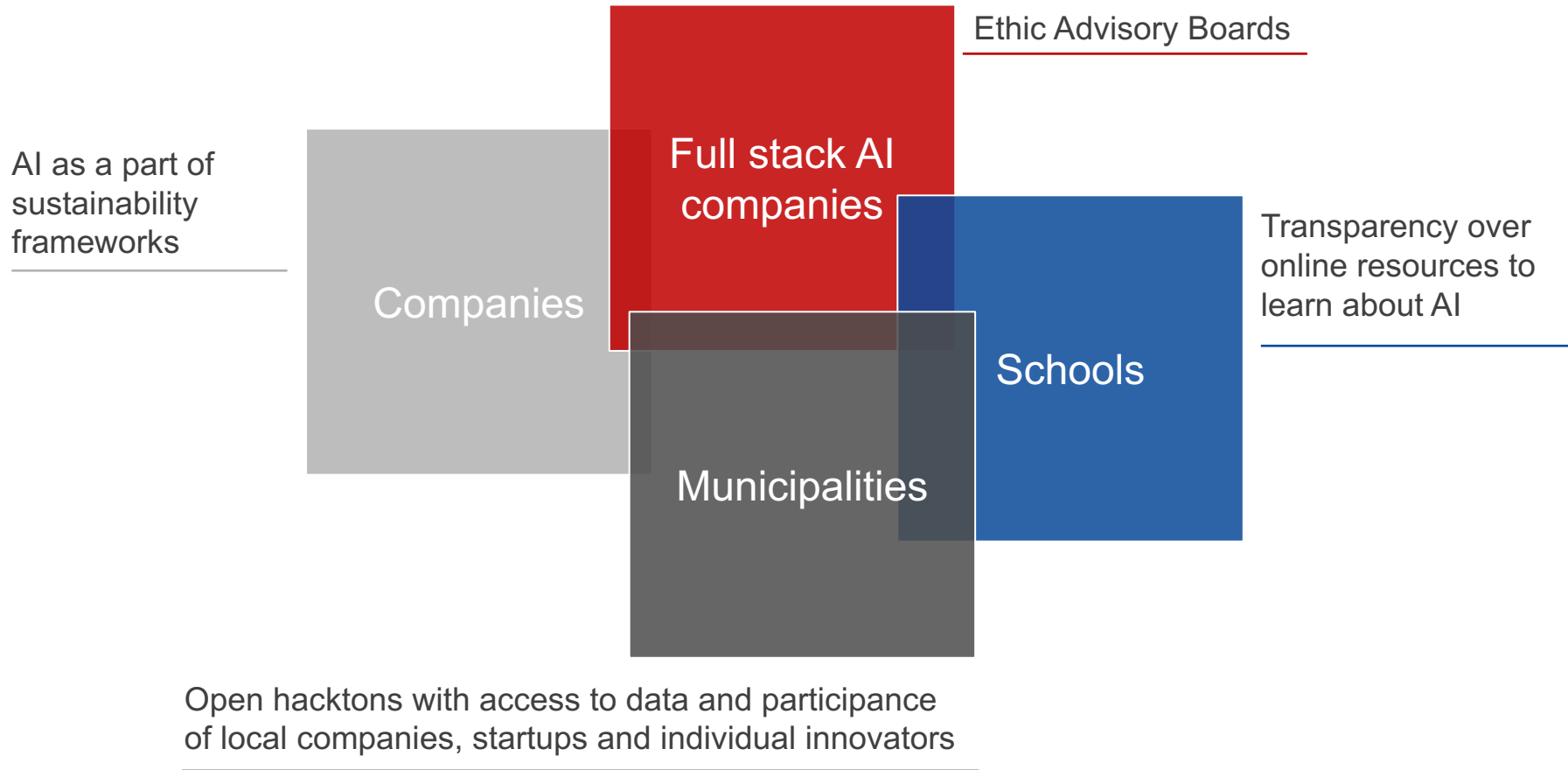
"People can't understand new ideas if their livehood depends on the old ones."

Upton Sinclair