Societal, ethical and trustworthiness issues

Presented by: Andrea Renda   For: ETSI Summit on Artificial Intelligence
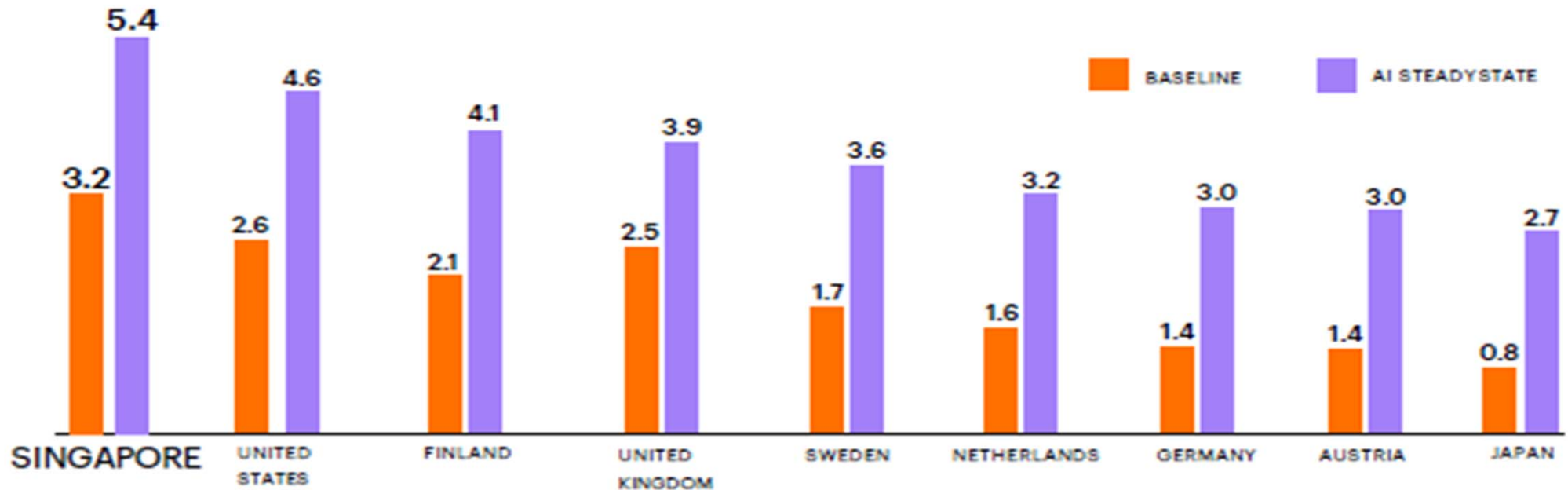
04.04.2019

A huge potential, in need of direction and standards

There is consensus on AI's enormous potential to improve prosperity

- Several authors believe productivity will (eventually(increase
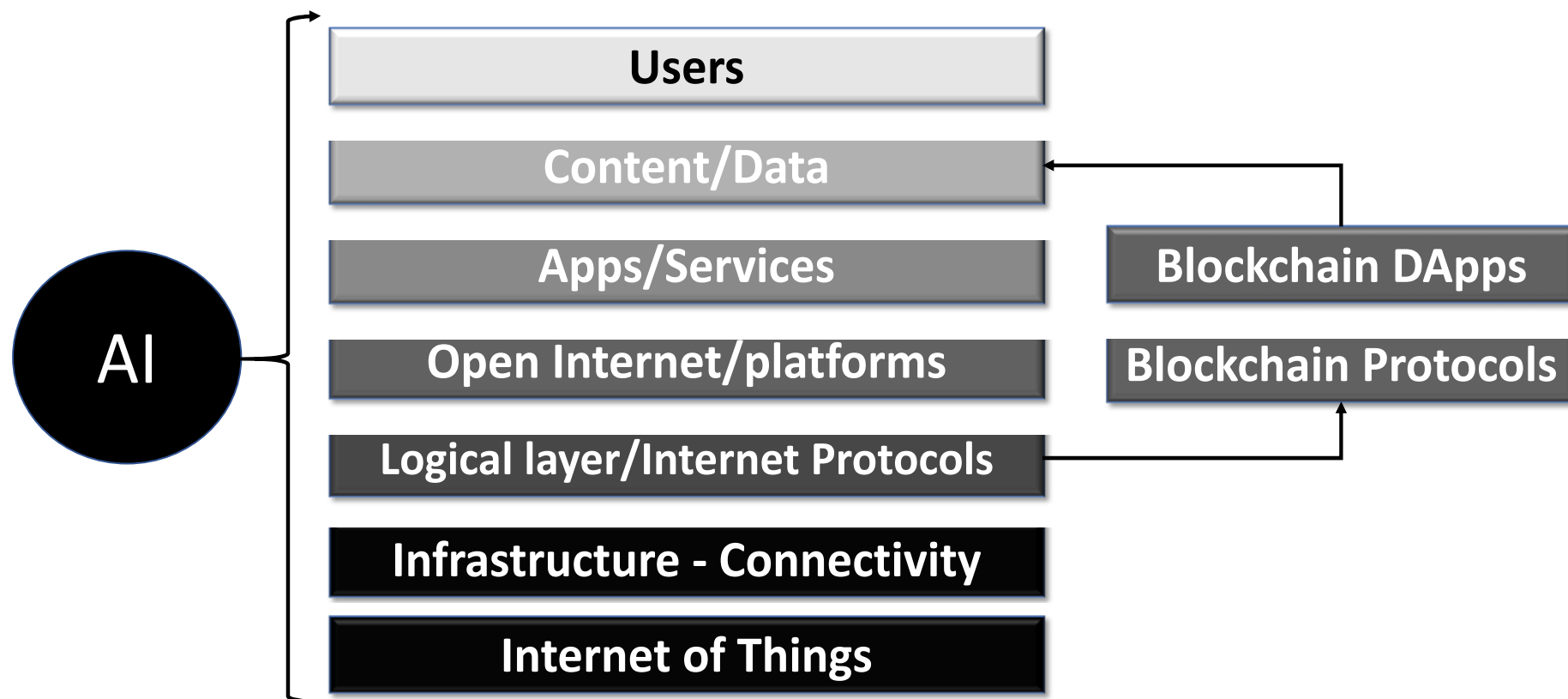- Need to look at the whole technology stack

AI CAN DOUBLE GROWTH RATES

A comparison of baseline annual gross value added growth (%) in 2035 to a scenario where AI has been absorbed.

	BASELINE	AI STEADYSTATE
SINGAPORE	3.2	5.4
UNITED STATES	2.6	4.6
FINLAND	2.1	4.1
UNITED KINGDOM	2.5	3.9
SWEDEN	1.7	3.6
NETHERLANDS	1.6	3.2
GERMANY	1.4	3.0
AUSTRIA	1.4	3.0
JAPAN	0.8	2.7

Source: Accenture/Frontier Economics (2017)

The new «technology stack»

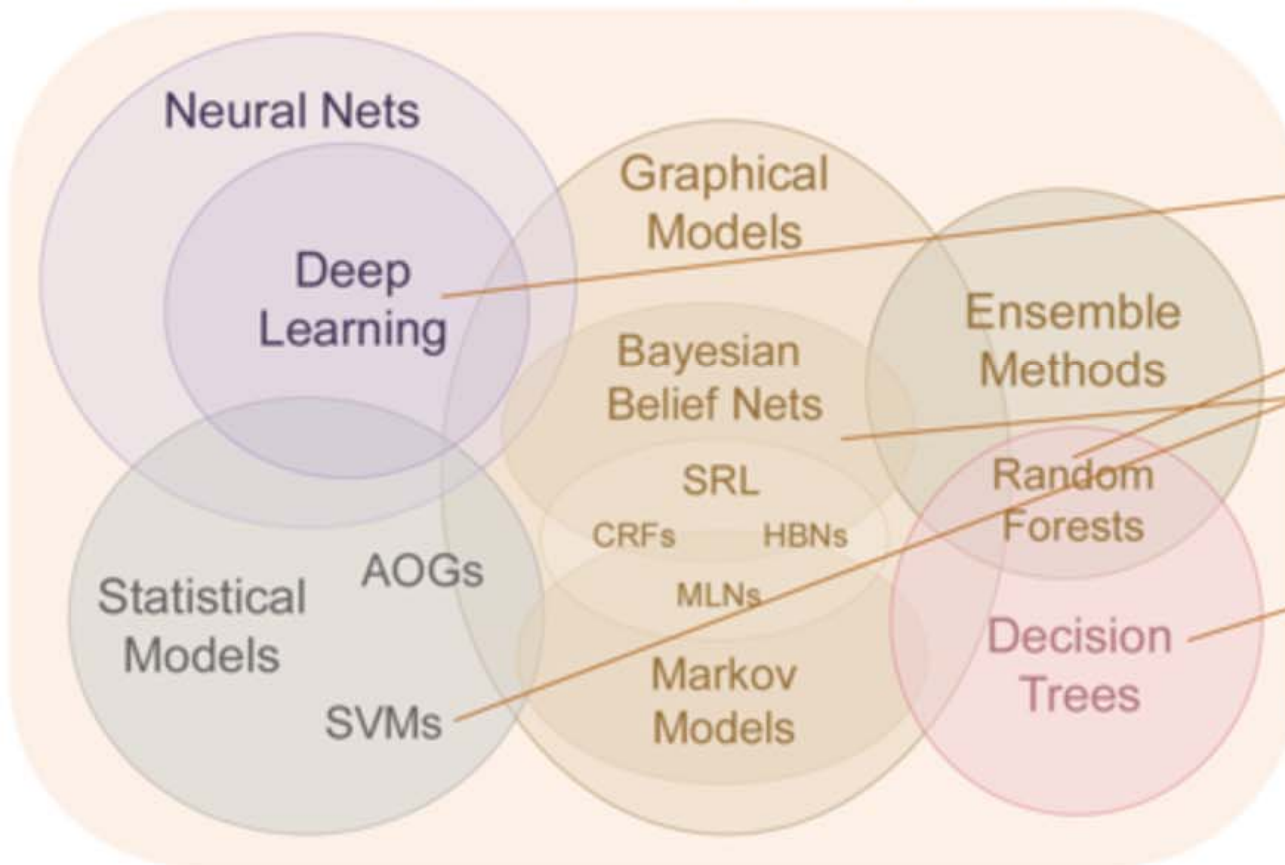# A huge potential, in need of direction and standards (2)

There is consensus on AI's enormous potential to improve prosperity

- ᛦ Several authors believe productivity will (eventually increase)
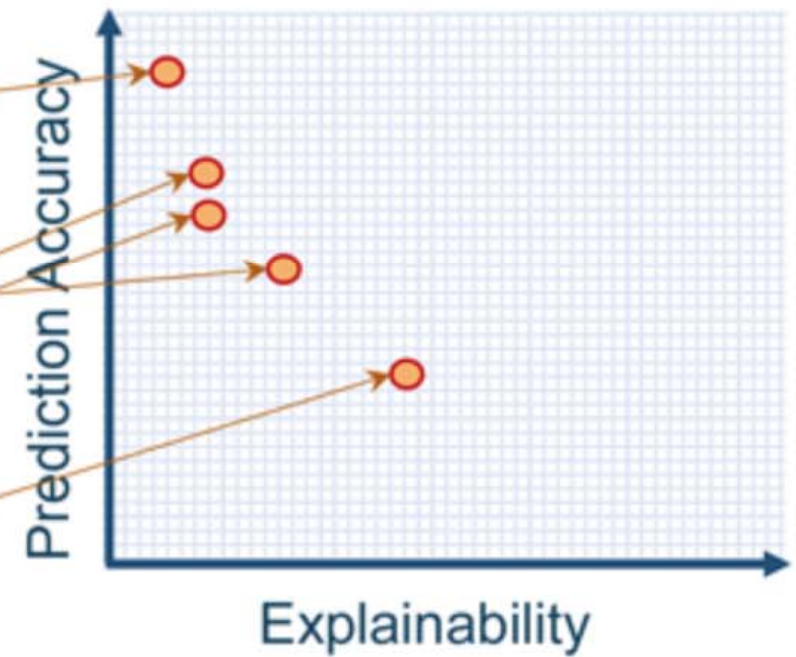- ᛦ Need to look at the whole technology stack to appraise the full potential (and risks)

Emerging trade-offs will depend on the use case

- ᛦ Accuracy/Privacy/Explainability trilemma
- ᛦ What is acceptable bias? What is acceptable nudging? What is acceptable surveillance?
- ᛦ What liability in complex human-machine interactions?
- ᛦ What data governance schemes are secure and privacy-preserving?

Learning Techniques (today)

Neural Nets

Deep Learning

Graphical Models

Ensemble Methods

Bayesian Belief Nets

SRL

CRFs   HBNs

MLNs

Statistical Models

AOGs

SVMs

Markov Models

Random Forests

Decision Trees

Explainability (notional)

Prediction Accuracy

Explainability

Source: Zelros AI
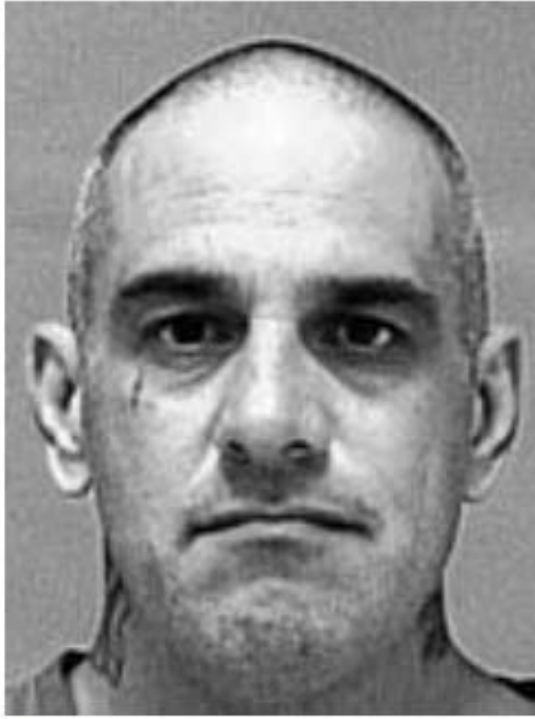
The Intersect

Google's algorithm shows prestigious job ads to men, but not to women. Here's why that should worry you.

By Julia Carpenter   July 6, 2015   ✉

A recent screenshot of Google images for "CEO."

Fresh off the revelation that Google image searches for "CEO" only turn up pictures of white men, there's new evidence that algorithmic bias is, alas, at it again. In a paper published in April, a team of researchers from Carnegie Mellon University claim Google displays far fewer ads for high-paying executive jobs...
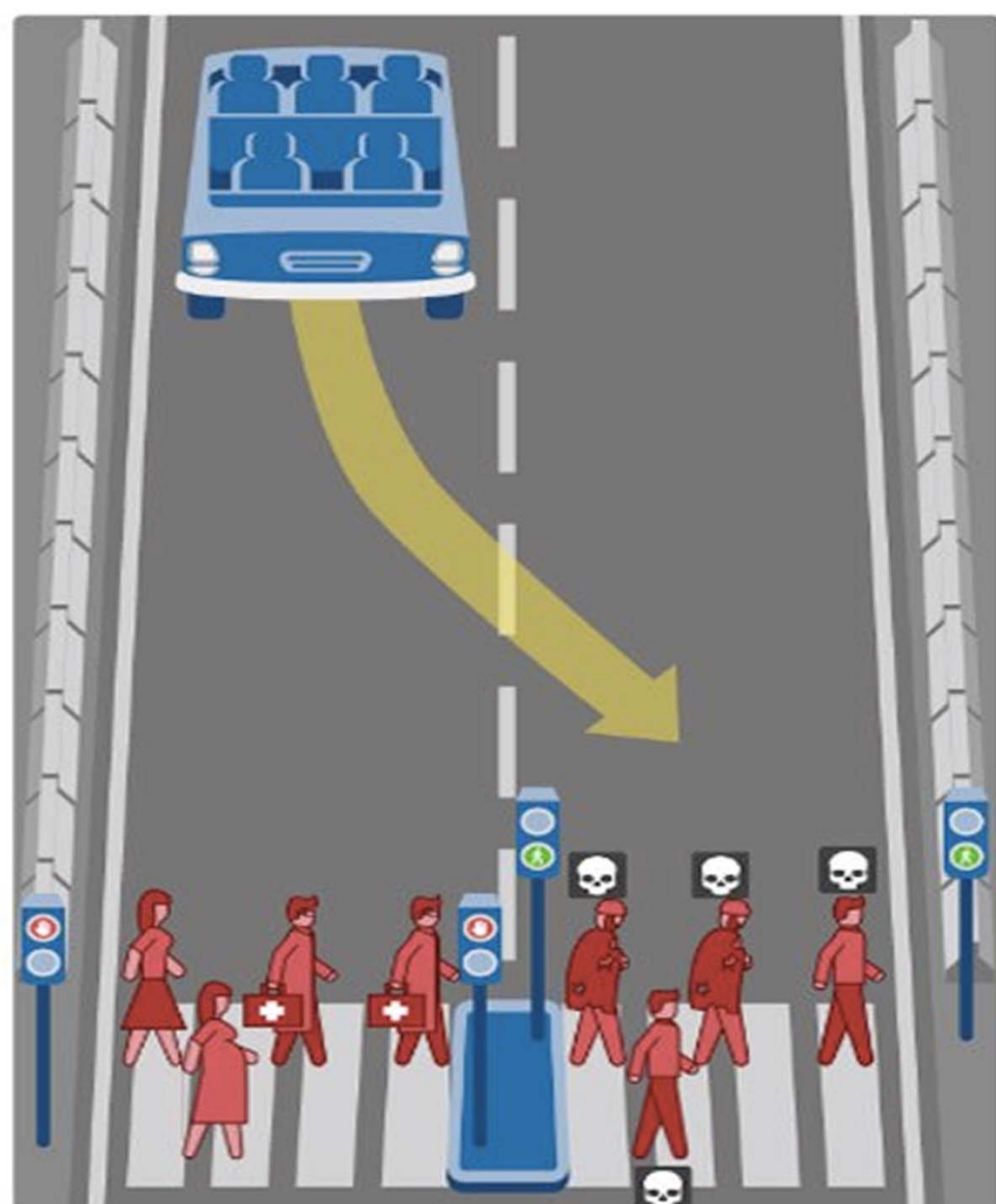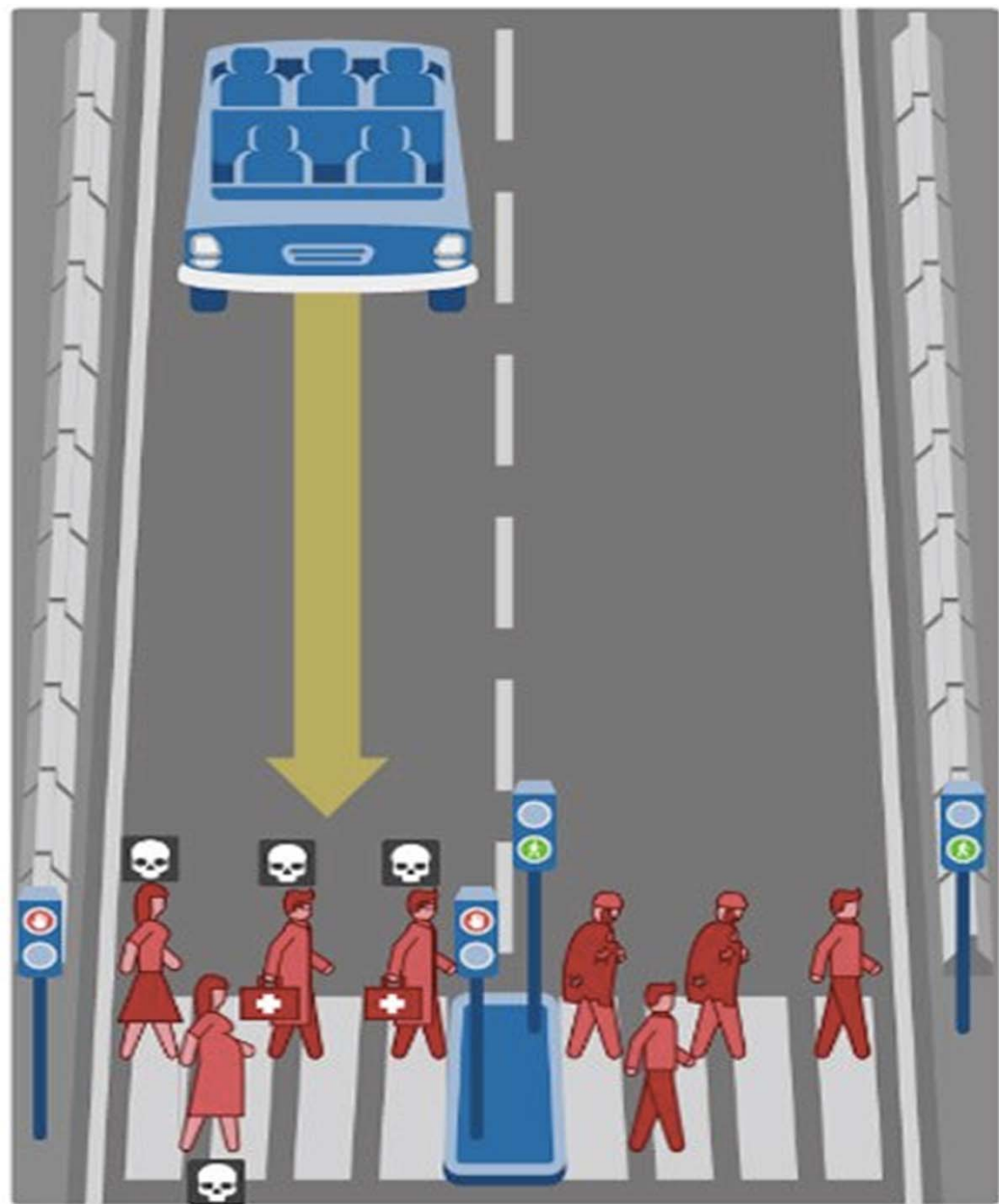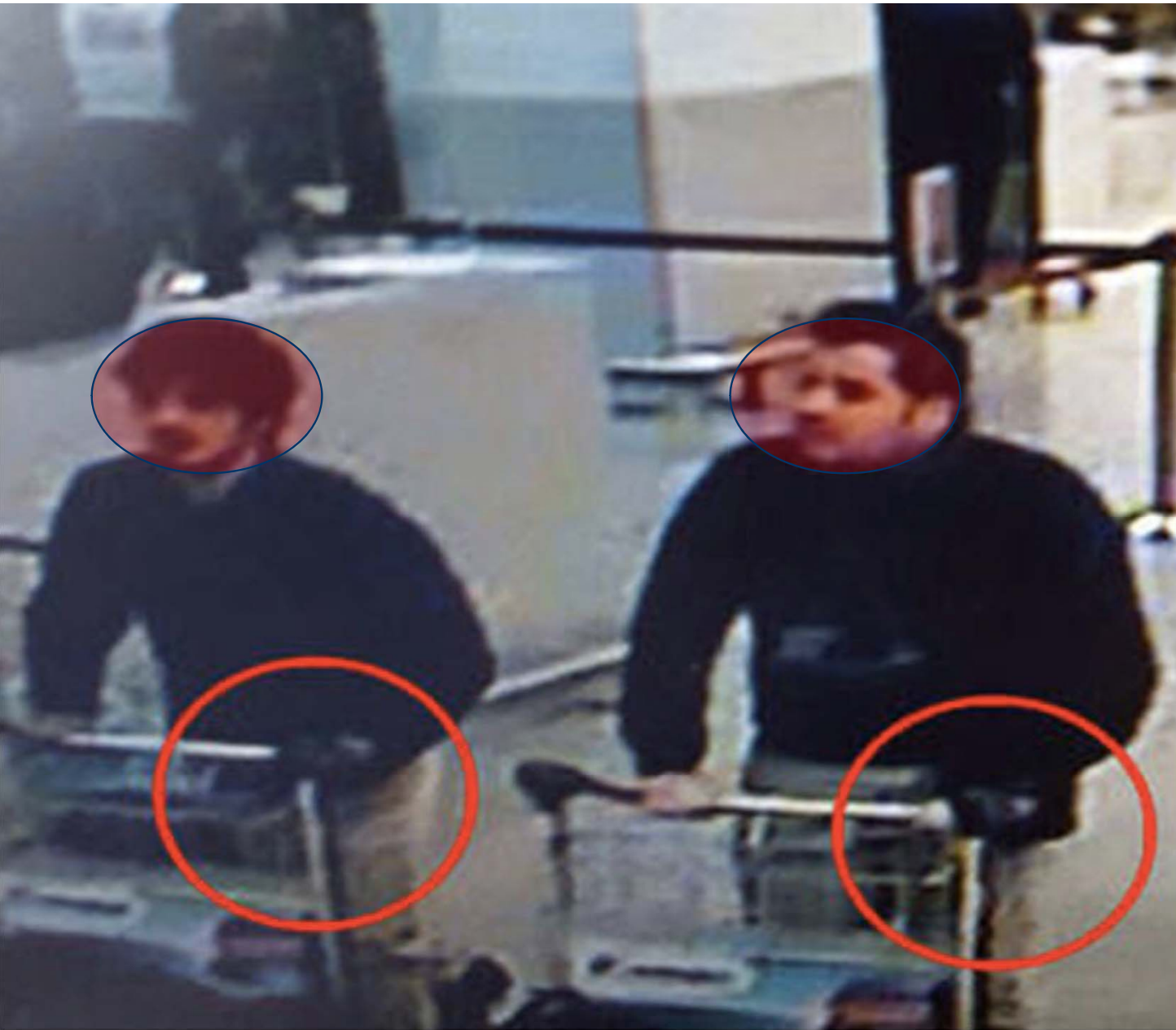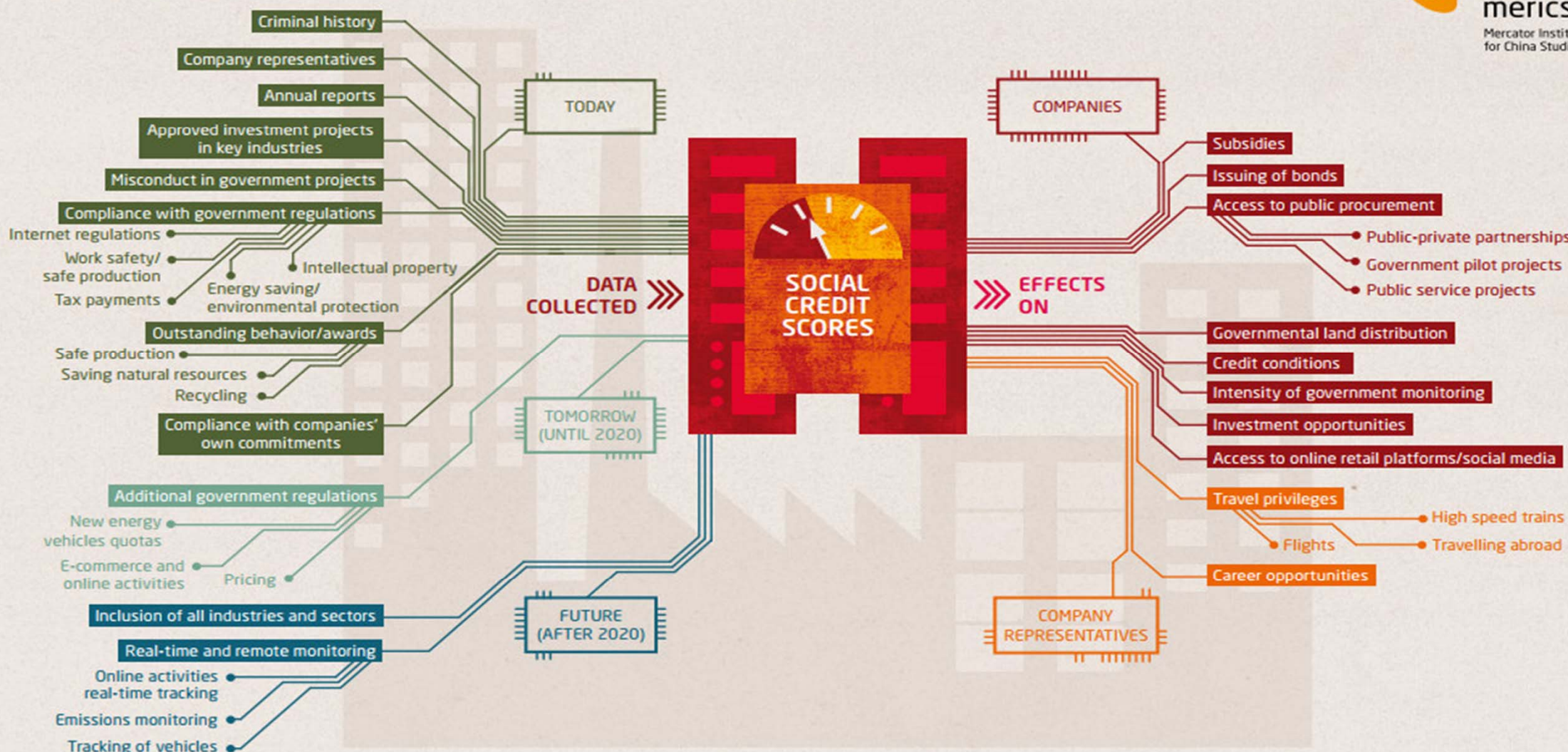
Vernon Prater
3 Low Risk

Brisha Borden
8 High Risk

China's tight grip on enterprises
Influencing business decisions via Social Credit Scores*

merics
Mercator Institut
for China Studie

DATA COLLECTED »

SOCIAL CREDIT SCORES

EFFECTS ON »

TODAY
- Criminal history
- Company representatives
- Annual reports
- Approved investment projects in key industries
- Misconduct in government projects
- Compliance with government regulations
  - Internet regulations
  - Work safety/safe production
  - Tax payments
  - Intellectual property
  - Energy saving/environmental protection
- Outstanding behavior/awards
  - Safe production
  - Saving natural resources
  - Recycling
- Compliance with companies' own commitments

TOMORROW (UNTIL 2020)
- Additional government regulations
  - New energy vehicles quotas
  - E-commerce and online activities
  - Pricing

FUTURE (AFTER 2020)
- Inclusion of all industries and sectors
- Real-time and remote monitoring
  - Online activities real-time tracking
  - Emissions monitoring
  - Tracking of vehicles

COMPANIES
- Subsidies
- Issuing of bonds
- Access to public procurement
  - Public-private partnerships
  - Government pilot projects
  - Public service projects
- Governmental land distribution
- Credit conditions
- Intensity of government monitoring
- Investment opportunities
- Access to online retail platforms/social media

COMPANY REPRESENTATIVES
- Travel privileges
  - High speed trains
  - Flights
  - Travelling abroad
- Career opportunities

*Selection of data collected and exemplary effects of Social Credit Scores.
Source: Policy documents and regulations released by the Chinese central government since 2014.

A huge potential, in need of direction and standards (2)

There is consensus on AI's enormous potential to improve prosperity

- Several authors believe productivity will (eventually increase)
- Need to look at the whole technology stack to appraise the full potential (and risks)
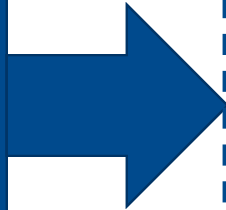
Emerging trade-offs will depend on the use case

- Accuracy/Privacy/Explainability trilemma
- What is acceptable bias? What is acceptable nudging? What is acceptable surveillance?
- What liability in complex human-machine interactions?
- What data governance schemes are secure and privacy-preserving?

Towards a new Trustworthy AI Assessment framework

- Need for piloting/testing
- Need for standards that ensure adaptive, agile governance of the system

Core Principles

Definitions and examples

TAI assessment

Respect for Human Autonomy

Prevention of Harm

Fairness

Explicability and Responsibility

Trustworthy AI

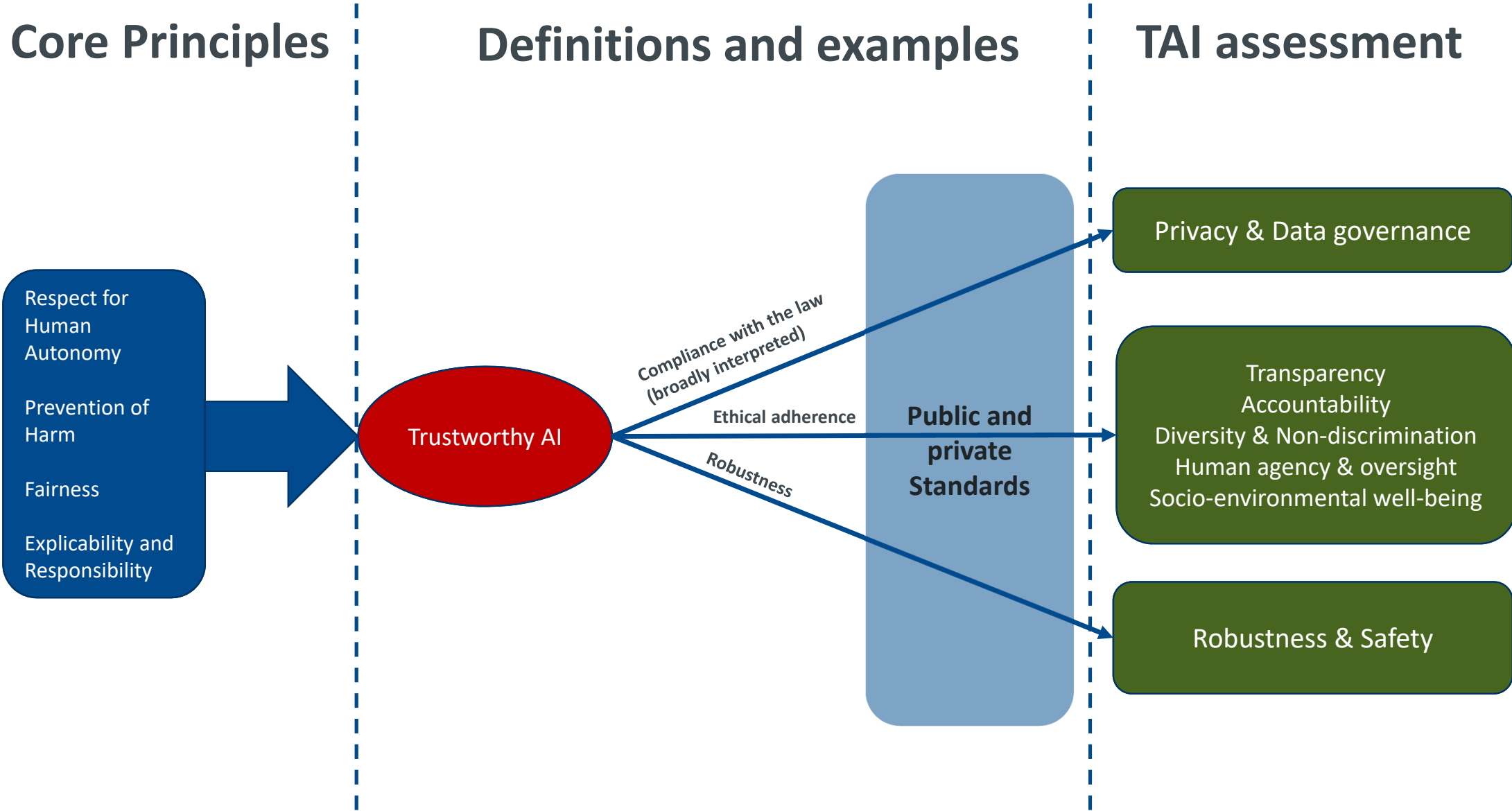Compliance with the law (broadly interpreted)

Ethical adherence

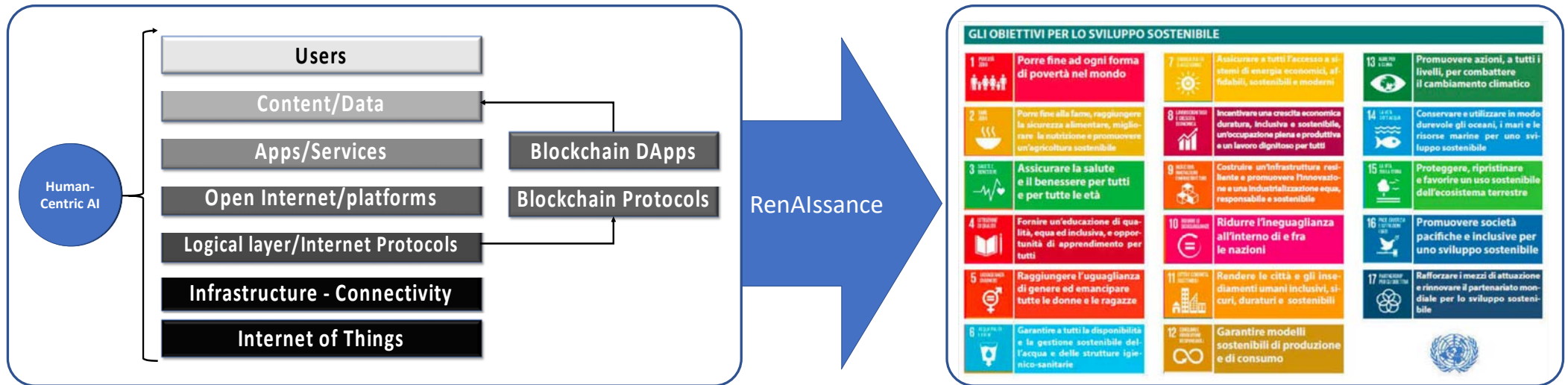Robustness

Public and private Standards

Privacy & Data governance

Transparency
Accountability
Diversity & Non-discrimination
Human agency & oversight
Socio-environmental well-being

Robustness & Safety

AI for what?

Human-Centric AI

| Users |
| Content/Data |
| Apps/Services |
| Open Internet/platforms |
| Logical layer/Internet Protocols |
| Infrastructure - Connectivity |
| Internet of Things |

Blockchain DApps
Blockchain Protocols

RenAIssance

GLI OBIETTIVI PER LO SVILUPPO SOSTENIBILE