Artificial Intelligence
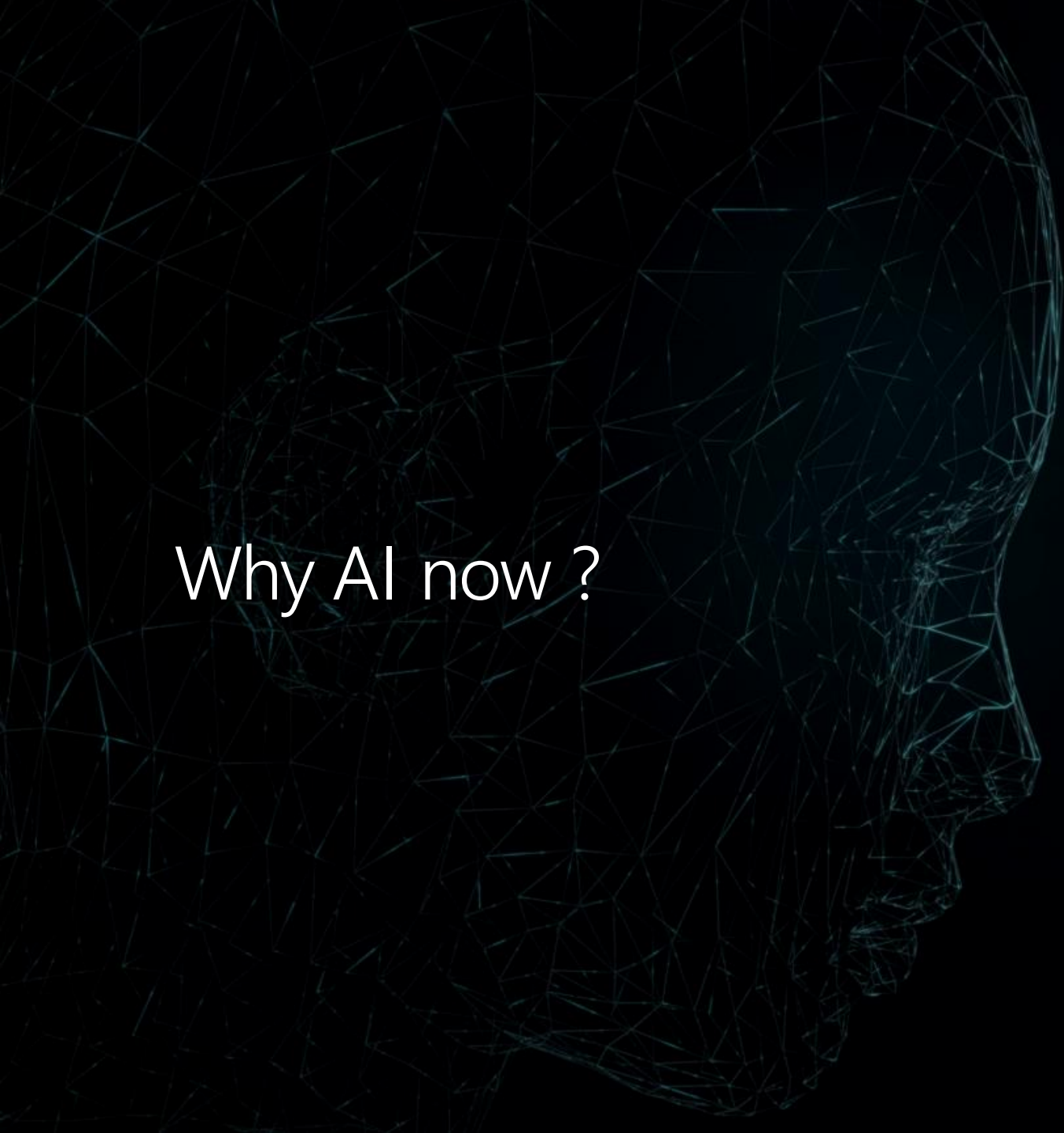## Growing reality & standardization needs

Candi Carrera
Country Manager – Microsoft Luxembourg

ILNAS Afterwork – December 12th 2019

Microsoft

Why AI now ?

The path of human progress

1784
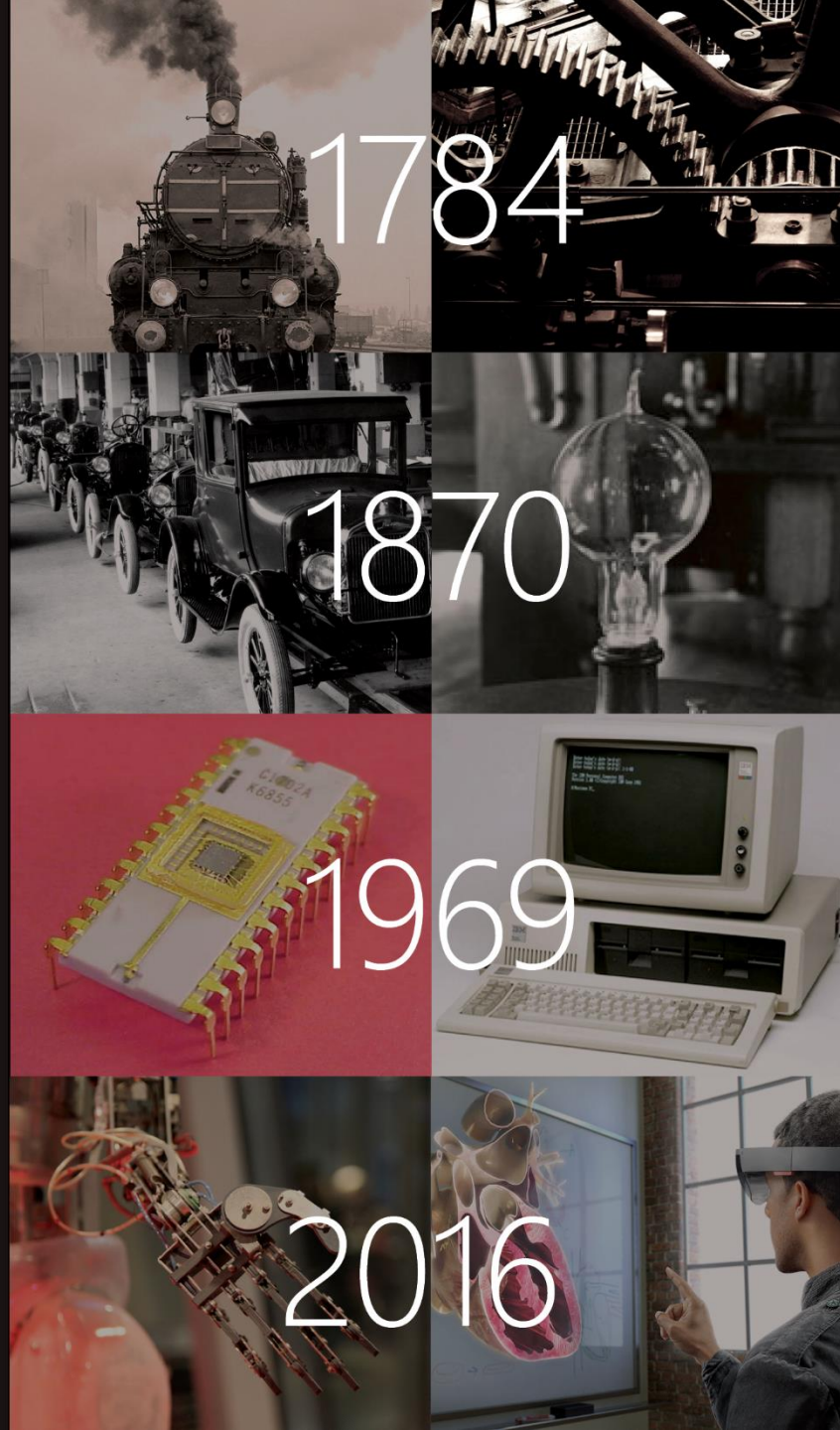
First Industrial Revolution
Water & steam, production mechanization

1870

Second Industrial Revolution
Division of labour & electric power to create mass production

1969

Third Industrial Revolution
Electronics & information technology to automate production

2016

Fourth Industrial Revolution
Fusion of technologies blurring the lines between physical, digital & biological spheres (cyber-physical systems)

Common to 4 revolutions

initiated by people to achieve certain objectives

making money
becoming famous
simply to overcome challenges
removing inefficiencies

2nd industrial revolution

1920

1930

3rd industrial revolution

+19.2m jobs
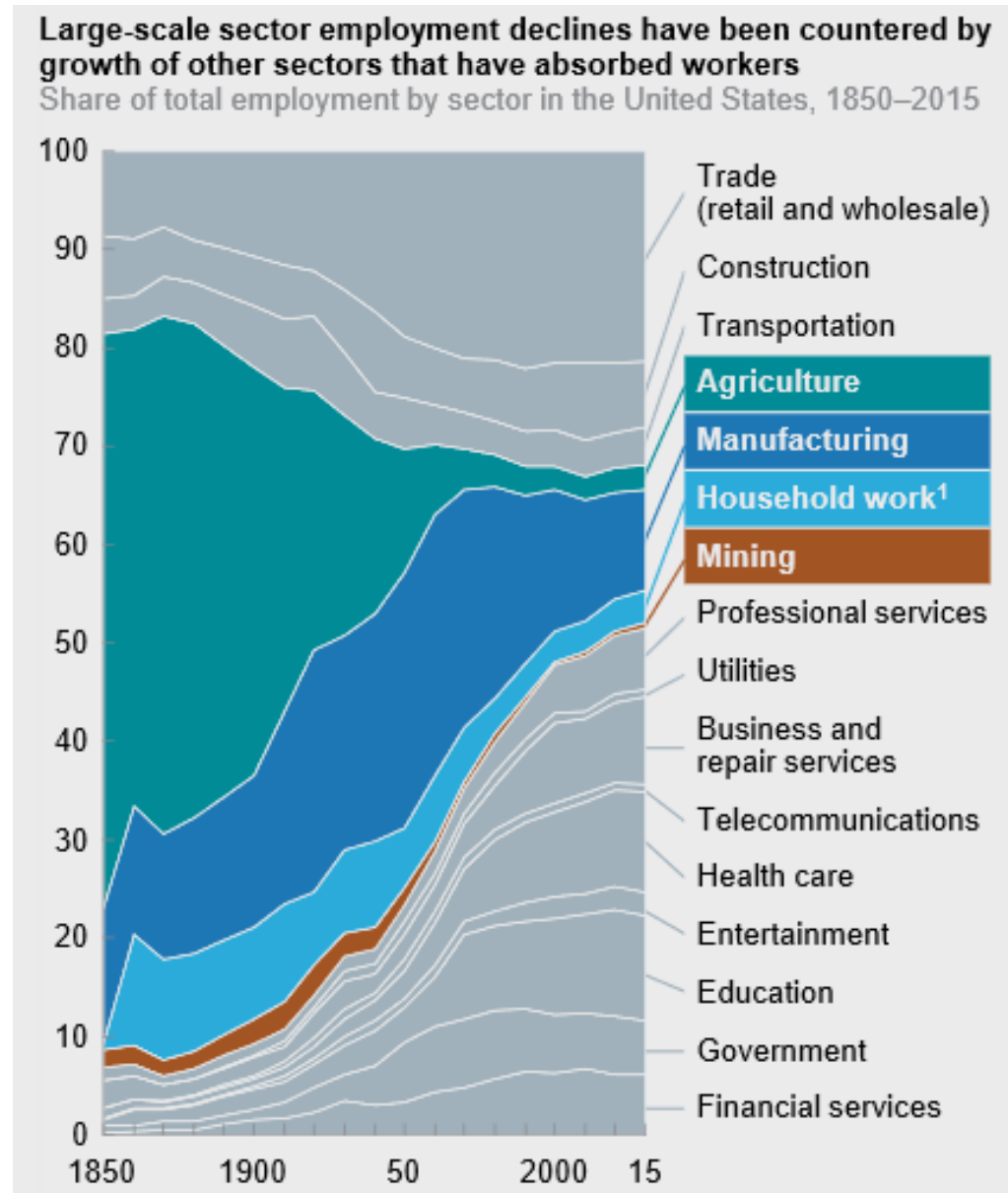
4.5x

Net +15.7m

-3.5m jobs

Global picture 1850 - 2015

Large-scale sector employment declines have been countered by growth of other sectors that have absorbed workers

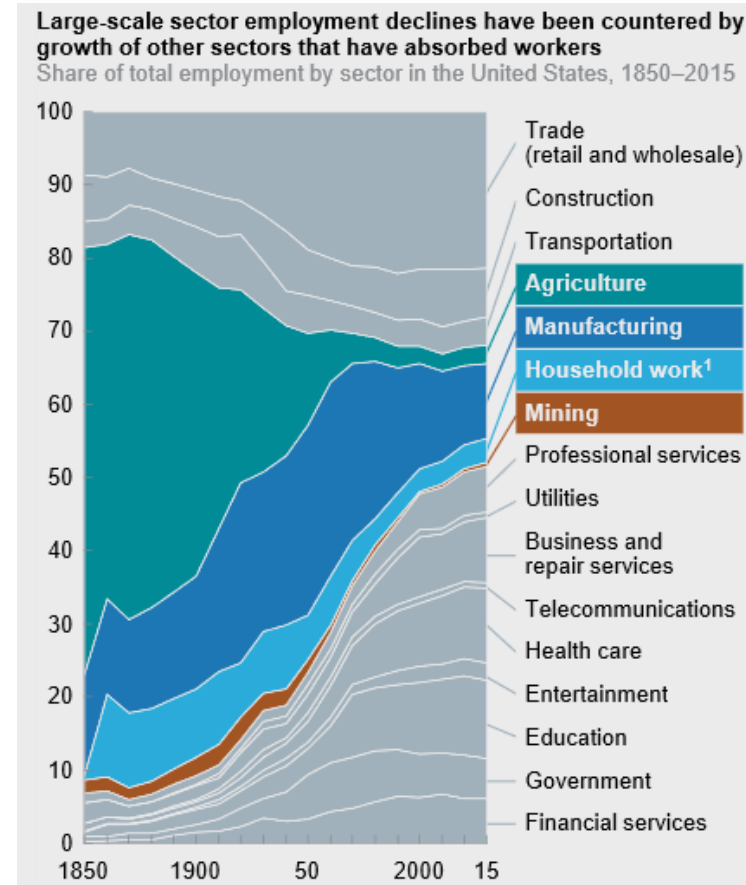Share of total employment by sector in the United States, 1850–2015
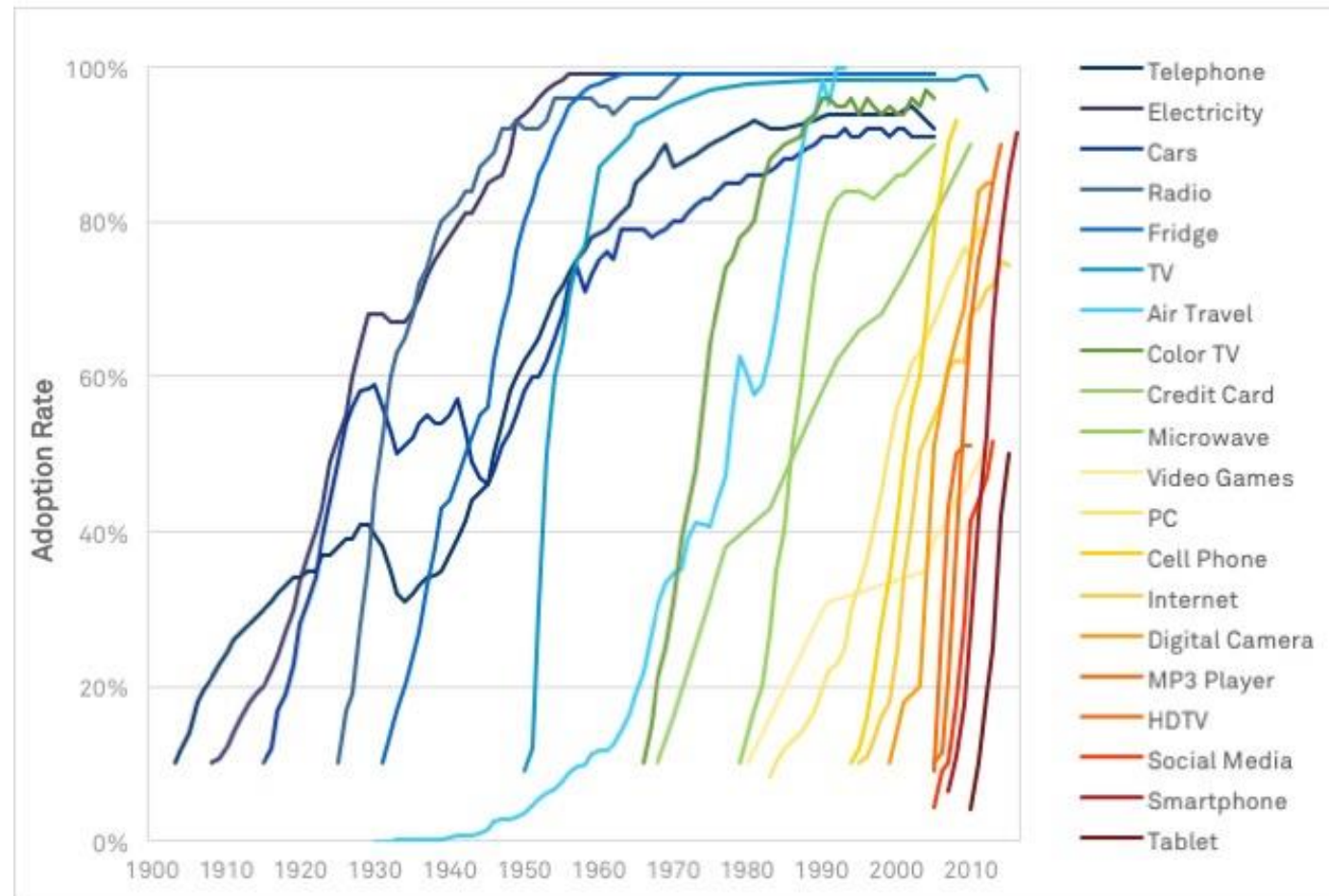
Global picture 1850 - 2015

Top new jobs

Blogger
Digital Marketing Specialist
Social Media Managers
Cloud Computing Specialist
Drone Operators
Mobile App Developers
Sustainability Managers
User Experience Designers
YouTube Content Creators
Data scientist

...

Large-scale sector employment declines have been countered by growth of other sectors that have absorbed workers
Share of total employment by sector in the United States, 1850–2015

Source : McKinsey Global Institute, *Jobs Lost, Jobs Gained: Workforce transitions in a time of automation*, December 2017

Time to adapt is shrinking

Source: Asymco

BLACKROCK

Microsoft

Why AI now ?

History of AI

Microsoft

Turing test invented

collapse in
perception by
government & VCs
Moravec Paradox

First AI winter

market collapse of
for specialised AI
hardware in 1987

Second AI winter

1956

1980

2012

1950

1973

1988

2019

Boom times

Deep learning
revolution
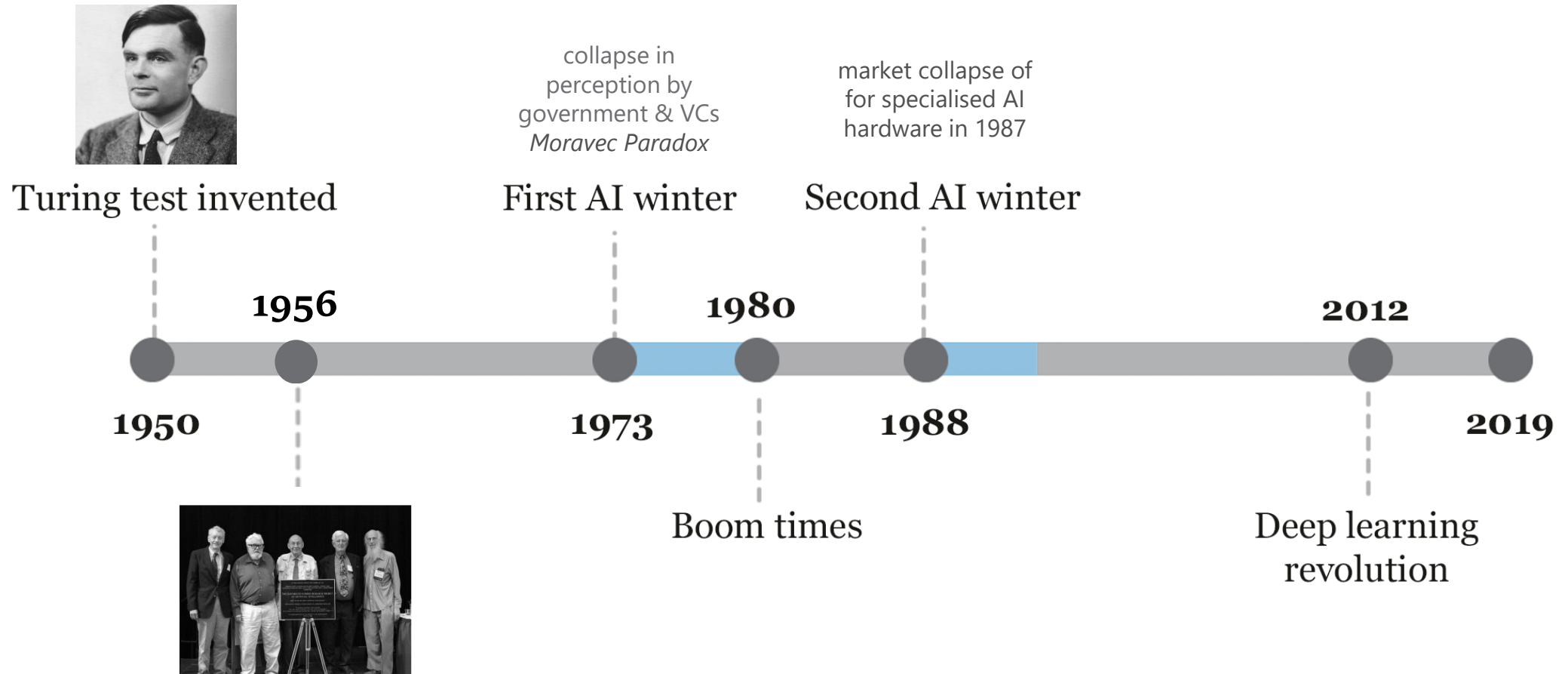
3rd AI wave

Convergence of several trends

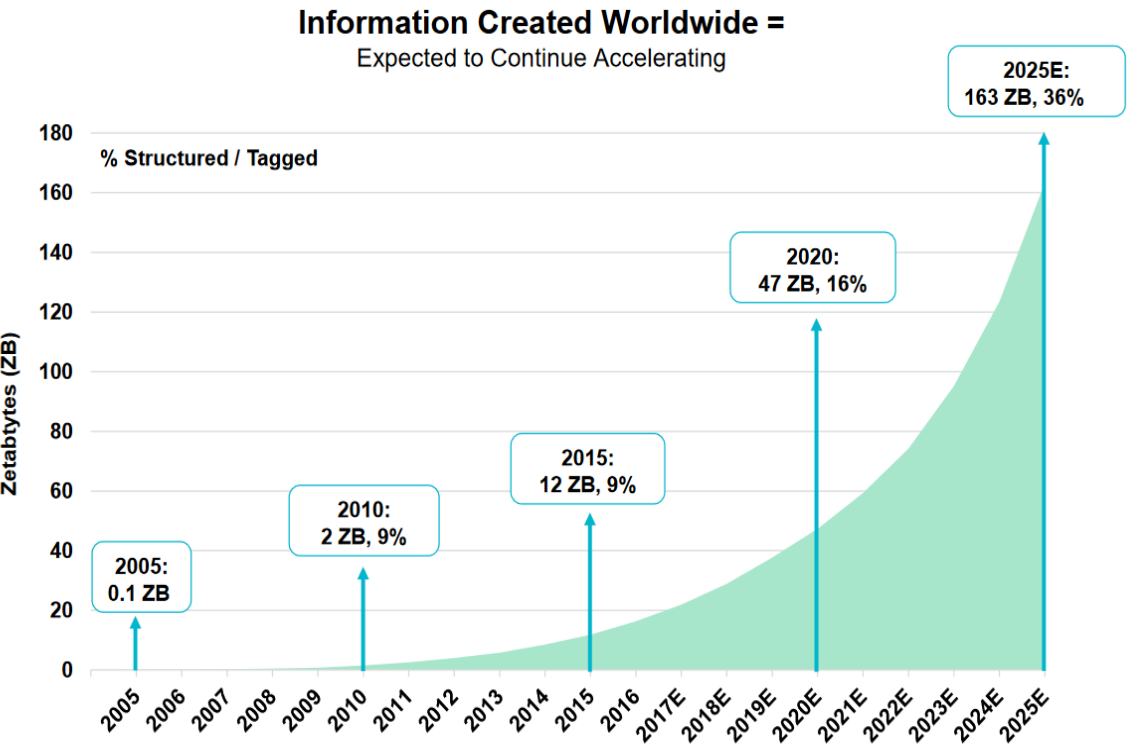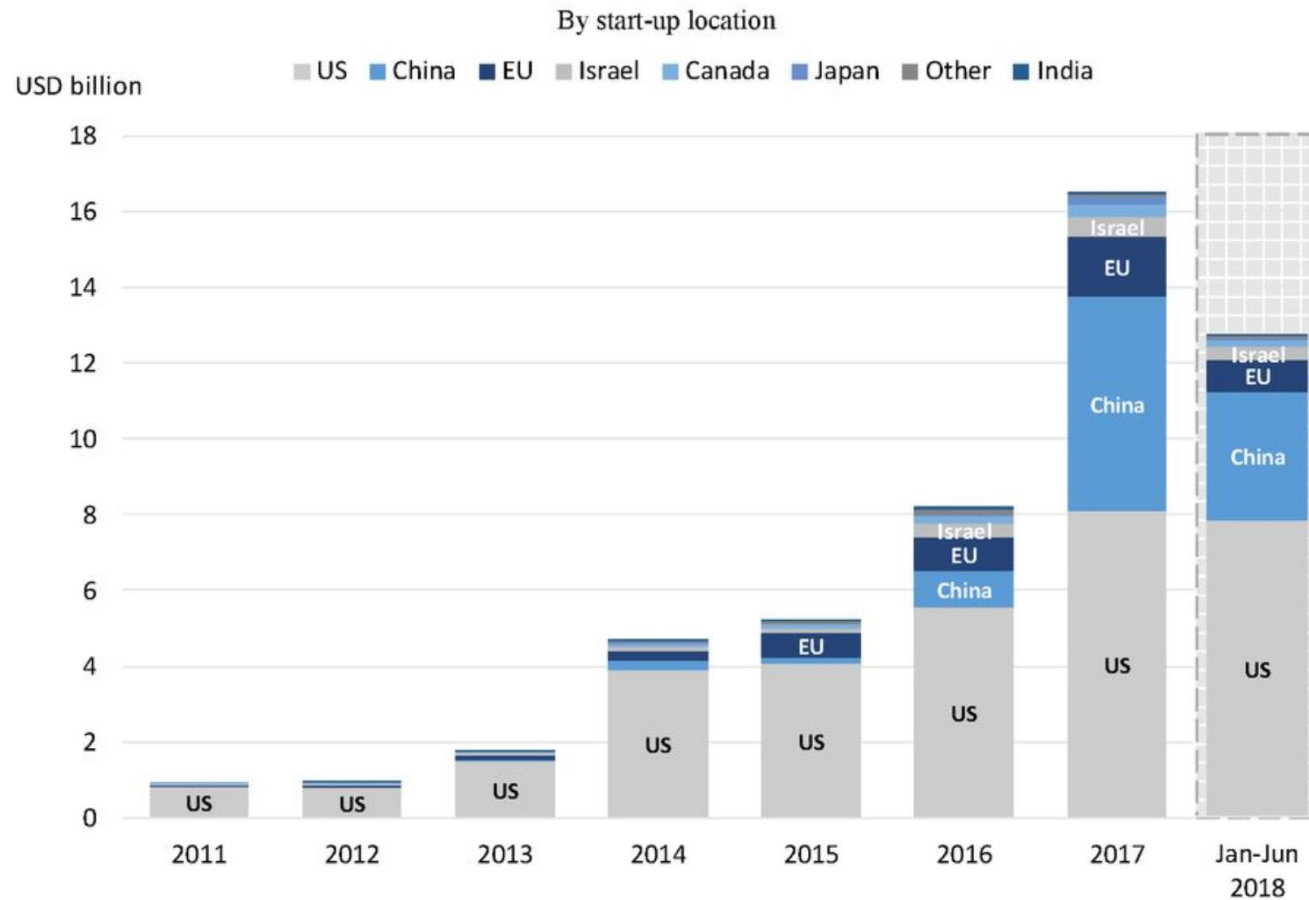- Increased **migration** of socio-economic activities to the Internet

- **Miniaturization & exponential decline** in cost of data collection, sensors, storage & processing (Big Data, IoT, Cloud Computing)

- **Breakthroughs** in machine learning & **pretrained cognitive models** by HSCP

- Exponential growth of **VC investments in AI startups**

Digital universe growth

Information Created Worldwide =
Expected to Continue Accelerating

2025E:
163 ZB, 36%

% Structured / Tagged

2020:
47 ZB, 16%

2015:
12 ZB, 9%

2010:
2 ZB, 9%

2005:
0.1 ZB

Source: IDC DataAge 2025 Study, sponsored by Seagate (3/17)
Note: 1 petabyte = 1MM gigabytes, 1 zeta byte = 1MM petabytes

KP INTERNET TRENDS 2017 | PAGE 132

AI fuel

AI investments per country
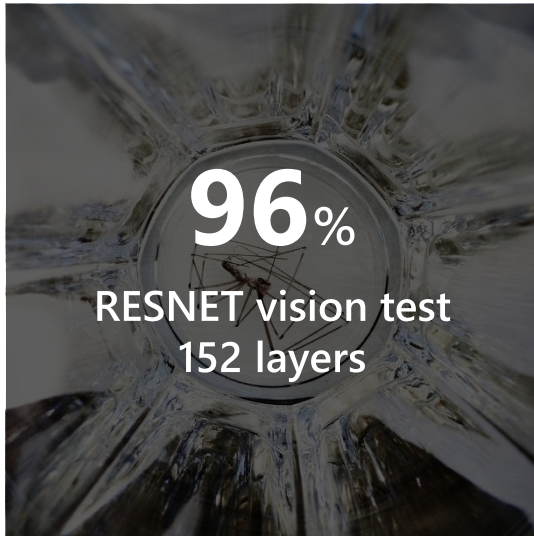

U.S. Executive Order
Maintaining American Leadership in Artificial Intelligence > U.S. leadership on international technical standards as a priority

China
AI Standardization White Paper published by the China Electronics Standardization Institute (CESI)

AI breakthroughs in cognitive functions
Moravec paradox broken

Microsoft

96%
RESNET vision test
152 layers

5.1%
Switchboard speech
recognition test

88.5%
SQuAD reading
comprehension test

69.9%
MT
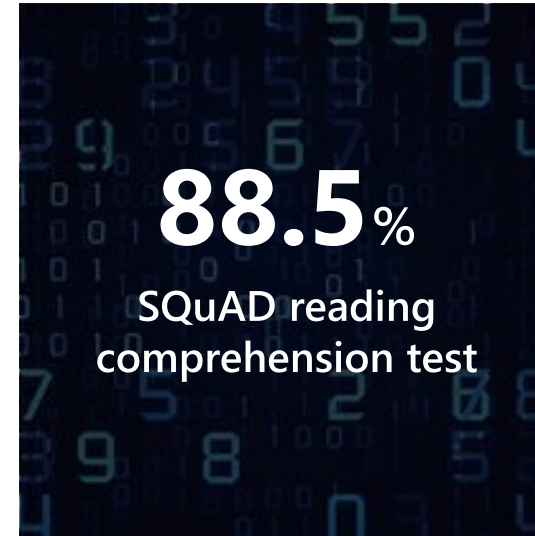research system

2016

Object recognition
Human parity
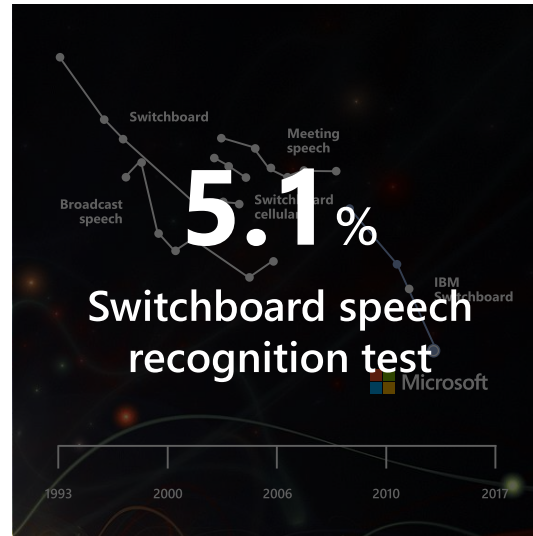
2017

Speech recognition
Human parity

January 2018

Machine reading comprehension
Human parity

March 2018

Machine translation
Human parity

AI implementations

Microsoft

AI use in high stakes decisions

Credit

Admissions

Sentencing

Hiring

Driving

{LawGeex}
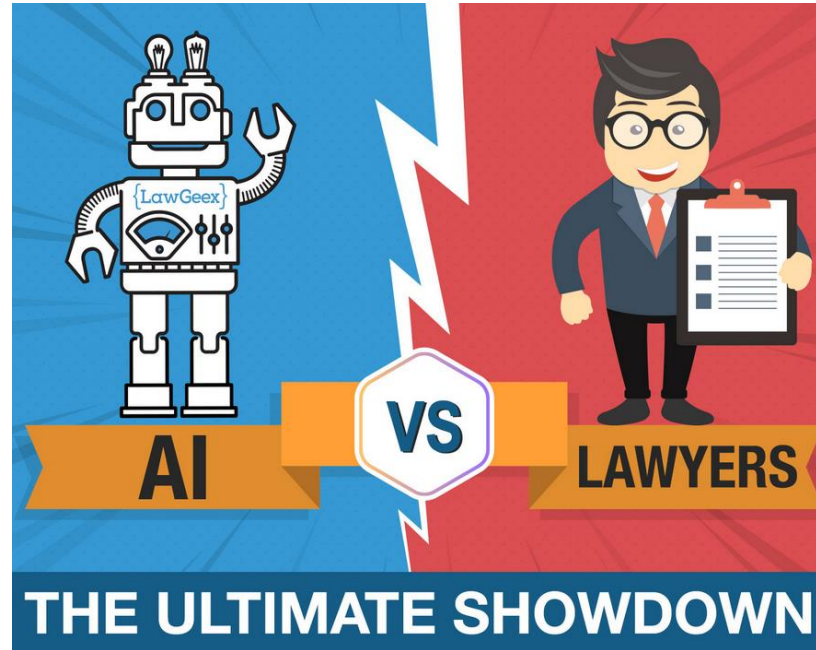
White-collar case study

NDA benchmark AI vs lawyers

NDA experiment

20 US-trained lawyers

AI trained with thousands of NDA

AI untrained on the experiment

4 hours

5 NDAs

30 issues to spot

Source : www.lawgeex.com/resources/AIvsLawyer
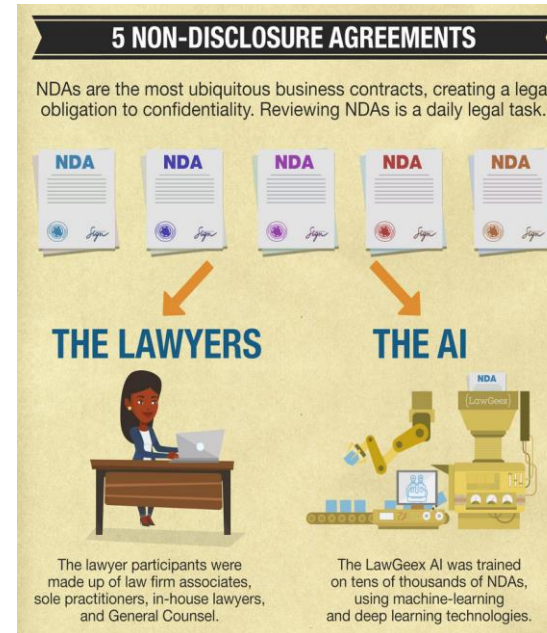
Legal NDAs – results

| Coffees | **0** | 12 |

Legal NDAs – results

Coffees	**0**	12
Accuracy	**94%**	85%

Legal NDAs – results

Microsoft

	AI	Lawyers
Coffees	**0**	12
Accuracy	**94%**	85%
Time	**26 s**	5.520 s (92 min)

Unfair & types of harm : QoS

The New York Times

Facial Recognition Is Accurate, if You're a White Guy

By STEVE LOHR   FEB. 9, 2018

1%

Gender was misidentified in **up to 1 percent of lighter-skinned males** in a set of 385 photos.

12%

Gender was misidentified in **up to 12 percent of darker-skinned males** in a set of 318 photos.

7%

Gender was misidentified in **up to 7 percent of lighter-skinned females** in a set of 296 photos.

35%

Gender was misidentified in **35 percent of darker-skinned females** in a set of 271 photos.

Unfair & types of harm : over/under representation & stereotyping
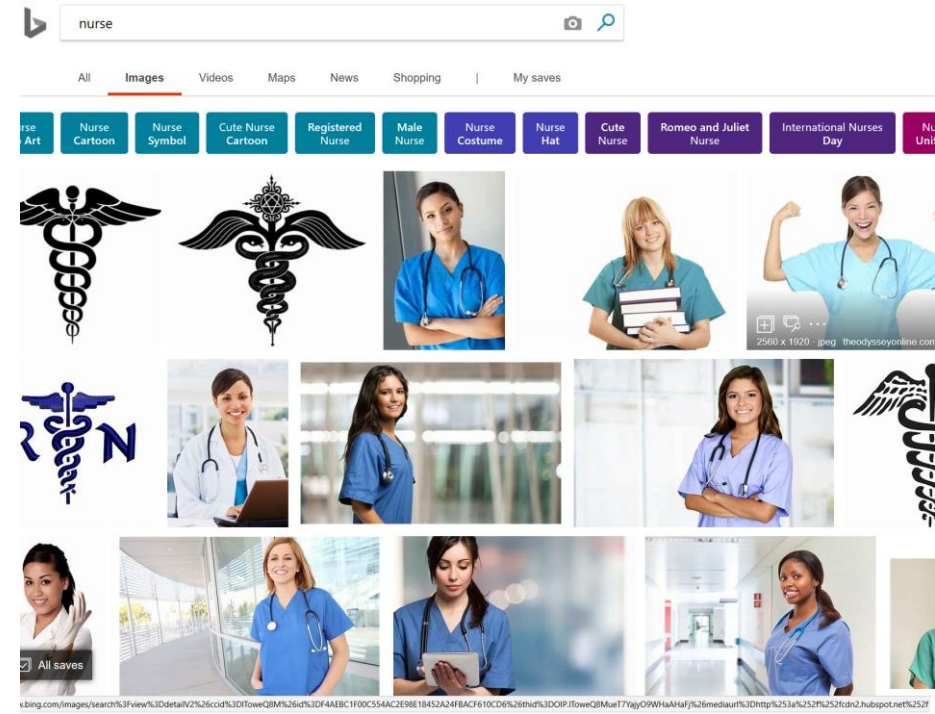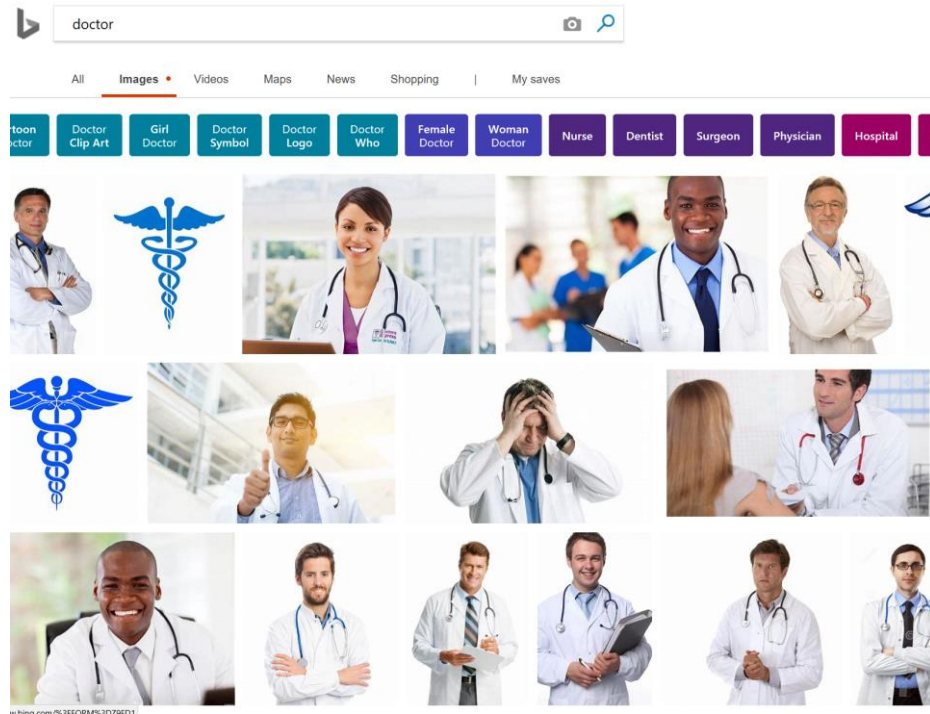
Morality - Infamous "trolley" problem

Microsoft
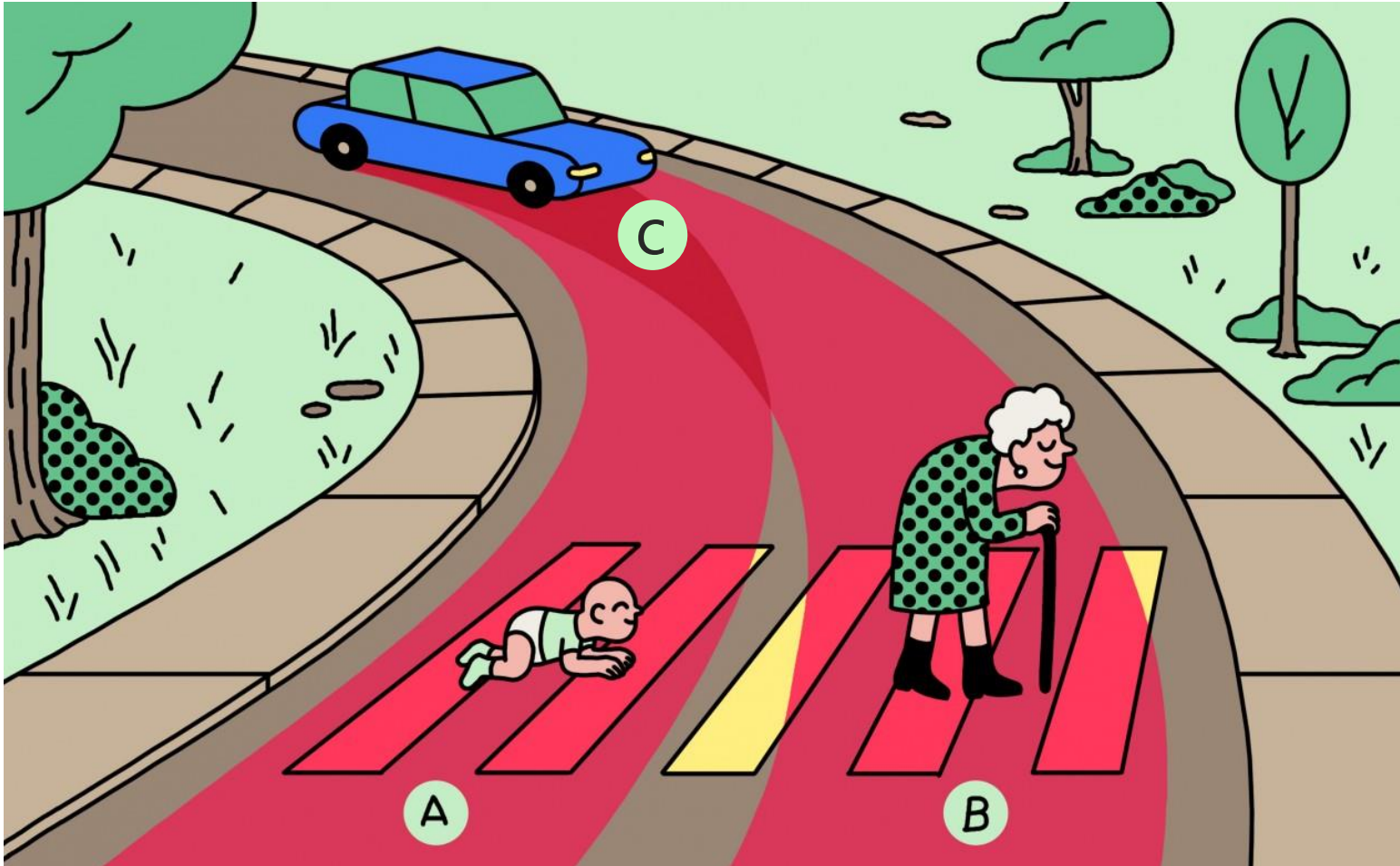
Should a self-driving car kill the baby (A) or the grandma (B) or the driver (C) ?

Ethics & morality

- From **niche/narrow AI to general AI** – major risk

- AI expands circle of moral agents beyond humans to artificially intelligent systems called **artificial moral agents (AMAs)**

- Challenge of designing agents respecting set of values & laws demanded by **human moral agents (HMA)**

Figure 2.1. Two Dimensions of AMA Development.

Microsoft
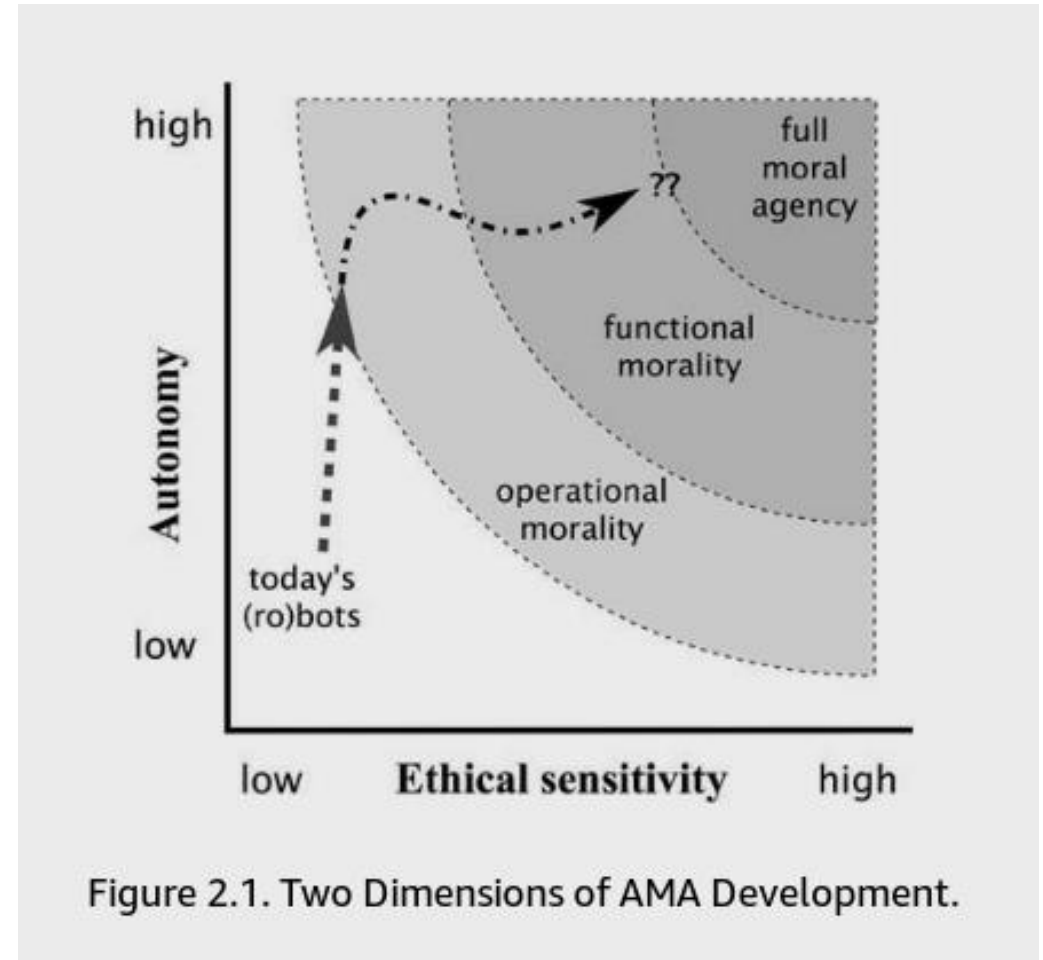
What does it take to trust machine decision-making?

Is it.....

Accurate?
Fair?
Interpretable?
Tamper-Proof?
Accountable?

Microsoft

Ethics of AI

- **Fairness:** AI systems should treat people fairly

- **Reliability and Safety:** AI systems should perform reliably and safely

- **Privacy and Security:** AI systems should be secure and respect privacy

- **Inclusiveness:** AI systems should empower everyone and engage people

- **Transparency:** AI systems should be understandable

- **Accountability:** AI systems should have algorithmic accountability
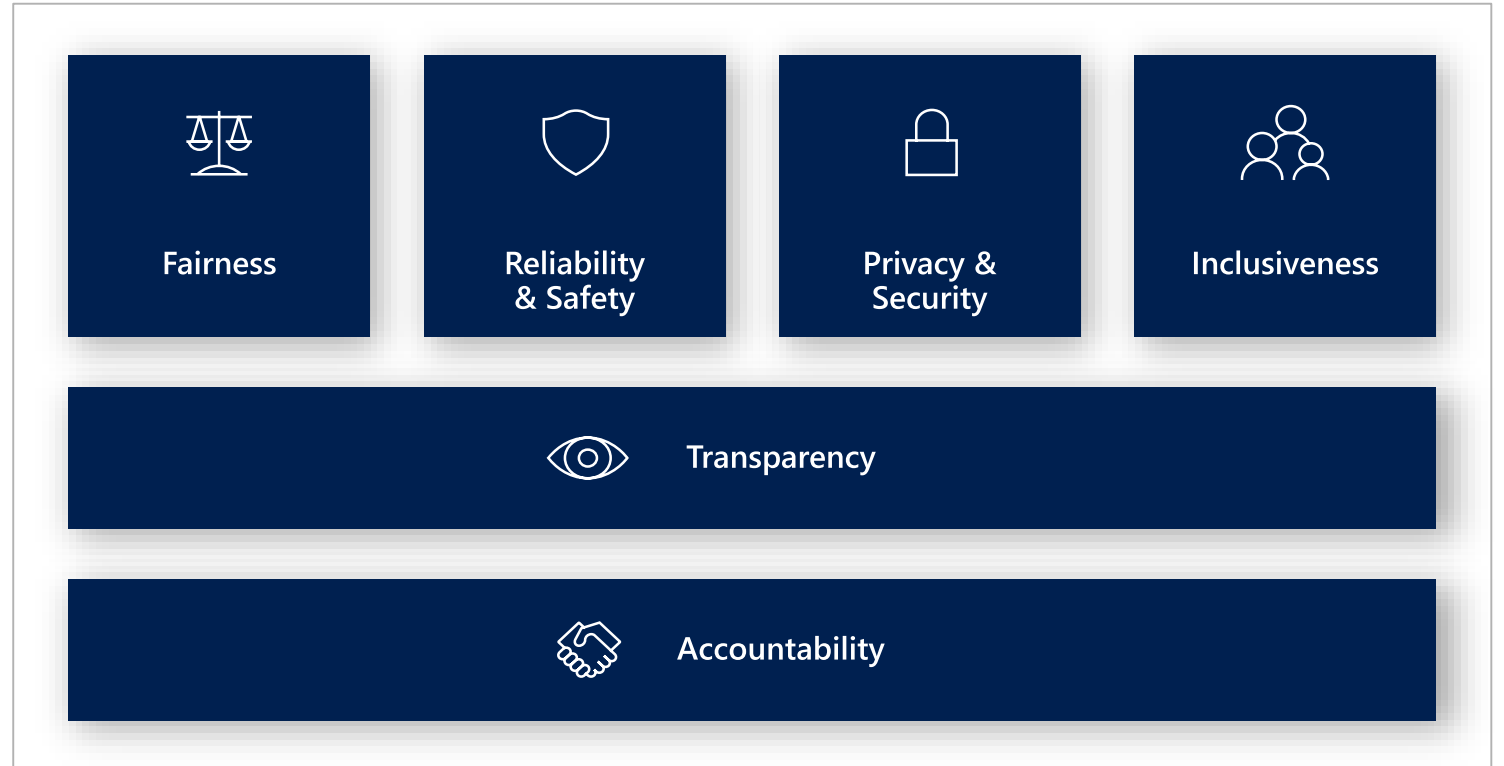
Microsoft

Microsoft Aether Committee

AI, Ethics, and Effects in Engineering and Research

Sensitive Uses

Reliability and Safety

Human-AI Collaboration & Interaction

Fairness and Bias

Intelligibility & Explanation

Engineering Practices

Human Attention & Cognition

Ethics of AI

Fairness	Reliability & Safety	Privacy & Security	Inclusiveness

Transparency

Accountability

Microsoft
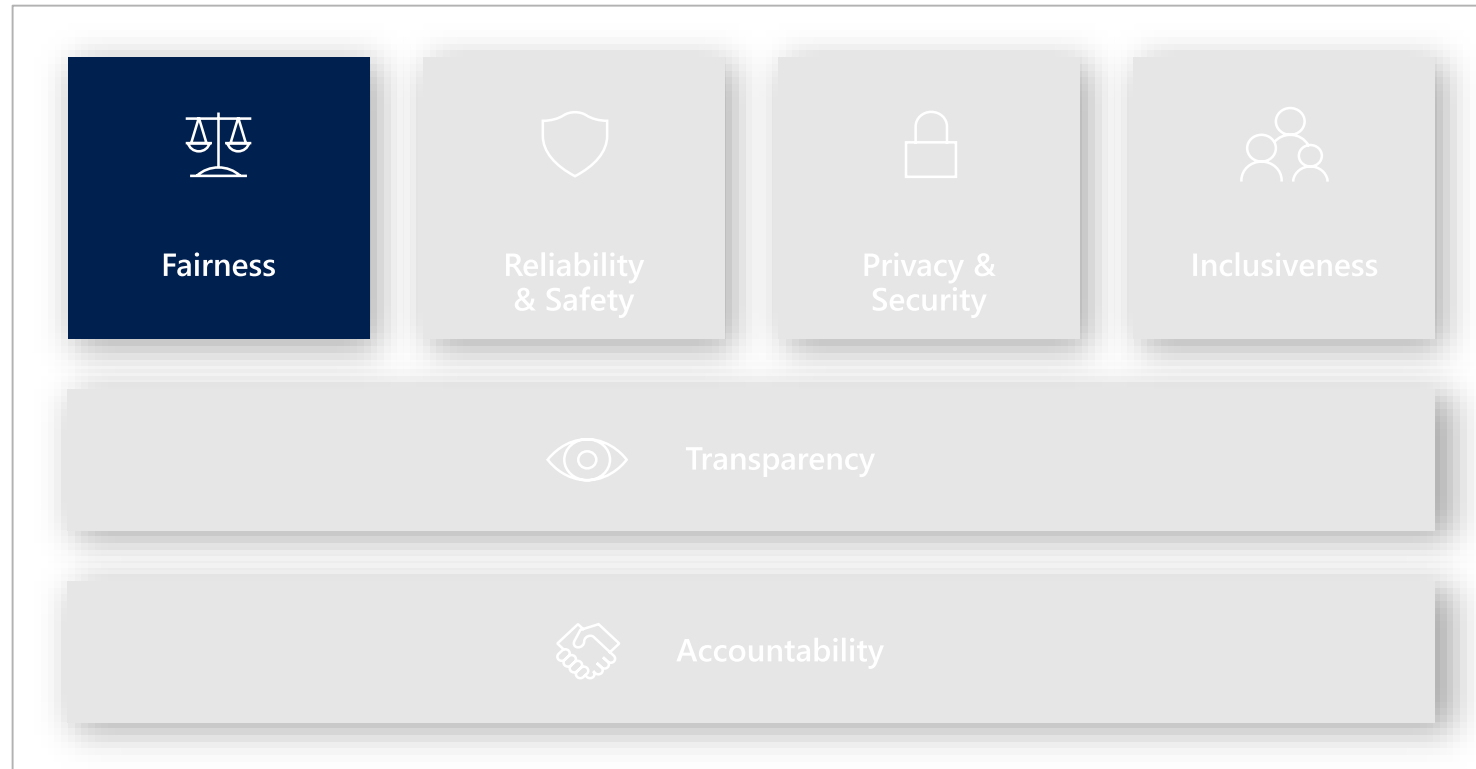
Development process example

Fairness

Does the model encourage feedback loops that can produce increasingly unfair outcomes?

Is an algorithm an ethical solution to our problem?
Could algorithm be misused in other contexts?

Are we deploying our model on a population that we did not train/test on?

Does our data include enough minority samples? Is the data skewed? Can we collect more data or reweight?
Are there missing/biased features?
Was our historical data generated by a biased processed that we reify?
Do our labels reinforce stereotypes?
Do we need to apply debiasing algorithms to preprocess our data?

Have we evaluated the model using relevant fairness metrics?
Do our selected fairness metrics capture our customers needs?
Can we evaluate the model on other datasets beyond test set?

Is the objective function in line with ethics?
Do we need to include fairness constraints in the function?
Do our proxies really measure what we think they do?
Do we need to model minority populations separately?

Feedback

Problem Formation

Deployment

Dataset Construction

Testing Process

Algorithm Selection

Training Process

Possible misuse

Minority samples

Untrained population

Measuring different types of fairness - community

Metrics Project	Description	Repository
AI Fairness 360	Comprehensive collection of fairness metrics, pre- and post-processing debiasing algorithms.	http://aif360.mybluemix.net/
MSR tool kit	Offers similar functionality to AI Fairness 360 with more performant debiasing algorithms.	In development. Contact Jenn Wortman Vaughan
Fairness Measures	Framework to test given algorithm on a variety of datasets	https://github.com/megantosh/fairness_measures_code
Fairness Comparison	Compares ML algorithms with respect to fairness measures.	https://github.com/algofairness/fairness-comparison
Themis-ML	Python library implementing fairness-aware machine learning algorithms	https://github.com/comicBboy/themis-ml
FairML	Quantifies dependence of model outputs on inputs	https://github.com/adebayoj/fairml
Aequitas	Web and python auditing tool. Generates bias report for model/dataset	https://github.com/dssg/aequitas
Fairtest	Audits algorithms impact on protected subpopulations	https://github.com/columbia/fairtest
Themis	Designs test cases to explore where algorithm might be exhibiting group-based discrimination	https:.://github.com/LASER-UMASS/Themis
Audit-AI	Python library to audit scikit-learn models	https://github.com/pymetrics/audit_ai

Microsoft

Ethics of AI

Fairness	Reliability & Safety	Privacy & Security	Inclusiveness

Transparency

Accountability

Microsoft

Transparency & intelligibility

$$f$$

$$\boldsymbol{x}$$

Input

?

$$f(\boldsymbol{x})$$

Classification result

Microsoft

Transparency & intelligibility

f

x

Input

?

$f(x)$

Classification
result

logistic   GAMs   GA2Ms   boosted   neural
regression   trees   networks

interpretable   flexible

Microsoft
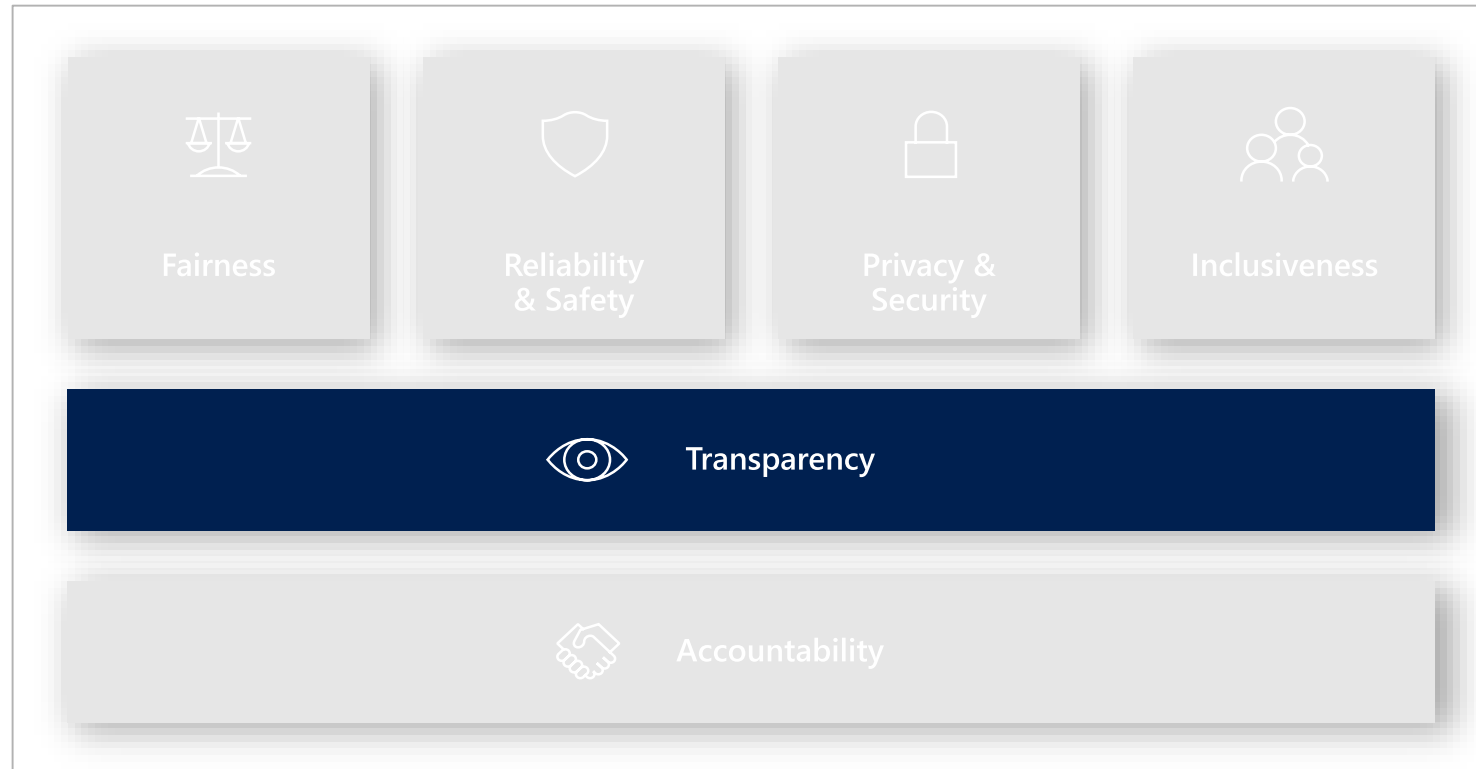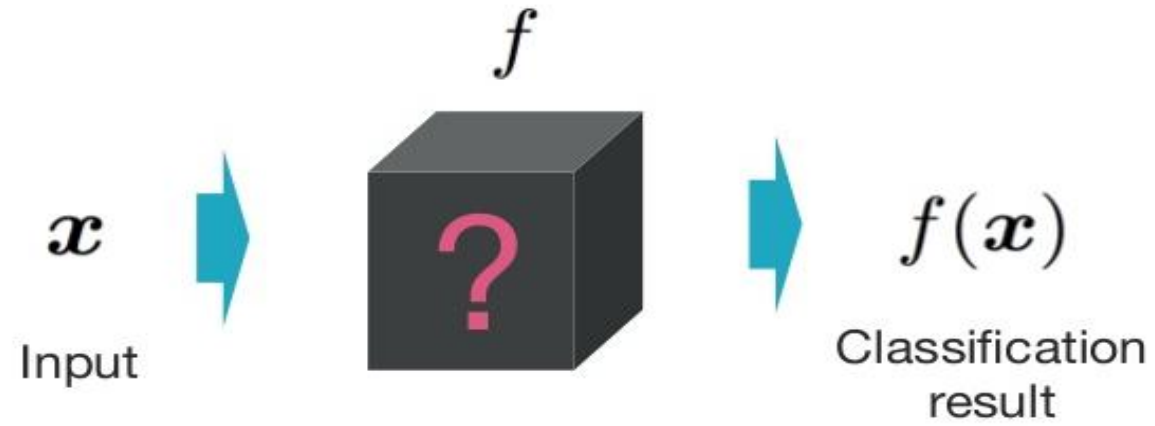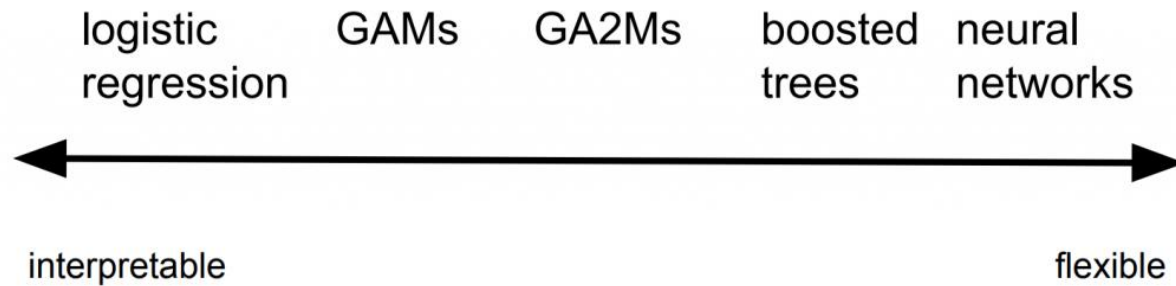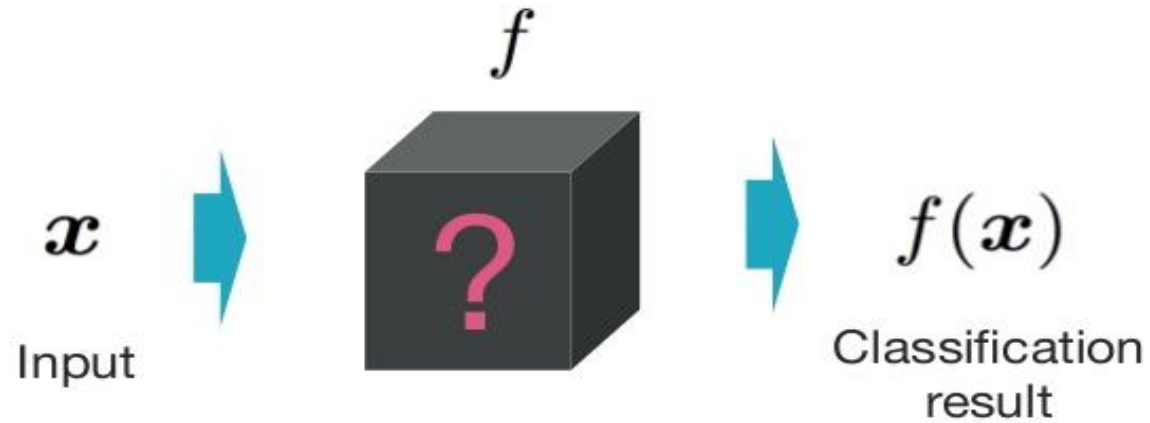
T&I – Personal medicine

Transparency & intelligibility

f

x → Input

$f(x)$ Classification result

logistic regression   GAMs   GA2Ms   boosted trees   neural networks

interpretable ←——————————————————→ flexible
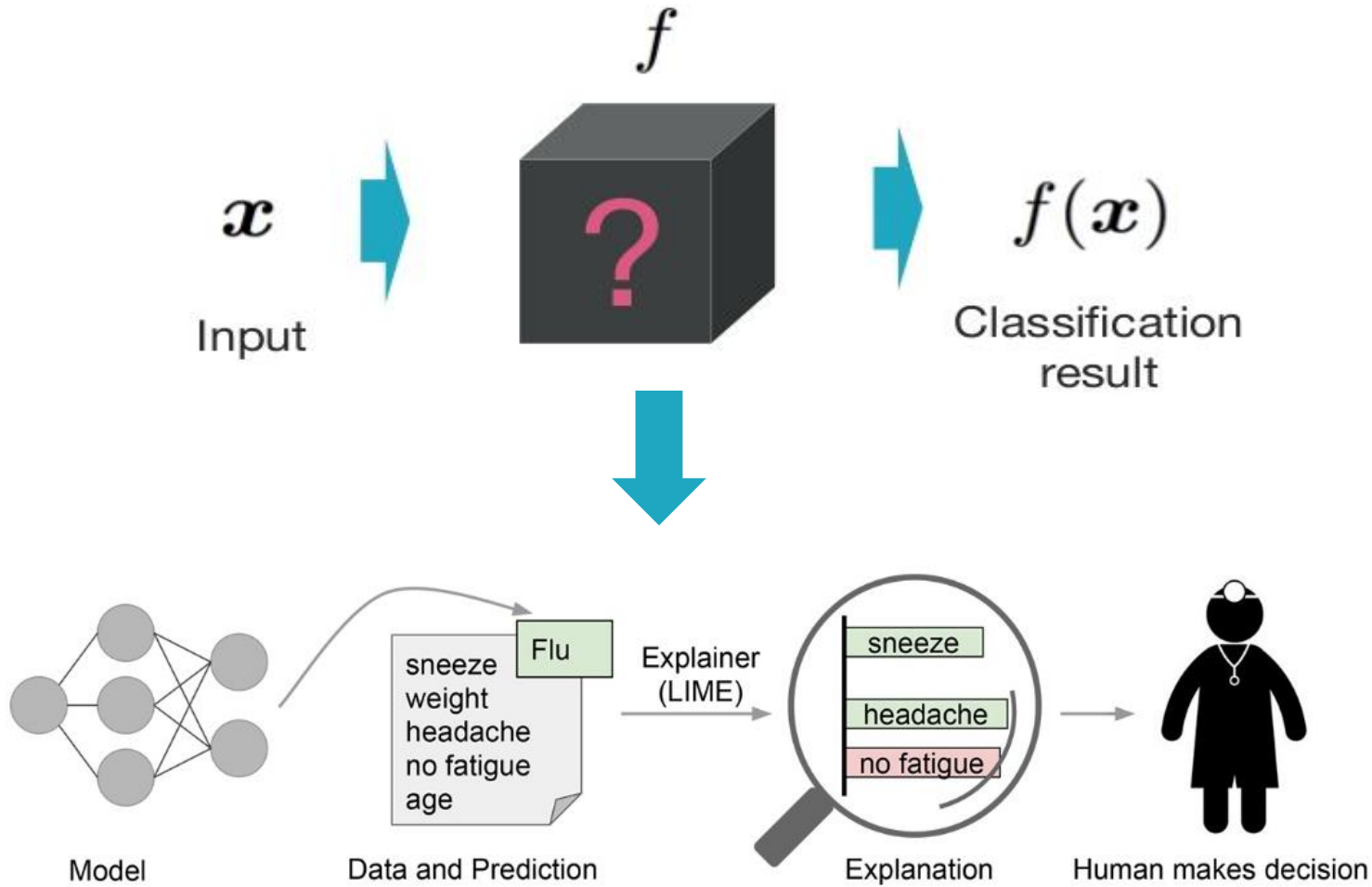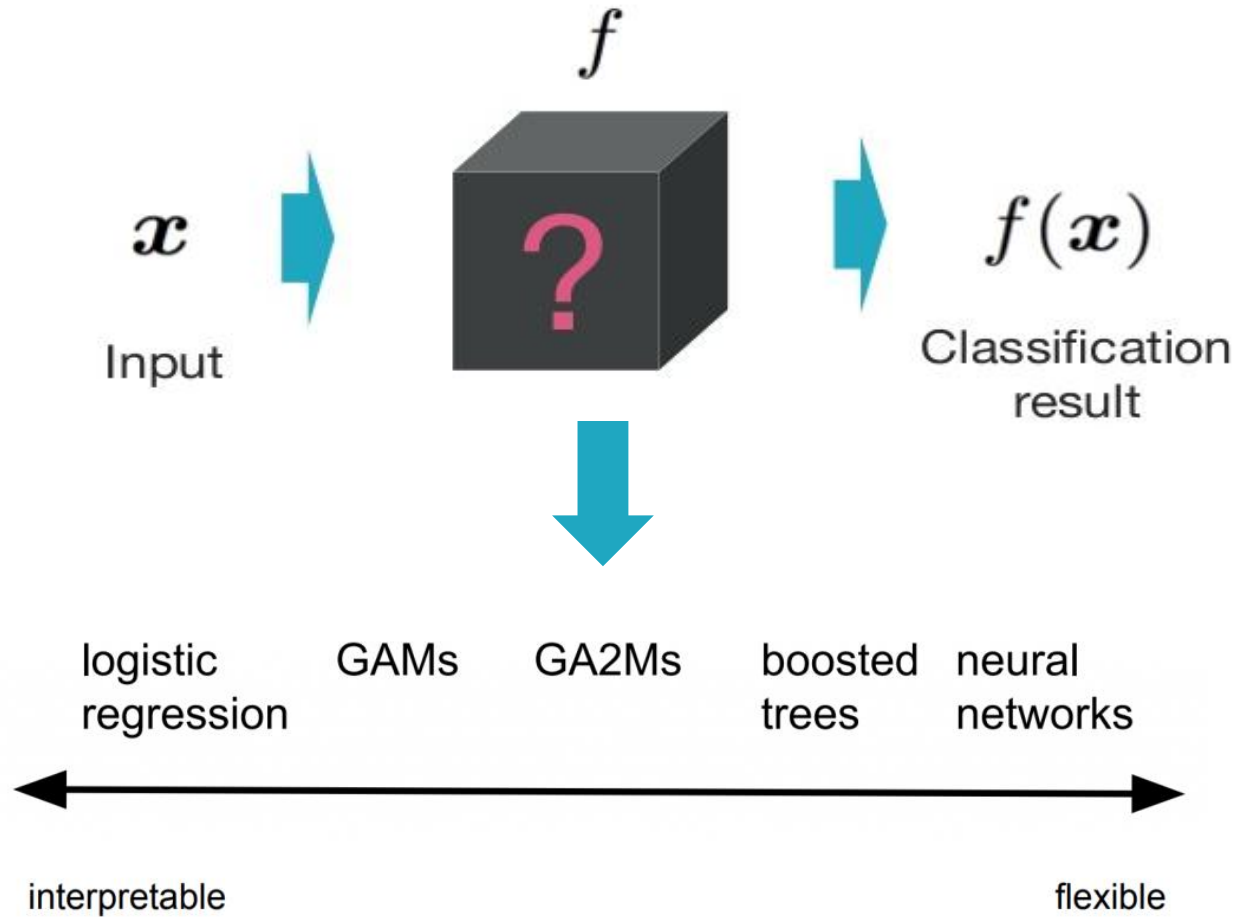
- highest accuracy through complex models **difficult to interpret**

- **Creates tension** between accuracy and interpretability

- various methods proposed to interpret predictions but often unclear how methods are related & preferable

Microsoft

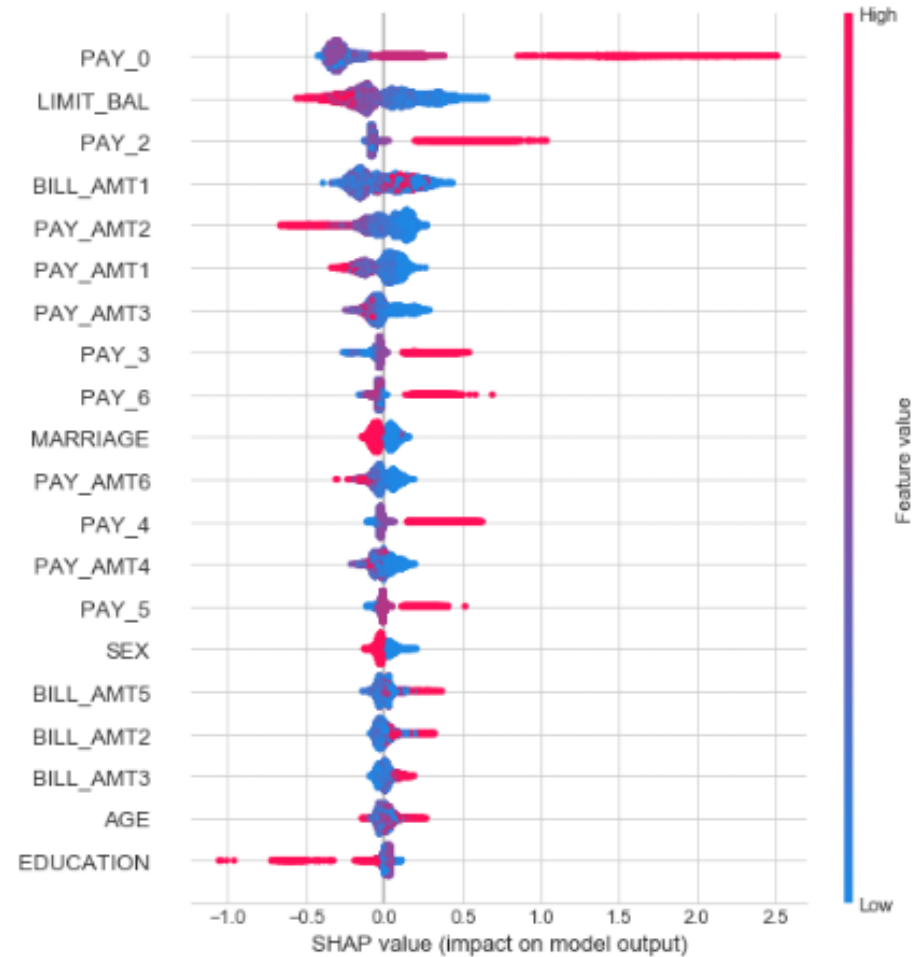# Transparency & intelligibility – post-hoc explanations

- Methods: LIME, SHAP, Sensitivity Analysis, Model Distillation

- SHAP – Shapley Addictive exPlanations
  - addresses this problem
  - incorporates six existing, individual model interpretation methods

Transparency & intelligibility

Understanding **why** a model makes certain predictions is as crucial as the prediction **accuracy**

AI skills, standardization & regulation

AI in Luxembourg

Microsoft

Virtual agents KYC (insurance)

Virtual agents customer QnA (administration)

Telesales campaign conversion-rate augmentation (insurance)

Self-quote bot (logistics)

Skills scoring (administration)

Process output prediction (industry)

Ad positioning (B2C)

...

AI in Luxembourg

Microsoft

Many **unanswered** questions at Luxembourg corporations

- Who is **skilled** internally on AI ? HR, CTO/CxO, Sales, Marketing, Compliance, …

- Where can I **skill** myself on AI ?
  University of Luxembourg/AISE, AI Academy Luxembourg, ILNAS SC42 mirror committee

- Who monitors the **decision** process of AI ? Luxembourg regulator, compliance
  manager, AI learning manager, data scientists

- Who is looking if the AI implementation is **responsible & ethical** ? AI compliance
  manager, AI ethical committee, AI internal & external auditors

- Who **certifies** the AI implementation ?

Lack of skills vs GDP

Forecasted GDP impact

Potential Forgone Growth Premium.
Absolute additional cumulative value at risk, 2018-2028, $US bn.
And additional % points of GDP growth at risk every year.
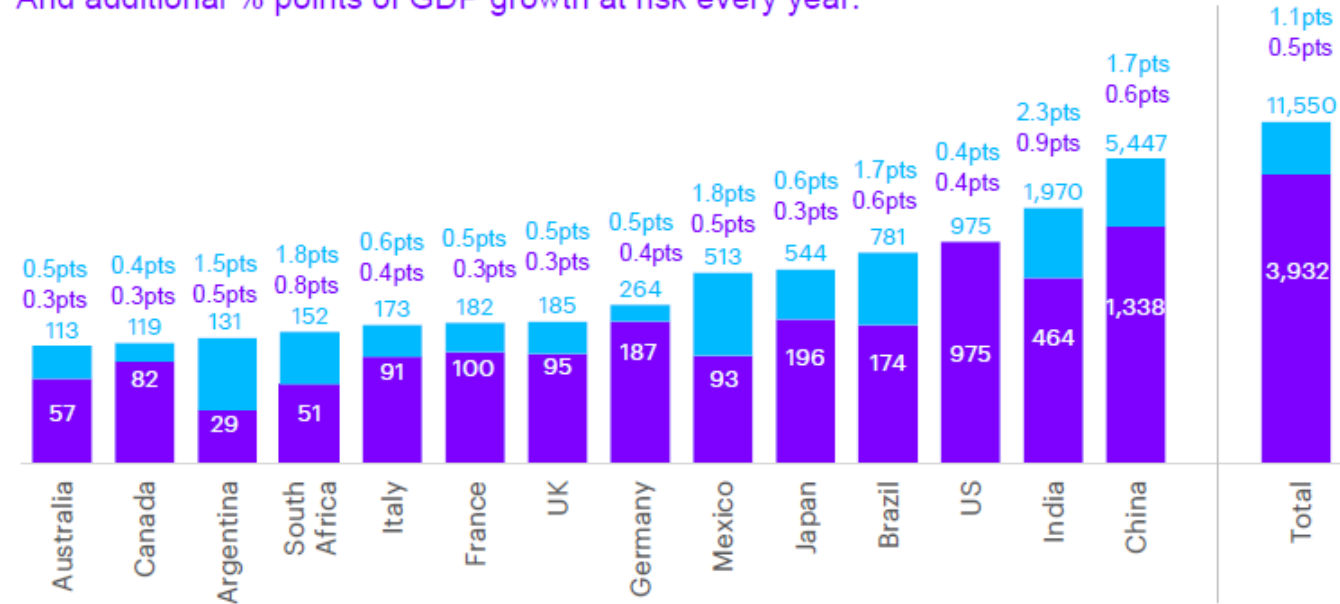
Scenario 1: Investments in intelligent technologies per worker in each country reach the country's current investment levels in traditional technologies per worker.

Scenario 2: Investments in intelligent technologies per worker in each country reach current US investment levels in traditional technologies per worker.

Source : How to accelerate skills acquisition in the age of intelligent technologies, *Accenture*

International & national developments

Figure 2. International AI standards under development

Network - Product	Network - Process
• Foundational Standards: Concepts and terminology (SC 42 WD 22989), Framework for Artificial Intelligence Systems Using Machine Learning (SC 42 WD 23053) • Transparency of Autonomous Systems (defining levels of transparency for measurement) (IEEE P7001) • Personalized AI agent specification (IEEE P7006) • Ontologies at different levels of abstraction for ethical design (IEEE P7007) • Wellbeing metrics for ethical AI (IEEE P7010) • Machine Readable Personal Privacy Terms (IEEE P7012) • Benchmarking Accuracy of Facial Recognition systems (IEEE P7013)	• Model Process for Addressing Ethical Concerns During System Design (IEEE P7000) • Data Privacy Process (IEEE P7002) • Methodologies to address algorithmic bias in the development of AI systems (IEEE P7003). • Process of Identifying and Rating the Trustworthiness of News Sources (IEEE P7011)
Enforced - Product	**Enforced - Process**
• Certification for products and services in transparency, accountability, and algorithmic bias in systems (IEEE ECPAIS) • Fail-safe design for AI systems (IEEE P7009)	• Certification framework for child/student data governance (IEEE P7004) • Certification framework for employer data governance procedures based on GDPR (IEEE P7005) • Ethically Driven AI Nudging methodologies (IEEE P7008)

MNCs like Google & Microsoft participate in AI SC42 developments

Certification scheme / AIMS
to support consumer trust on
products, services & processes

Source : *Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development*, Peter Cihon, Research Affiliate, Center for the Governance of AI, Future of Humanity Institute, University of Oxford

International & national developments

Standardization drives **new skill opportunities**

AI auditors
AI certifiers
AI trainers
AI regulators
...

Source : *Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development*, Peter Cihon, Research Affiliate, Center for the Governance of AI, Future of Humanity Institute, University of Oxford

"AI will be either the best, or the worst thing, ever to happen to humanity."

———

STEPHEN HAWKING

Responsible AI to **amplify human** ingenuity

Microsoft

soil temperature
71°

pH 6.0

6CO₂ + 6H₂O + ENERGY = C₆H₁₂O₆ + 6O₂