

AI ethics, governance, and regulation

Francesca Rossi

IBM Fellow and AI Ethics Global Leader



Current AI Applications

- Digital assistants:
 - Home assistants (Alexa)
 - Travel assistants (Waze)
- Driving/travel support:
 - Auto-pilot (Tesla)
 - Ride-sharing apps (Uber, Lyft)
- Customer care:
 - Client service chatbots
- Online recommendations:
 - Friend recommendations (Facebook)
 - Purchase recommendations (Amazon)
 - Movie recommendations (Netflix)
- Media and news:
 - Ad placement (Google)
 - News curation
- Healthcare:
 - Medical image analysis
 - Treatment plan recommendation
- Financial services:
 - Credit risk scoring
 - Loan approval
 - Fraud detection
- Job market:
 - Resume prioritization
- Judicial system:
 - Recidivism prediction (Compas)

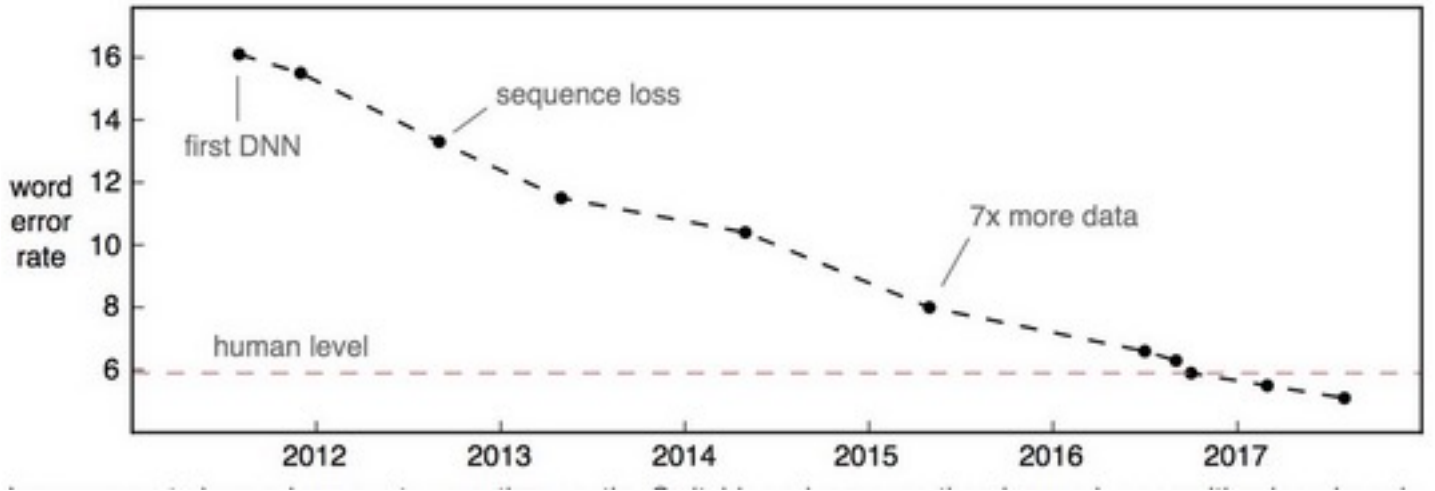
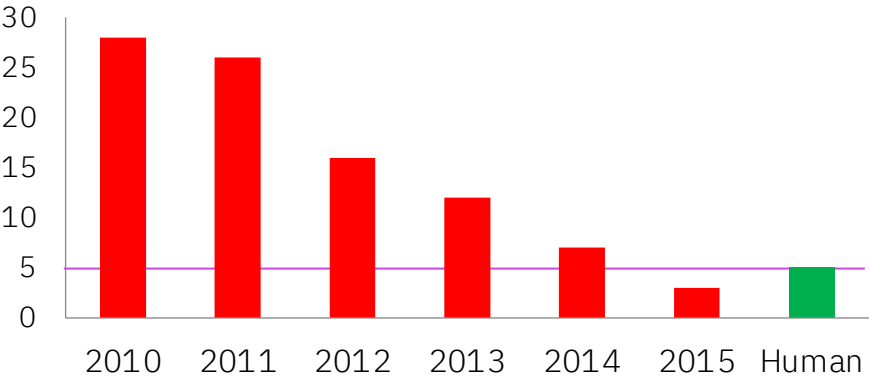
AI capabilities: Image and natural language interpretation



Woman holding a cask of bananas



A group of young people playing fresbee



Main AI Ethics issues

AI needs data

- Data privacy and governance

AI is often a black box

- Explainability and transparency

AI can make or recommend decisions

- Fairness and value alignment

AI is based on statistics and has always a small percentage of error

- Who is accountable if mistakes happen?

AI can profile people and manipulate their preferences

- Human and moral agency

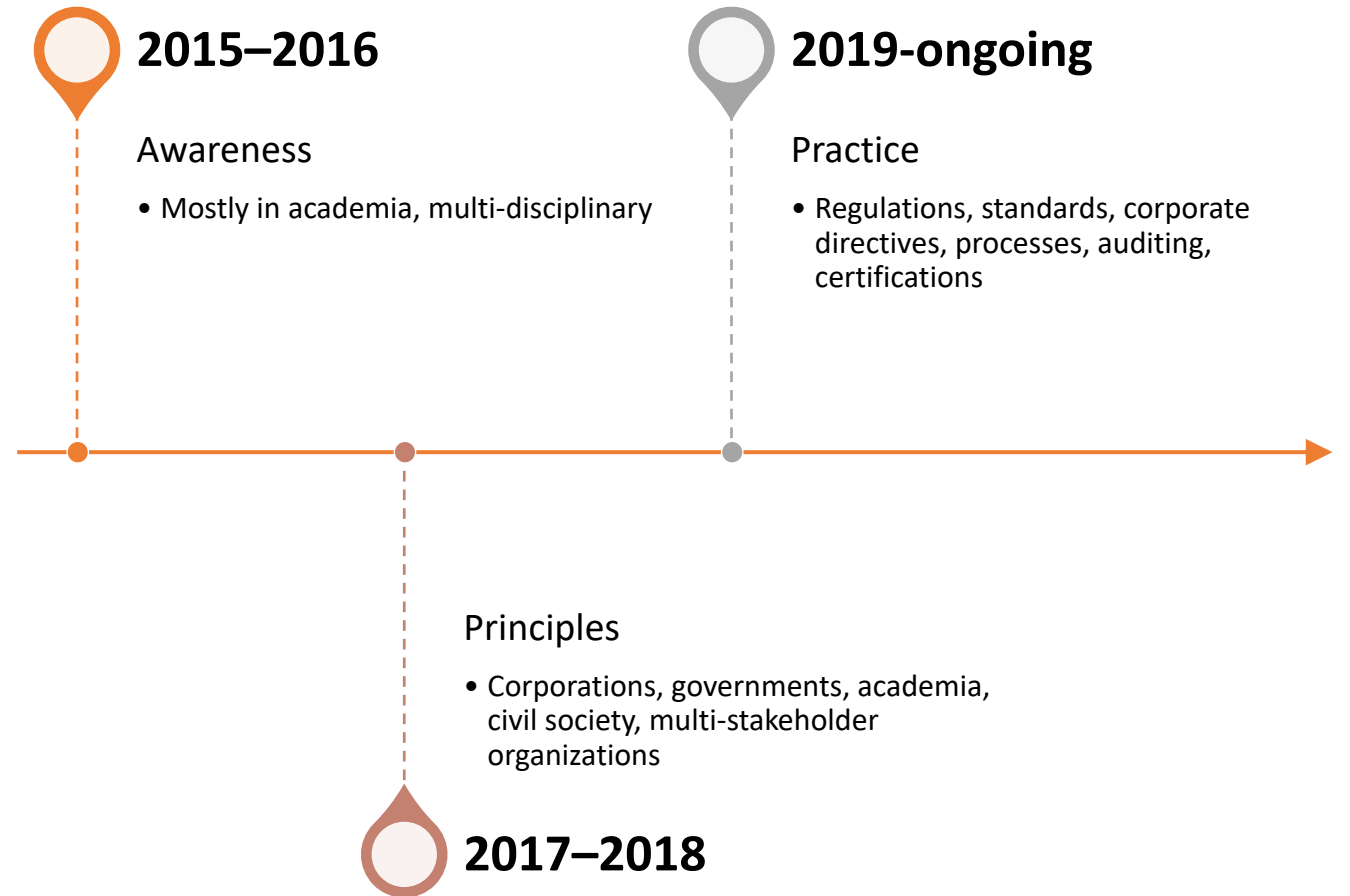
AI is very pervasive and dynamic

- Larger negative impacts for tech misuse
- Fast transformation of jobs and society

Good or bad use of the technology

- Autonomous weapons and mass surveillance
- UN Sustainable Development Goals

AI Ethics 3.0



The EU AI Act -- A regulation proposal

April 2021



- Regulating AI applications, not AI itself
- Regulation based on AI applications' risk
 - 4 levels of risk: unacceptable, high, low, minimal
- Building trust in AI and the whole AI ecosystem
- Technical documentation, record keeping, and transparency/explainability for high-risk AI systems
- Transparency obligations for lower-risk AI systems
- Human oversight
- Obligations on providers, users, and other stakeholders

IBM Policy Lab

Precision Regulation for Artificial Intelligence



Human agency
and oversight



Diversity, non-
discrimination and
fairness
Societal &
environmental
well-being
Accountability



Technical Robustness
and Safety



Privacy and data
governance

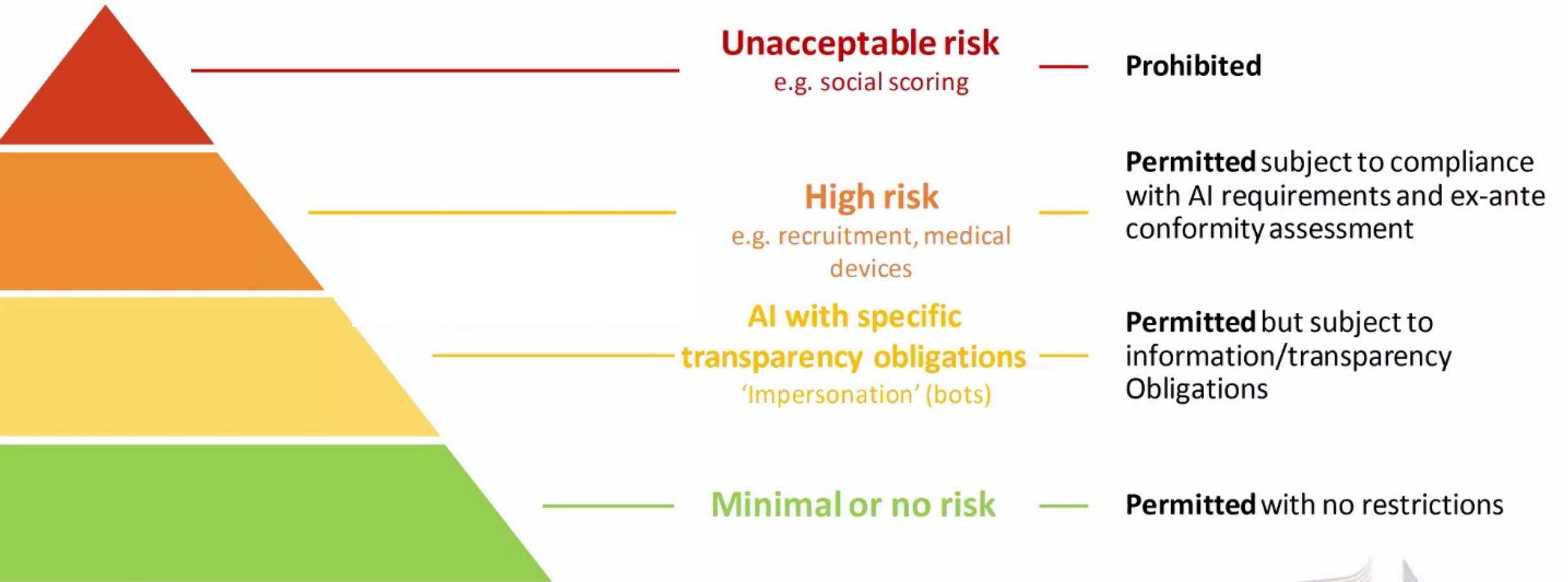


Transparency



IBM

The European AI Act: a risk-based regulation of AI uses



AI that contradicts EU values is prohibited (Title II, Article 5)

X

Subliminal manipulation
resulting in physical/
psychological harm

Example: An **inaudible sound** is played in truck drivers' cabins to push them to **drive longer than healthy and safe**. AI is used to find the frequency maximising this effect on drivers.

X

**Exploitation of children
or mentally disabled persons**
resulting in physical/psychological harm

Example: A doll with an integrated **voice assistant** encourages a minor to **engage in progressively dangerous behavior** or challenges in the guise of a fun or cool game.

X

**General purpose
social scoring**

Example: An AI system **identifies at-risk children** in need of social care **based on insignificant or irrelevant social 'misbehavior'** of parents, e.g. missing a doctor's appointment or divorce.

X

**Remote biometric identification for law
enforcement purposes in publicly accessible
spaces (with exceptions)**

Example: All faces captured live by video cameras checked, in real time, against a database to identify a terrorist.

High-risk Artificial Intelligence Systems (Title III, Annexes II and III)



Certain applications in the following fields:

1 SAFETY COMPONENTS OF REGULATED PRODUCTS

(e.g. medical devices, machinery) which are subject to third-party assessment under the relevant sectorial legislation

2 CERTAIN (STAND-ALONE) AI SYSTEMS IN THE FOLLOWING FIELDS

- ✓ Biometric identification and categorisation of natural persons
- ✓ Management and operation of critical infrastructure
- ✓ Education and vocational training
- ✓ Employment and workers management, access to self-employment
- ✓ Access to and enjoyment of essential private services and public services and benefits
- ✓ Law enforcement
- ✓ Migration, asylum and border control management
- ✓ Administration of justice and democratic processes

Why should
a company
building or
using AI care
about AI
ethics and
regulation?

Company values

Company reputation

Existing or expected laws

Social justice and equity

Client demands

Media pressure

Differentiators

Business opportunities

Legal issues

Impact of the EU AI Act on companies

Obligations for providers and users

- Conformity assessment
- Certification
- Auditing
- AI systems' registration
- Transparency

For high-risk AI systems

- Trustworthy AI
- Human oversight

AI Regulation: a private- public partnership

Governments

- Hard laws, incentives, certifications, students' education

AI corporations

- Internal policies and audits, processes, education, tools

Standard bodies

- National and international

Multi-stakeholder organizations

- Stakeholder identification and convening, awareness, inclusion

AI Ethics at IBM

Principles for Trust and Transparency

- AI should augment human intelligence
- Data belongs to the creator
- AI should be transparent and explainable

Pillars of Trustworthy AI

- Fairness, Explainability, Transparency, Robustness, Privacy

Governance

- AI Ethics board, focal points, advocacy network

Operationalization

- Tools, education, offerings' evaluation, glossary

Partnerships

- Partnership on AI, Rome call, IBM-Notre Dame tech ethics lab, World Economic Forum, Global Partnership on AI

AI Ethics in a company: lessons learnt

Not just a person or a team

- Company-wide approach

Not just a discussion body

- A governance body with the power to make decisions for the company

Not (just) external advisors

- Multi-stakeholder partnerships

Not just principles

- Full operationalization

Technical tools are not enough

- Processes, education, guidelines

Don't wait for regulations

- Anticipate them, internally and for clients

AI ethics: beyond compliance to regulation

Thank you!

To know more about IBM's approach to AI Ethics and trustworthy AI, visit <https://ibm.biz/ibmaiethics>

