Artificial Intelligence: an American Perspective

Francesca Rossi

IBM Research

IBM AI Ethics Global Leader

AAAI President-elect
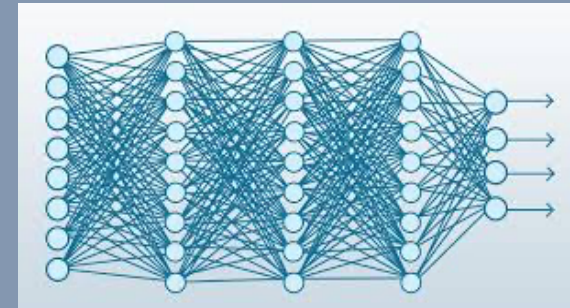
AI research: Europe vs USA vs Asia

- Europe – rules, less funds
  - Traditionally more on theoretical/rigorous/symbolic side
  - Applications mostly in the B2B space
- USA – funds, innovation
  - More applied/experimental/data driven side of AI
  - B2C and B2B business models
- Asia – data, funds
  - Mostly applied
  - More and more research, mostly data-driven
  - B2C

IBM

THE FUTURE OF/WITH AI

General
- Able to solve many different problems and adapt to new environments

Trustworthy
- Reliable, robust, explainable, aware of its bias, compliant with ethics principles
- Also in presence of uncertainty and disruptive events

Collaborative
- Fosters effective human-machine teaming
- Especially in complex scenarios with high-stake decisions
- Cooperation and autonomy

Sustainable and computationally scalable
- Energy and data efficient
- Real-time decisions

IBM

Current AI:
capabilities, limitations, ethical issues

Capabilities

Machine Learning

- Learning from data (Deep, Reinforced, Supervised/Unsupervised/Self Supervised)
- Hidden patterns in huge amounts of data
  - Prediction, perception tasks
  - Correlation, pattern discovery, data mining
- Flexible, can handle uncertainty

Rule-based, symbolic, and logical approaches

- Explicit procedure to solve a problem
- Reasoning, planning, scheduling, optimization for complex problems
- Symbolic, traceable, explainable

Limitations

- Generalizability and Abstraction
- Robustness and Resiliency
- Contextual awareness
- Multi-agent cooperation
- Resource efficiency (examples, energy, computing power)
- Adaptability
- Causality

AI ethics issues

- Trust
  - Fairness, robustness, explainability, causality, transparency
- Data governance, privacy, liability, human agency, impact on work and society
- AI autonomy vs augmented intelligence
- Real vs online life, metrics of success/goals

IBM

Neuro-symbolic AI
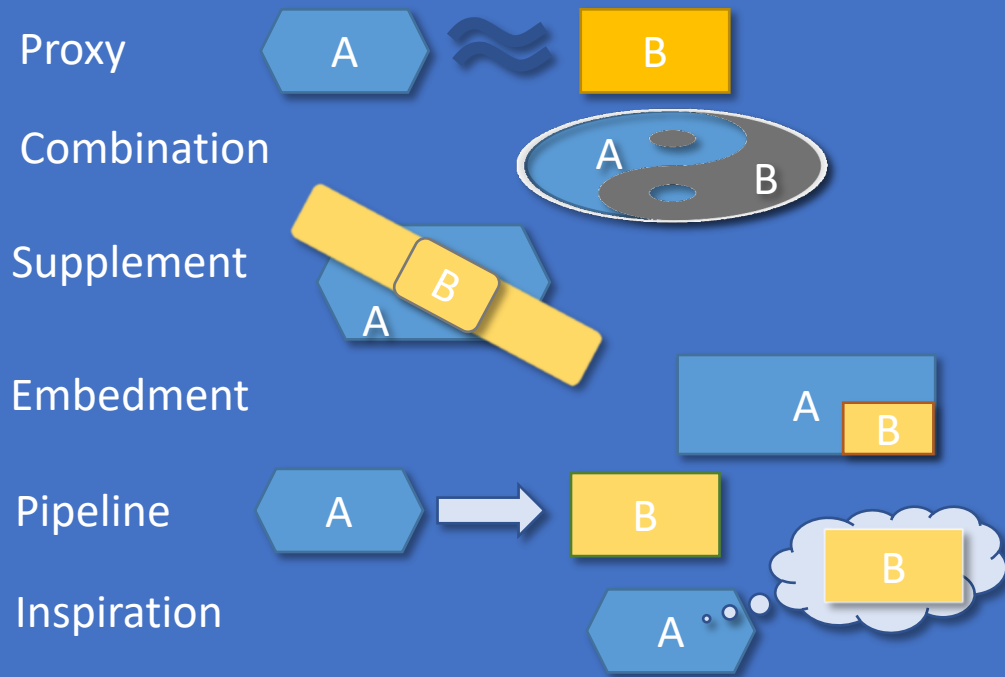
- Beyond predictions and correlation
  - Effective embedding into human decision making processes
- From data to knowledge
- Data "Understanding"

AI ethics

Multi-disciplinary and multi-stakeholder approach

- Technical solutions: algorithms, toolkits, libraries
- Non-technical solutions: Guidelines, best practices, standards, incentives, funds, audit, certifications, policies, laws, governance, education

AAAI / ACM conference on
ARTIFICIAL INTELLIGENCE, ETHICS, AND SOCIETY

**ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)**

- AI needs data
  - Data privacy and governance
- AI is often a black box
  - Explainability and transparency
- AI can make autonomous decisions
  - Fairness and value alignment
- AI is based on statistics and has always a small percentage of error
  - Accountability
- AI can infer our preferences and manipulate them
  - Human and moral agency
- AI is very pervasive
  - Larger negative impacts for tech misuse
  - Impact on jobs and society

Human-AI teaming and Explainable AI

- AI should support humans in making better decisions
- Effective teams need trust and collaboration
  - Humans should be trained too
- Standards, certifications, evaluation protocols
  - Not just for the AI system, but also for the human and the human-AI team
- Explainability is essential for building an effective team
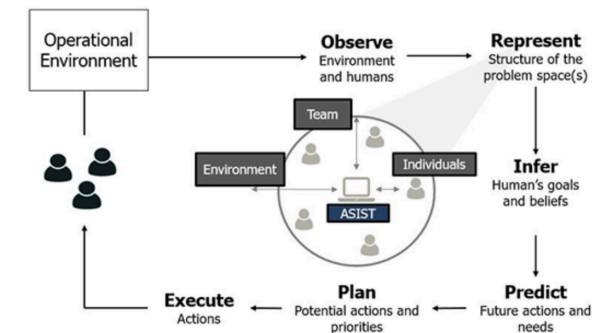
Explainable Artificial Intelligence (XAI)

Dr. Matt Turek
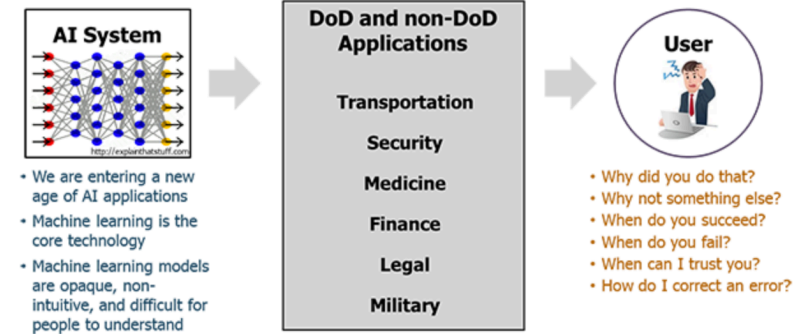
Using AI to Build Better Human-Machine Teams

Program seeks to demonstrate the basic machine social skills needed to generate effective human-machine collaborations

OUTREACH@DARPA.MIL
3/21/2019

IBM AI Research pillars

Neuro-symbolic AI — Machine learning combined with knowledge reasoning

Secure and Trusted AI — Fairness, explainability, robustness, transparency

AI engineering — Tools to simplify and automate key tasks in the AI pipeline
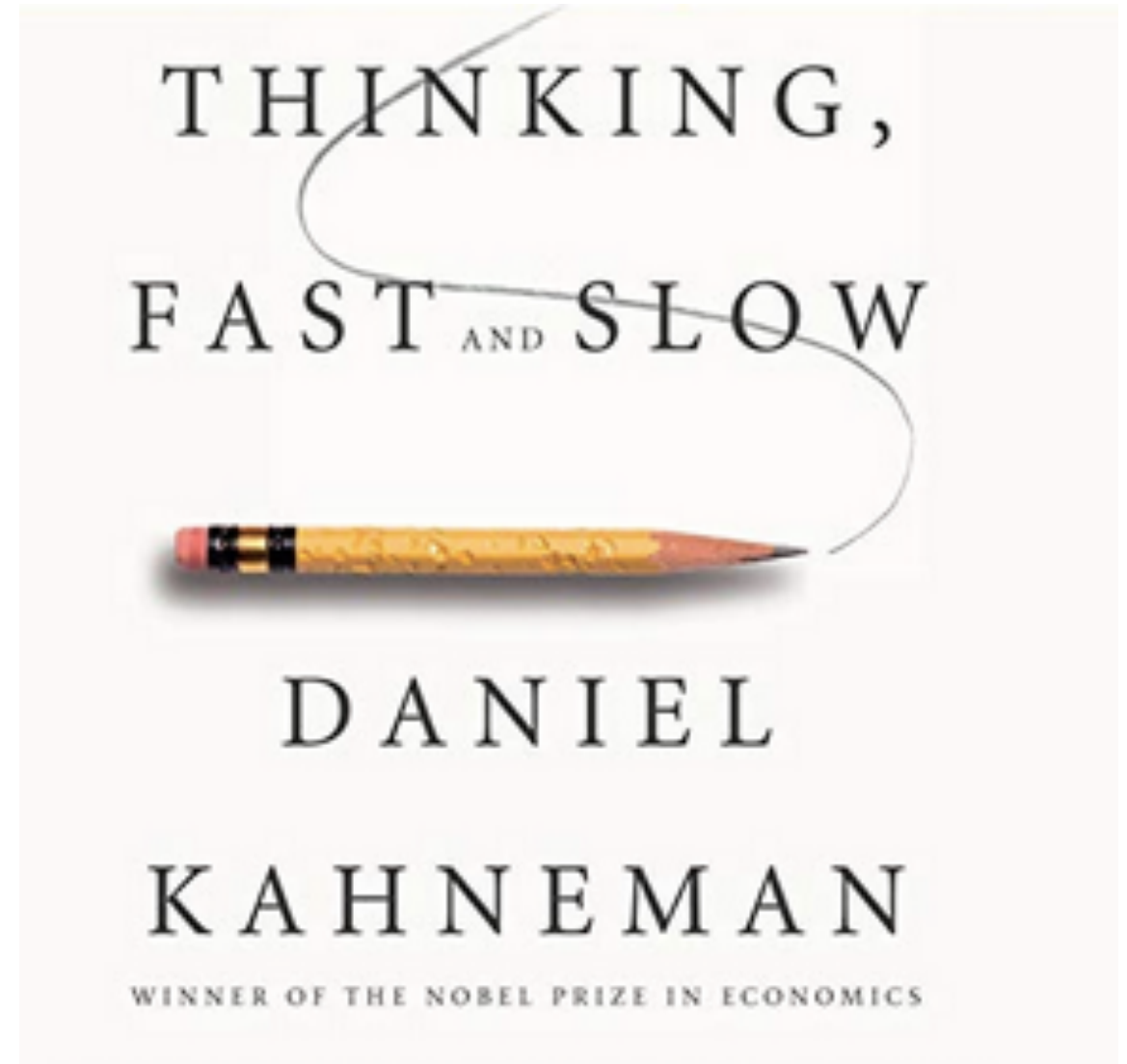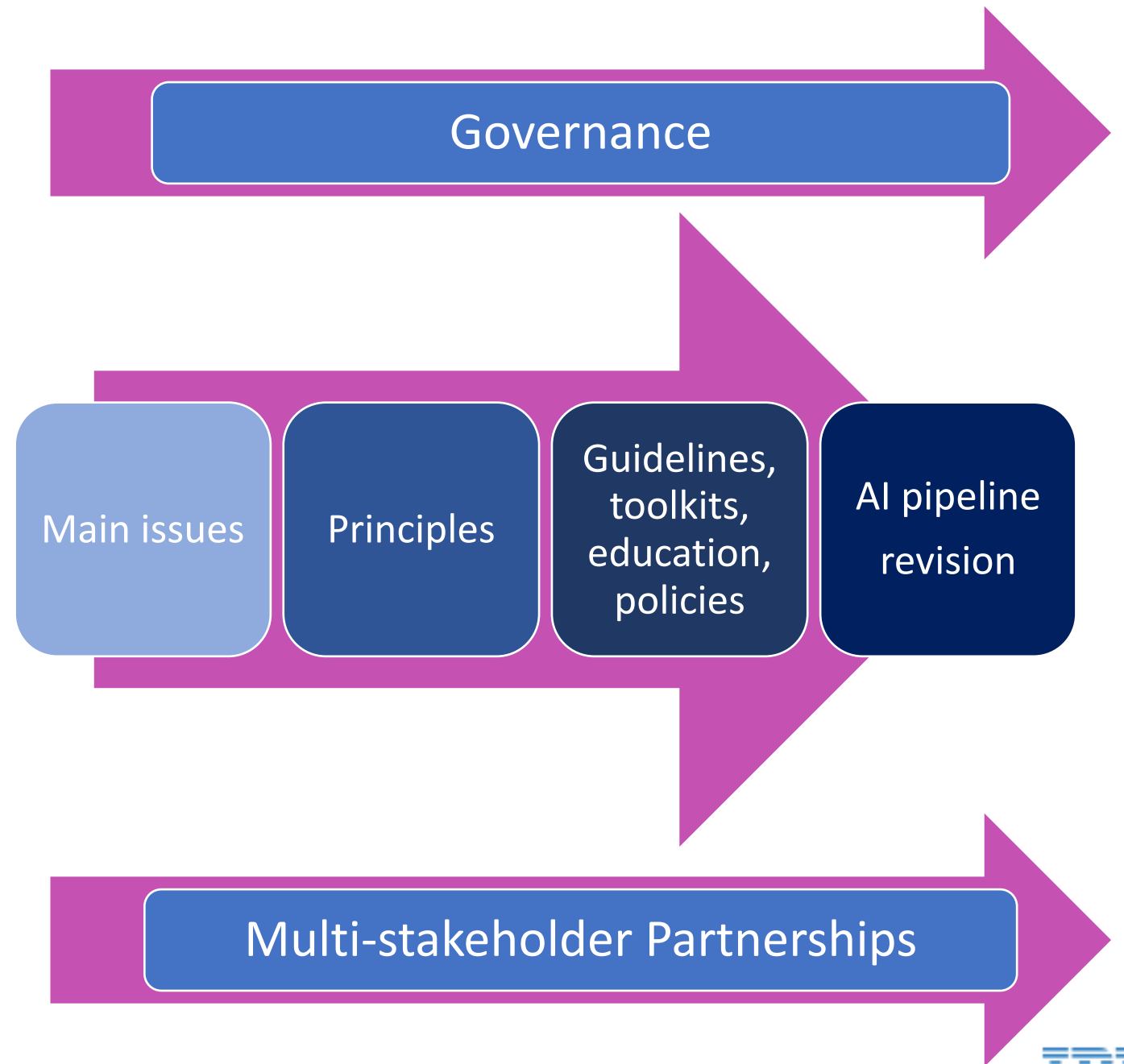
AI hardware — Energy-efficient hardware, quantum computing

A multi-disciplinary approach towards more general AI

- Exploit what we know about the human mind
  - Causes of human behavior
- How can humans adapt easily?
  - Abstraction, model building, generalization, explicit knowledge
  - Inject these capabilities in machines
- Beyond neuro-symbolic narrow AI
- IBM Exploratory Science program

THINKING, FAST AND SLOW

DANIEL KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS

IBM

From AI ethics principles to practice

Governance

Main issues → Principles → Guidelines, toolkits, education, policies → AI pipeline revision

Multi-stakeholder Partnerships

Internet days 2020

IBM
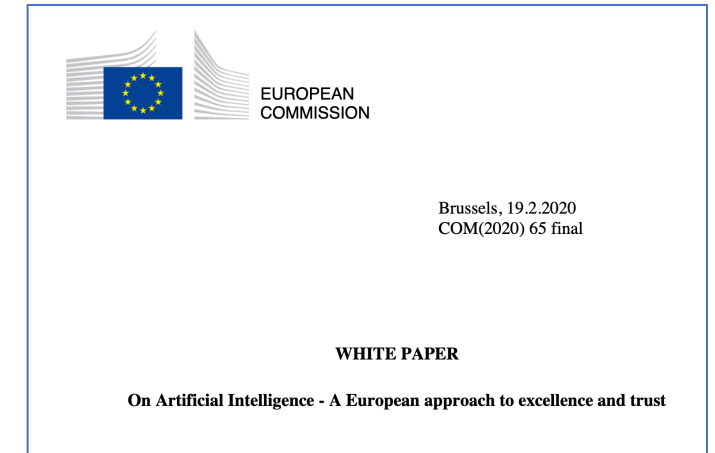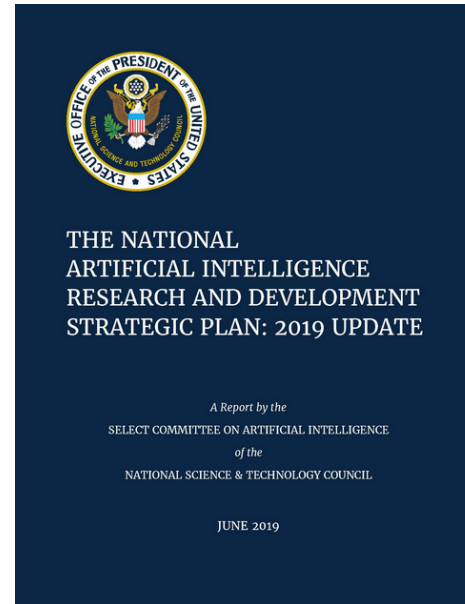
Regulating AI applications: EU vs USA

- Risk-based approach
- Focus on uses, not the technology
- Definition of high-risk
  - Sector-based or more fine-grained
- EU: future regulation only for high-risk uses of AI
- Self-assessment, standards, labelling for uses with less risk

THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN: 2019 UPDATE

A Report by the
SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE
of the
NATIONAL SCIENCE & TECHNOLOGY COUNCIL

JUNE 2019

EUROPEAN COMMISSION

Brussels, 19.2.2020
COM(2020) 65 final

WHITE PAPER

On Artificial Intelligence - A European approach to excellence and trust

IBM Policy Lab

Precision Regulation for Artificial Intelligence

IBM

Thanks!

IBM