Machines that decide, humans that obey?

Magdalena Ortiz
Knowledge-Based Systems Group
Institute of Logic and Computation

October 29th, 2019
Data and Learning

- Data and computing power almost unbounded
- Better techniques in data science and machine learning
- Readily available tools

Learning from data easier and more powerful than ever
AI in Decision Making

Patterns discovered in data are used to make predictions

- *Cheaper, more efficient, and more reproducible* than human decision making
- At the core of *business intelligence*
- Already reaching all aspects of our lives:
  - employment, credit, healthcare, policing, justice, education
  - even the movies we watch, where we go on holidays…
Data alchemy?

Algorithms can find patterns that humans cannot

Promises of turning data into gold
Science, not magic

- Optimism becoming a blind spot
- Scrutiny and scientific rigour falling short

*AI is getting much more trust than it has earned*
Who gets healthcare?

- In Arkansas, new algorithm drastically cut care for hundreds e.g., woman living alone, cerebral palsy: from 56 to 32 hours
- No one could explain the decision
- Eventually taken down after a legal battle

Pitfalls of Artificial Intelligence Decisionmaking Highlighted In Idaho ACLU Case

By Jay Stanley, Senior Policy Analyst, ACLU Speech, Privacy, and Technology Project

JUNE 2, 2017 | 1:30 PM

Who gets bail / parole?

COMPAS software to assess potential recidivism, used in US courts

Who gets the job?

Austria’s employment agency rolls out discriminatory algorithm, sees no problem
Learning in Healthcare

*Predict risk in case of pneumonia symptoms to support hospitalisation decisions*

Caruana et al., 2015

The system learned that **asthma causes lower risk**

*Correlation is not causation!*

*AI can only learn correlation*
Much more A than I

- No understanding
- No need to fear superintelligence yet
- **Autonomy** must be strictly **limited**

If AI sees something unusual, it’s probably not super-powers…

*AI is stupid*
What can we do?

Our job: better AI

- beyond the current paradigms
  - knowledge, symbolic manipulation
  - common-sense ...

- trustworthiness, transparency, fairness
What can you do?

- Scrutinise, question, transparency
- Awareness, spec. personal data
- Demand regulation