

#### ALMA MATER STUDIORUM Università di Bologna

# Al nella Giustizia: esperienze e problemi

## **Giuseppe Contissa** Ricercatore in informatica giuridica

#### **Predictive Justice**

 Predictive policing (e.g. Predpol, Xlaw, Keycrime etc.

 Individual Risk Assessment (suspect, accused, sentenced; e.g. COMPAS, HART, etc.)





### **Predictive Justice**

## Judicial decision modeling

technologies, targeting

court decisions

 other aspects of the litigation (time, costs, judge's or party's behaviours, etc) PeerJ Computer Science
Predicting judicial decisions of the
European Court of Human Rights: a
Natural Language Processing perspective



2018 numbers are year-to-date. Open dots are full-year estimates

Cases by Type

Trademark

Cases

### Main benefits and issues

- Consistency, quality in decision-making
- efficiency of the judicial system
- Issues related to:
  - Fairness / Non-Discrimination
  - Transparency / Explainability
  - Decision-Making Authority
  - Protection of fundamental rights



### Fairness / Non-Discrimination

"...decisions made by computers **may enjoy an undeserved assumption of fairness or objectivity.** However, the design and implementation of automated decision systems can be vulnerable to a variety of problems that can result in **systematically faulty and biased determinations**."

• Kroll et al. (2016) Accountable algorithms.

#### Many Causes:

- inadequate feature selection, biases embedded in the predictors
- Low quality or incompleteness of source data
- training set that reflect past prejudice or implicit bias, or that offer a statistically distorted picture of a set of events, of groups composing the population, etc.
- proxy discrimination

**BUT algorithms may also help correcting human cognitive biases, preventing unfair outcomes, "...**making it far easier to know whether discrimination has occurred [...] with the right safeguards in place, they have the potential to be a positive force for equity"

• (Kleinberg, Ludwig, Mullainathan, Sunstein 2018)



## Transparency / explainability

- ex-ante and ex-post transparency
- Limited explainability of the AI systems' functioning and outcomes

#### GDPR:

...such [automated] processing should be subject to suitable safeguards, which should include **specific information** to the data subject and the right to obtain human intervention, **to express his or her point of view**, **to obtain an explanation of the decision reached after such assessment and to challenge the decision.[..]** (GDPR, Recital 71)

[The controller shall provide the data subject with information concerning] the existence of automated decision-making, [..and...] **meaningful information about the logic involved**, as well as the significance and the envisaged consequences of such processing for the data subject. (GDPR, Art 13 & 15)



#### JUDICIAL DECISIONS ON TRANSPARENCY AND ACCESS TO ALGORITHMS

**Italy (2017)** – **Tar Lazio** (Administrative Regional Court of Lazio) allows access to the algorithm that manages the assignment of professors within Italian provinces.

**Italy (2018) - Tar Lazio** declares that an administrative process involving discretion cannot be fully delegated and adopted by an automated system

(Decision 9230/2018)

**Italy (2019-2020) – Consiglio di Stato**: principle of **transparency**. The algorithm must be known in all aspects... The "technical formula", which in fact represents the algorithm, need to be accompanied by explanations that translate it into the "legal rule" underlying it; principle of **non-exclusivity**; principle of **non-discrimination** 

(Decisions 2936/2019; 8472/2019; Decision 881/2020)

**France (2016)** – Decision of the Commission d'accès aux documents administratifs (CADA) on the access to source code of the Plateforme Admission post bac (A.P.B.)



DECISION-MAKING AUTHORITY in socio-technical systems



The model described (or prescribed) by the law is "under human control" / human responsibility, e.g.:

- **GDPR:** Right not to be subject to **(fully)** automated individual decisionmaking (Art 22 GDPR): "[oversight of the decision] should be carried out by someone who has the authority and competence to change the decision" (Art29WP)
- Aviation: ICAO Annex 2, sec. 2.3.1 Responsibility of pilot-in-command (ultimate responsibility)
- Vienna Convention on Road Traffic, Art. 1(v) "Driver" means any person who drives a motor vehicle or other vehicle (but recent amendment of art 8(5) for ADS)



## **EFFECTIVE DECISION-MAKING AUTHORITY ?**

What about decisions to be taken jointly with AI, in conditions of limited resources – time, information, explanations? Examples:

- **Medical diagnosis** assisted by Al (Lagioia, Contissa 2020)
- Frontex border controls: «12 seconds to decide»



Machine intelligence is fundamentally **alien**, and often, the entire purpose of an Al system is to learn to do or see things in ways humans cannot[..]

Ultimately, the lack of a principled basis to contradict AI predictions implies that the reasonableness of an action in individual cases must be tied to the decision to use AI as a general matter. (Selbst 2019)

Owing to the **evidence** in their favor (stipulated by definition), it is more appropriate to think of **expert robots as above average in their ability to make decisions that will produce desirable outcomes [...]** 

This fact suggests that granting a general decision-making authority to human experts will be problematic once expert robots are properly on the scene. (Millar, Kerr 2018)



#### COE: European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems

**2018, CEPEJ**, European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment ("**European Ethical Charter**")

- 1. Respect for fundamental rights during the design and implementation of AI,
- 2. Non discrimination;
- 3. Quality and security when processing judicial decisions and data;
- 4. Transparency, impartiality, and fairness;
- 5. "Under user control"

2019: Certification system for JDM tools (COE +IEEE)

#### COUNCIL OF EUROPE



#### CONSEIL DE L'EUROPE



### Implementation of the CEPEJ AI Charter As Applied to Judicial Decision Modeling Technologies

**Step 1:** analysis of each of the Charter Principles into its component elements;

**Step 2:** formulation of questions pertinent to assessing adherence to each of the elements identified in Step 1; and

**Step 3:** identification of the specific empirical evidence that can serve as the basis for answering each of the questions formulated in Step 2.

#### Perspectives:

- Lifecycle of AI application
- Agents engaged in development /operation of AI within the STS
- «Informed Trust» conditions: effectiveness, competence, accountability, transparency, data agency & security

