

**TEAM ITALY 2 – THEMIS 2019**  
**SEMIFINAL D**  
Judicial Ethics and Professional Conduct

TITLE  
**THE JUDGE OF THE FUTURE: ARTIFICIAL INTELLIGENCE  
AND JUSTICE**



**MEMBERS**  
Camilla OVI  
Luigi SALVIA  
Pierluigi TONNARA

**TRAINER**  
Francesco PERRONE

# Table of Contents

## Chapter 1: Artificial intelligence

- 1.1 Aspiration of predictable justice
- 1.2 Artificial intelligence
- 1.3 Artificial intelligence: some applications to justice
- 1.4 Expected benefits

## Chapter 2: Human intelligence

- 2.1 How does the judge decide?
  - 2.1.1 *The juridical realism*
  - 2.1.2 *The mechanism by which judges make their decisions*
  - 2.1.3 *The emotional judge: the role of the emotions in judicial decision-making*
- 2.2 The distinctive element of human intelligence: the balancing test
  - 2.2.1 *The ECHR case law*
- 2.3 The limits of A.I. compared to H.I.

## Chapter 3: Dealing with Ethics and AI in the justice field

- 3.1 The Ethical charter on the Use of Artificial Intelligence in Judicial Systems
- 3.2 Perspectives and concerns about using AI tool
- 3.3 Which risks arise from a robot-judge?
- 3.4 European Commission approach on implementing AI
- 3.5 To the future and beyond

\*\*\*

## Chapter 1. The artificial intelligence

### 1.1 Aspiration of predictable justice

*As we wait for a decision in a trial, do we aspire to a predictable justice more than to a correct decision?*

It is a general idea, which was long-established, that certainty, predictability, rationality, coherence, uniformity must be essential characteristics of law and judicial decisions.

Generally, discretion is perceived as a natural element of law-making, but it is not much-appreciated vis-à-vis the decision-making process.

This approach to justice is rooted in French Revolution cultural tradition, according to which judges have to be the “*bouches de la loi*”: they should apply legal provisions, taking them far from the creation of law through interpretation. In other terms, once the community, represented by an elected

assembly, set the legal order, the judge's role is confined to declaring the meaning of the law and, therefore, the will of its community in the specific case. According to this view, the equality and the freedom of people in front of legal order require subjectivism to be replaced by the "mechanical" application of rules that cannot be interpreted. This leads to the aspired certainty. In sum, to pass from disorder to predictability, one needs to aim to a new model of decision-maker "without a heart" and cognitive subjectivism.

The idea of decision making as a mechanical application of rules was further developed and found in capitalist society its most fertile soil. Max Weber pointed out that the modern economy and society pretend legal order to be rational, calculable and predictable "like a machine"<sup>1</sup>.

Nowadays, is this just a bizarre suggestion or anything which both contemporary society and political institutions tend towards? Any robot-judge on the horizon?

To this regard, artificial intelligence ("AI") seems to play the key role.

## 1.2. Artificial intelligence

When referring to AI, the mind goes to the ability of a machine to replicate human intelligence. Actually, intelligence is concerned mainly with rational action. Ideally, an intelligent agent is expected to take the best possible response in a given situation.

Philosophers<sup>2</sup> considered that the human brain is in some ways comparable to a machine, which operates on knowledge encoded in some internal language, and that thought can be used to select what actions to take. This idea brings to mind the Greek philosopher Aristotle, who was one of the first to attempt to codify "right way of thinking", that is the universal structure of the correct way of reasoning. Each type of syllogism guarantees that, when given certain premises, one always reaches correct conclusions.

Moreover, cognitive psychologists<sup>3</sup> adopted the idea that both humans and animals can be considered as information processing machines. Linguists<sup>4</sup> showed that language use fits into this model. In

---

<sup>1</sup> M. Weber, *Economy and Society* (G. Roth & R. Wittich ed. 1968).

<sup>2</sup> Aristotle, 384–322 B.C., syllogism; Ramon Lull, d. 1315, idea of mechanical artifact; Thomas Hobbes, 1588–1679, artificial animal: "For what is the heart but a spring; and the nerves, but so many strings; and the joints, but so many wheels"; Blaise Pascal, 1623–1662, "the arithmetical machine produces effects which appear nearer to thought than all the actions of animals", Gottfried Wilhelm Leibniz (1646–1716); René Descartes, 1596–1650, rationalism; David Hume's, 1711–1776, principle of induction; Vienna Circle (Ludwig Wittgenstein, Bertrand Russell), doctrine of logical positivism.

<sup>3</sup> William James, 1842–1910. Stuart J. Russell and Peter Norvig, *Artificial Intelligence. A Modern Approach*, Third Edition, Prentice Hall, 2010, p. 13 refer that "in the United States, the development of computer modeling led to the creation of the field of cognitive science. The field can be said to have started at a workshop in September 1956 at MIT. (We shall see that this is just two months after the conference at which AI itself was "born.") At the workshop, George Miller presented *The Magic Number Seven*, Noam Chomsky presented *Three Models of Language*, and Allen Newell and Herbert Simon presented *The Logic Theory Machine*. These three influential papers showed how computer models could be used to address the psychology of memory, language, and logical thinking, respectively. It is now a common (although far from universal) view among psychologists that "a cognitive theory should be like a computer program" (Anderson, 1980); that is, it should describe a detailed information processing mechanism whereby some cognitive function might be implemented".

<sup>4</sup> B. F. Skinner, 1904–1990; Noam Chomsky, 1928.

particular, their suggestion is that language with its syntactic structure is formal enough that it could be programmed.

Further, a significant boost to AI was given by mathematics<sup>5</sup>, which provided the tools to manipulate statements of logical certainty as well as uncertain, probabilistic statements. In other terms, once pointed out some fundamental ideas on AI, the leap to a formal science required a level of mathematical formalisation in three key areas: logic, computation, and probability<sup>6</sup>.

In this regard, algorithm<sup>7</sup> plays an important role. This concept comes from computer science domain, rather than the field of law, and refers to a set of mathematical instructions that must be followed in a fixed order so to pass from the given premises to the correct and straightforward output.

Which kind of connection may be found between the functioning of human intelligence and the algorithm's usage?

The assumption is that reality is an aggregation of data, which can be represented through mathematical language; reasoning means establishing a net of relations amongst data and, therefore, it can be synthesised in numerical formulas. Such formulas permit that, where assumed given data, a straightforward and unequivocal output results.

Engineers were thus able to create machines having artificial neural networks (ANN), which are computing systems that are inspired by, but not necessarily identical to, the biological neural networks. Such networks "learn" to perform tasks by considering examples, generally without being programmed with any task-specific rules.

In a few words, AI's aim is to create, in the various human fields<sup>8</sup>, computers capable of: natural language processing, so to enable it to communicate successfully in a given language; knowledge representation, so to store data that it knows or hears; automated reasoning, so to resume the stored information to answer questions and to draw new conclusions; machine learning, so to adapt to new circumstances and to detect and extrapolate patterns. In particular, machine learning is an application of AI that provides systems with the ability of automatically learning and improving from experience without being explicitly programmed. It focuses on the development of computer programs that can access data and use it to learn for themselves.

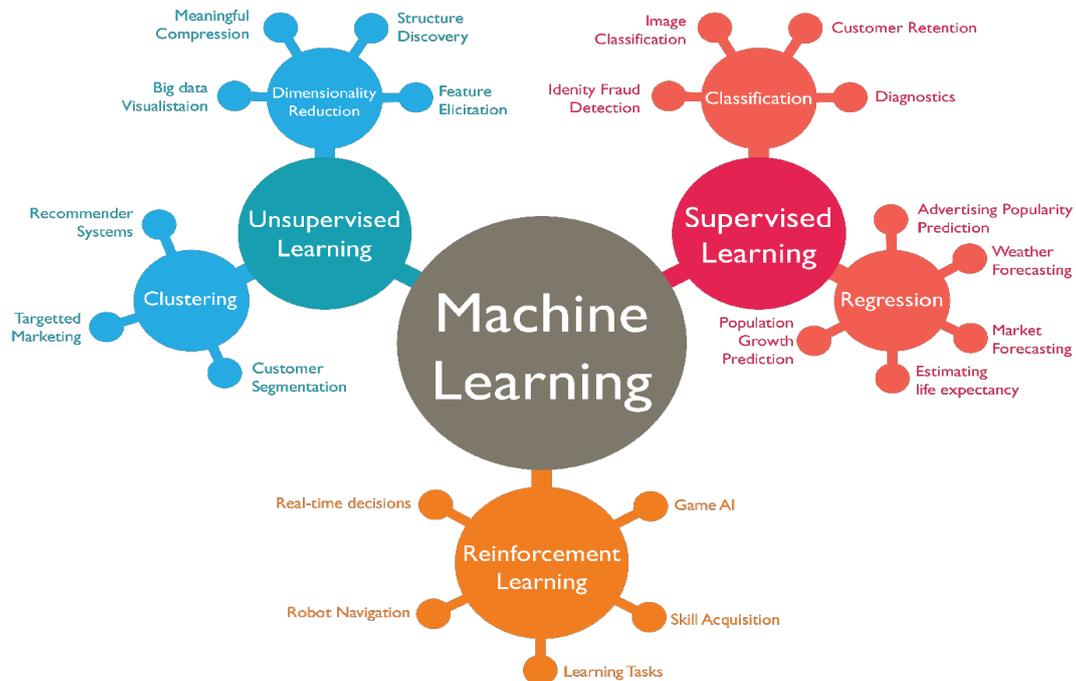
---

<sup>5</sup> George Boole, 1815–1864; Gottlob Frege, 1848–1925; Alfred Tarski, 1902–1983

<sup>6</sup> Stuart J. Russell and Peter Norvig, *Artificial Intelligence. A Modern Approach*, Third Edition, Prentice Hall, 2010, p. 7.

<sup>7</sup> The word algorithm comes from al-Khowarazmi, a Persian mathematician of the 9th century, whose writings also introduced Arabic numerals and algebra to Europe.

<sup>8</sup> We can daily experience the potentialities of AI applications in the various domains of our lives. Just to make some examples, AI is involved in agriculture (crop and soil monitoring, agricultural robots), finance (market analysis and data mining, personal finance, portfolio management), medicine (electronic medical record - EMR software, computer-aided interpretation of medical images, heart sound analysis, radiology), music (composing software named "Emily Howell" by David Cope), energy (smart grids), city planning (smart cities), transports (aviation, autonomous cars). AI may raise unprecedented problems to be solved. For instance, in 2018 an autonomous Uber car killed a woman in the street in Arizona. Can robot cars accurately predict human behavior? In other terms, problems seem to come when humans and robots interact.



*From [www.medium.com](http://www.medium.com)*

### 1.3. Artificial intelligence: some applications to justice

AI has also made its way into the legal world and further developments are certainly yet to come.

The extensive use of AI technologies is currently implemented in the US criminal trial, where predictive algorithms are applied to access to the recidivism risk, i.e. the likelihood of a sentenced individual to commit another crime based on an evaluation of actuarial data. In this respect, in the case *State v. Loomis*<sup>9</sup>, the Supreme Court of Wisconsin held that usage in a court trial of an algorithm for the recidivism risk assessment did not violate the defendant’s due process rights to a fair trial. Remarkably neither the court nor the defendants were really aware of the algorithm structure.

Furthermore, many prototypes have been testing.

In Germany, the project “argumentum”<sup>10</sup>: judicial argumentation requires sophisticated intellectual effort and the knowledge of many information; given that the fulfilment of this task finds a limit by the natural human information processing capacity, a software prototype was developed to support automated identification, analysis and selection of legal arguments structures.

In Russia, engineers tried to pattern a “Robot Lawyer” intellectual system<sup>11</sup>: the main goal is to assist lawyers and citizens in providing the necessary information regarding legal processes; “Robot Lawyer” includes an expert system that uses a set of rules to provide reference information and some neural network models to answer more complex questions.

In Brazil, scientists are studying a legal Information Retrieval system precision<sup>12</sup>: such a system is based on a model that uses artificial intelligence technique known as Case-Based Reasoning (CBR);

<sup>9</sup> 881 N.W.2d 749, Wis. 2016.

<sup>10</sup> More info to: <https://ieeexplore.ieee.org/document/6611332>.

<sup>11</sup> More info to: <https://ieeexplore.ieee.org/document/8674441>.

<sup>12</sup> More info to: <https://ieeexplore.ieee.org/document/5532453>.

according to the fundamental logic of this model, previous legal cases can be useful to provide the rule to solve a current problem, to the extent a certain degree of similarity is possible to be found amongst the involved cases.

In Italy, there is ongoing research aimed at defining theoretical and technological components enabling a citizen to obtain guidance and training on legal concepts starting from a textual description of a case<sup>13</sup>. The system is able to detect relevant legal concepts from the textual description also relying on an ontology and on the enrichment of the case text with common-sense knowledge. Detected concepts are used to generate a training path aimed at providing citizens with the basis for understanding legal issues the case deals with. The training path is then enriched with legal information like relevant laws and jurisprudence retrieved on an external legal repository.

#### **1.4. Expected benefits**

Predictive algorithms are expected to rationalise the decision-making process by summarising all relevant information in a more efficient way than the human brain is able to do. This should have an impact also in terms of saving time.

But the main argument in favour of implementing AI in the field of justice is the ability of algorithms to reduce discretion.

Quantification helps to hold judges and prosecutors more accountable for their decisions. The usage of algorithms and “smart statistics” appear to be an easy solution for making sentencing more consistent and efficient. A quantitative assessment provided by a software program is supposed to be perceived as more reliable, scientific, and legitimate than other sources of information, including one’s emotionality about the case.

## **Chapter 2: Human intelligence**

### **2.1 How does the judge decide?**

#### ***2.1.1 The juridical realism***

This is a big matter that has been questioned for a long time. Is the judge’s decision the result of a mere rational process or does it also depend on other factors? Jerome Frank<sup>14</sup>, an American judge exponent of the juridical realism, stated that the decision is not predictable because it is the result of intuition and not of reasoning and the judge comes to a decision before trying to motivate it. Realists argue that the rational application of legal reasons does not sufficiently explain judicial decisions and that psychological, political, and social factors influence rulings as well. The realist view is commonly caricaturized by the trope that justice is “*what the judge ate for breakfast*”. A study conducted by

---

<sup>13</sup> More info to: <https://ieeexplore.ieee.org/document/7057139>.

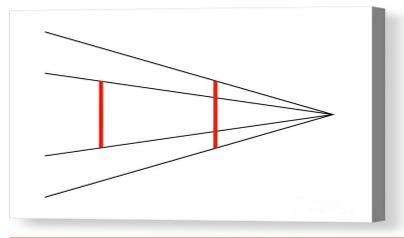
<sup>14</sup> J. Frank, *Law and the Modern Mind*, New York, Coward-McCann, 1949.

Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso<sup>15</sup>, has shown that “*the likelihood of a favourable ruling is greater at the very beginning of the workday or after a food break than later in the sequence of cases. ... the likelihood of a ruling in favour of a prisoner spikes at the beginning of each session — the probability of a favourable ruling steadily declines from  $\approx 0.65$  to nearly zero and jumps back up to  $\approx 0.65$  after a break for a meal*”. The reason for these results is that when judges make repeated rulings, they show an increased tendency to rule in favour of the status quo. A short break can help to overcome mental fatigue.

Is it true? Is the judicial system completely arbitrary?

The juridical realism focused for the first time on the judge and not on the rules of law. Every human being has bounded rationality especially in conditions of particular uncertainty and stress, like the decisions taken in Court. Often psychologists talk of cognitive illusions that are assumed to be interactions based on assumptions about the world, which lead to unconscious inferences. When people have to deal with complex problems, make use of heuristics, that are shortcuts of thinking, but often these tools are less reliable.

A typical example is the Ponzo Illusion: our mind is brought to see the two segments of a different length unless they are exactly alike.



### 2.1.2 The mechanism by which judges make their decisions

Is it the judge affected by these distortive mechanisms of the mind?

The first thing we have to understand is how the human mind thinks and reasons, in particular the judge when he takes a decision. The study of logic, attributable to Aristotle and subsequently developed by other philosophers, divided the reasoning into three types of inferences: induction; abduction; deduction.

#### ***deduction***

<u>MAJOR PREMISE</u>	<u>MINOR PREMISE</u>	<u>CONCLUSION</u>
<u>case</u>	<u>rule</u>	<u>result</u>
Socrates is a man	All men are mortal	Socrates is a mortal

<sup>15</sup> Shai Danziger, Jonathan Levav and Liora Avnaim-Pessoa, *Extraneous factors in judicial decisions*, in [www.pnas.org/cgi/doi/10.1073/pnas.1018033108](http://www.pnas.org/cgi/doi/10.1073/pnas.1018033108)

***abduction***

<u>MAJOR PREMISE</u> <u>result</u>	<u>MINOR PREMISE</u> <u>rule</u>	<u>CONCLUSION</u> <u>case</u>
Socrates is mortal	All men are mortal	Socrates is probably a man

***induction***

<u>MAJOR PREMISE</u> <u>case</u>	<u>MINOR PREMISE</u> <u>result</u>	<u>CONCLUSION</u> <u>rule</u>
Socrates is a man	Socrates is mortal	All men are mortal

Civil judges and criminal judges, using presumptions and clues to prove the fact, give rise to abductive reasoning, in which certain assumptions lead to information not contained in them. The model of the judicial reasoning is the abduction because “judgment in fact”, having historical nature, operates backwards.

Most of the cognitivists state that the reasoning receives the contribution of two separate cognitive systems, connected to each other. These two are the intuitive system and the analytic system. The first one is a universal form of cognition shared by human beings and animals; the second one is used only by human beings and is not based on instinct, but it’s characterised by a slow and controlled process, and it is based on rules. These two minds at the service of the reasoning also assist the judge. The intuitive system can support the detection of the category that this event belongs to and the activation of the associated procedure. The analytic system will intervene to modulate the procedure only when the situation so requires. However often the intuitive system, not properly guarded by the analytic system, makes the judge fall into a fallacy. Is it possible to improve the reasoning process?

Koehler, Brenner e Griffin<sup>16</sup> asked if lawyers, public prosecutors and judges are less vulnerable to confirmation bias than non-experts. In the opinion of the three authors the answer depends on how extensive their experience is. Experience can help to reason more neutrally and can suggest generating alternative ideas compared to the hypothesis taken initially in consideration. The Italian criminal procedure code (art. 358) requires public prosecutors during the investigations to look for evidence also in favour of the suspect, but often it doesn’t happen because they continue to seek clues or proof that corroborate their allegations in order to build a case, ignoring evidence that could bring to an acquittal. However, in a research conducted on judges with different grades of experience has been

---

<sup>16</sup> Koehler, D.J., Brenner, L. e Griffin, D., *The calibration of expert judgment: Heuristics and biases beyond the laboratory*, in Gilovich, Griffin e Kahneman, 2002, pp. 686-715.

observed that the more expert judges during the trial took into consideration more options about the course of events than the less experts<sup>17</sup>.

And what about deductive reasoning? Are the judges able to apply this kind of reasoning to the trial? In the opinion of Johnson-Laird and Byrne<sup>18</sup> deductive reasoning can be defined as a particular way of reasoning from one or more statements (premises) to reach a logically certain conclusion. If all premises are true, then the conclusion reached is necessarily true. The deductive logic is used to make a rule of law: if A then B, if A really happened then B is the true conclusion. If the prepositions of law are true, the legal system can be built by applying the logic to the prepositions of law already known in order to achieve others. The legislator has to establish only the general principles, then the intermediate rules can be obtained by the general rules. This is only a utopic purpose, because when people are reasoning they do not follow the schemes of deductive logic *tout court*. In syllogistic reasoning, the content of the premises is irrelevant, only the structure of the topic counts. Nevertheless, the content affects the performance of the subjects called upon to reason deductively, also the judges.

Definitively even if the legal system has been created through deductive inferences, when the judge decides ("judgment in law") the deduction cannot be rigorous as a rule described in the provision because the effective applicability of the rule of law to the case depends on the interpretation of the provision. In other words, while in the abstract dimension the rule of law is emphasized in the formal sense, that is, as a hypothetical connection between a fact and an effect, in the concrete dimension the rule of law is emphasized in the substantive sense, that is its effective applicability to the case, which depends on the interpretation.

Often the rules of law require judges a more complex mental activity than deductive reasoning, because they have to choose between more options, as when they have to assess evidence or decide the punishment or the amount of damages<sup>19</sup>. As we will see in the next paragraph, sometimes the judge must balance between different freedoms; the balancing test cannot be included in a perfect deduction because the judge has to evaluate too many factors. If the AI is able to reproduce deductive reasoning where the minor premise doesn't need to be interpreted and the relevant facts are straight, certainly cannot balance between different rights and select the critical aspects of the fact (included in

---

<sup>17</sup> Catellani, P., *Il giudice esperto*, Bologna, Il Mulino, 1992.

<sup>18</sup> Johnson-Laird and Byrne, *Deduction*, Hillsdale, NJ, Erlbaum, 1991.

<sup>19</sup> The heuristic can influence the decision taken by the judge: if the plaintiff asks for higher compensation or the public prosecutor asks for a higher penalty, higher will be the damage or the punishment determined by the judge, influenced by these requests (Englich, B. e Mussweiler, T., *Sentencing under uncertainty: Anchoring effects in the courtroom*, in «*Journal of Applied Social Psychology*», 2001, 31, pp. 1535-1551; Marti, M.W. e Wissler, R.L., *Be careful what you ask for: The effect of anchors on personal-injury damages awards*, in «*Journal of Experimental Psychology*», 2000, 6, pp. 91-103). This is called anchoring heuristic: initial exposure to a number serves as a reference point and influences subsequent judgments. Another example of fallacy that can affect the judge while he's assessing evidence is the conjunction fallacy, which occurs when it is assumed that specific conditions are more probable than a single general one. When the lawyer describes to the judge a more accurate story, the judge could deem more reliable this version than another one more general.

the major premise) that allow determining the applicable rule. Aristotle called perfect syllogism what does not need anything else, so the necessity of deduction may be revealed, and imperfect syllogism the one which requires the addition of one or several objects, objects that are necessary but that have not been assumed through the premises. This second logical operation cannot be reproduced by an AI tool, that works using binary logic. If we consider the balancing between different freedoms, it often depends on the particular circumstances of the case that are not reducible to a perfect deduction.

### ***2.1.3 The emotional judge: the role of the emotions in judicial decision-making***

The emotions have often be considered the arbitrary part of the decision that a machine could certainly eliminate. The neuroscientist Antonio Damasio, a neurologist at the University of Iowa College of Medicine, in his most famous work “*Descartes’ error*”<sup>20</sup> describes the case of Elliot, a patient operated for heavy damage at the prefrontal cortex, that is the part of the brain responsible for the decision making process and planning process. Due to the surgery, Elliot lost the capacity to plan a decisional process. Through his research on patients with prefrontal cortex damage, Damasio discovered that reason, like almost all mental processes, is “embodied,” that is, based on the human being’s physical self. Emotions and other states that are rooted in physicality profoundly influence not only what people reason about, but how they reason. Without them, people either can’t make decisions or they make self-defeating ones.

Definitively, in the opinion of the author, the emotions weren’t merely the passionate element<sup>21</sup> that contaminated the rational thinking but they were an essential component of reasoning and decision. The emotions were a not replaceable cognitive guide. Without emotions, a man can perfectly reason but he’s not able to decide.

Brady<sup>22</sup>, a scientist-philosopher, studied the investigation about the emotions using the instrument of neuroscience. The emotions help to pay attention not only to some details, but also to some experiences and sensations lived. They push people to find reasons about psychological conditions, the emotions motivate us to find reasons but themselves are not reasons. In this respect they are reflexives. When we reflect, emotions may extend the agent’s perspective, correct prejudices, refine the judgment. This ability of the man is essential also for the judge. Some kinds of decisions taken in particular fields as family law, not only ask the judge to face with non-legal arguments but require particular sensitivity, because they have to be accepted by the parties to be effective. A simple exercise of rationality wouldn’t be desirable because people are more willing to accept a decision if

---

<sup>20</sup> Damasio, A.R., *Descartes’ Error: Emotion, Reason, and the Human Brain*, New York, Grosset/Putnam, 1994.

<sup>21</sup> Emotions can also create distortion but in order to understand in which way they can affect the judgment it is necessary to distinguish between: immediate emotions; anticipated emotions and incidental emotions. Immediate emotions generate from the same decision process; anticipated emotions are the ones that you can foreshadow as a result of the decision taken; incidental emotions are the ones influenced by previous moods, not related with the decision. All these kind of emotions can influence the decision of the judge: the judge can be sad or nervous when he must decide; the jury that has to determine on the amount of the damage can be affected by cruel pictures shown by the lawyers; the judge can overestimate the effect of his decision on the parties.

<sup>22</sup> Brady M., *Emotional insight: The Epistemic Role of Emotional Experience*, Oxford, Oxford University press, 2014.

they think it's fair. Sometimes the decision has to be adapted to the parties in order to allow them to share the content of the decision itself. In this perspective, emotions can be positive for the decision making process to enable the judge to take a more conscious decision. This is a fundamental part of the judgment that deductive reasoning, reproduced by an AI tool, couldn't include.

## **2.2. The distinctive element of human intelligence: the balancing test**

Most of the rights that the European Convention on human rights grants are not absolute, but have to be balanced with other fundamental rights. If we examine articles 8, 9, 10, 11 the same rules of law provide for the possibility of limiting these freedoms, also if the limitations have to be prescribed by law and necessary in a democratic society for saving specific interests.

Often the European Court faces with the assessment on whether the restrictive measure pursues a legitimate aim and whether it is proportionate and necessary to achieve that aim. The application of this balancing test seeks to strike a fair balance between concurrent interests at stake.

The European Court applied the balancing test in different fields. This kind of judgment implies the use of discretion because the rule of law doesn't provide when the restriction has to be deemed necessary compared with other rights. The value judgment is part of this test.

How can a machine balance between different rights and decide if the restriction of freedom is justified or not?

### **2.2.1 The ECHR and national Supreme Courts case law**

#### *Freedom of religion*

The most emblematic cases that show the importance of the balancing test - and the impossibility of the A.I. to emulate that pattern - are the judgments *Eweida* (ECHR), *Samira Achbita* (ECJ) and *Asma Bougnaoui* (ECJ). The European Courts were called upon to assess whether a blanket prohibition for employees to visibly wearing signs of religious beliefs is proportional and necessary for the aim to pursue a policy of corporate neutrality; in doing this, both the ECHR and ECJ seek to strike a fair balance between concurrent interests at stake.

In particular, in the case of *Eweida and others v. The United Kingdom* (2013), the Court had to establish whether some limitations to the freedom of religion (to wear a small cross on a chain visibly around the neck; belief that homosexual activity/relationships cannot be condoned) were justified in the light of the principles of reasonableness and proportionality. The Court assessed, on the one hand, the legitimate interest of the employer to pursue a policy of neutrality towards its customers by imposing a blanket prohibition to visibly wearing signs of religious beliefs. Then, on the other hand, the right of the employee not to be discriminated on the grounds of religious belief, because of its choice to wear an Islamic headscarf or a Christian cross at the workplace. The Court adopted a different decision depending on the specific circumstances of the four individual cases. In every judgment the Courts considered, case by case, whether the employee was required to interact with the

customers, whether the employer genuinely pursued the corporate policy consistently and systematically, whether the restriction to the rights of the employer was of crucial importance for the pursuit of the corporate policy. Anyways, every decision involved an assessment on whether the measure leading to the differential treatment pursues a legitimate aim and is proportionate and necessary to achieve that aim.

#### *Criminal cases: freedom of expression*

Freedom of expression constitutes one of the essential foundations of a democratic society and one of the primary conditions for its progress and self-fulfilment. Pursuant to article 10, § 2<sup>23</sup>, the restrictions of this freedom must be prescribed by law and necessary in a democratic society. The Court is therefore empowered to give the final ruling on whether a ‘restriction’ is reconcilable with freedom of expression as protected by Article 10 and to determine whether the interference was ‘proportionate to the legitimate aim pursued’ and whether the reasons adduced by the national authorities to justify it are ‘relevant and sufficient’. The nature and severity of the sanctions imposed are also factors to be taken into account when assessing the proportionality of the interference. In order to do this assessment it is necessary to balance the freedom of expression with the right to honour and privacy of people involved.

Although in many cases the Court<sup>24</sup> found that the domestic authorities were entitled to consider necessary to restrict the exercise of the applicant’s right to freedom of expression, the imposition of a custodial sentence for a media-related offence, albeit suspended, compatible with journalists’ freedom of expression as guaranteed by Article 10 of the Convention can only be in exceptional circumstances, notably where other fundamental rights have been seriously impaired, as, for example, in the case of hate speech or incitement to violence. Therefore the Court has to balance the freedom of expression with other rights, and once deemed the restriction necessary, it has to do another balancing test: to assess if the sanction imposed is proportional to the legitimate aim pursued.

#### *Housing rights*

An example of how the human reasoning can assess interest balancing values can be seen in the case law of the ECHR about housing rights (*Marzari v. Italy*, 4 May 1999) related to the specific way of life of nomad populations. The Court, who considers the housing right as guaranteed within Art. 8 and 18 of the Convention, recognised the positive obligation on the Member States to facilitate the gipsy

---

<sup>23</sup> Article 10 (freedom of expression) ECHR states that: “1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises. 2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary”.

<sup>24</sup> European Court of Human Rights, *Ricci v. Italy*, 8/10/2013; *Sallusti v. Italy*, 7/3/2019; *Belpietro v. Italy*, 24/9/2013; *Morice v. France*, 23/4/2015; *Riolo v. Italy*, 17/7/2008.

way of life (travelling around and settling in different places) as the vulnerable position of gypsies as a minority meant that some special consideration had to be given to their needs and their different lifestyle. This decision is clearly beyond the results of any mechanical application of the law, and it is the development of balancing between housing rights and property rights that also take into consideration the peculiar social and cultural context of the case submitted.

#### *Certainty of law and overruling*

Another example of balance test comes from the Allègre case (*Allègre v. France*, 12 July 2018) in which the ECHR found that different interpretation given by the Court and even Supreme Court's overruling are not in contrast with the right to a fair trial under the art. 6 of the Convention. In this case the ECHR recognised that jurisdictional differences are inherently the consequence of any judicial system that relies on a body of substantive jurisdiction, and that the right to a fair trial is not undermined if the national judicial system provides mechanism for the removal of inconsistency and profound divergences between different courts, if these mechanisms are efficiently applied (see even *Parish Greek Catholic Lupeni and others*, § 116, *Ferreira Santos Pardal v. Portugal*, 30 July 2015, § 42). In particular, the Court requires two conditions to avoid the breach of art. 6, which are the motivation of the overruling and the respect of the prevedibility principle.

Doing so, the ECHR admits that certainty of law is a fundamental right but that it should be balanced with the freedom and independence of a judge, which are the base of an effective jurisdiction and part of the right to a fair trial.

#### *Privacy and labour law*

Even in the field of privacy law the importance of the balancing test often emerges for the necessity to solve conflicts that involve equally protected rights. In particular, the Grand Chamber of the ECHR in the case *Bărbulescu v. Romania*, no. 61496/08, 5 September 2017, balanced the right to respect for an employee's private life and correspondence, on the one hand, and his employer's right to take measures in order to ensure the smooth running of the company, on the other. When those rights came to collide, the Court uses a proportionality test to decide if the behaviour of the employer, who controlled private electronic messages of the employee, was an acceptable impairment of the rights set in art. 8 of European Convention of Human Rights. The Court specifies the criteria to be applied by the national authorities when assessing whether a measure, taken to monitor employees' communications, is proportionate, considering that in the case submitted the domestic courts do not carry out a "*sufficient assessment of whether there were legitimate reasons to justify monitoring the applicant's communication*" (§ 135).

The case is relevant not only for the reasoning used by the Court but even because it shows how the conclusions of a balancing test may change from one judge to another: indeed, the Grand Chamber decided in a different way from the Chamber (decision of the 12 January 2016), which did not recognize any violation of art. 8 of ECHR.

Even in other and more recent decisions (see *López Ribalda v. Spain*, 9 January 2018; *Libert v. France*, 22 February 2018) the Court underlined how the balancing test is fundamental to assess interests at stake, focusing on the key role played by the national judge as the authority called to strike a fair balance of the interests protected by the provisions of the Convention.

#### *Privacy and administrative law*

The Italian Constitutional Court (sentence no. 20 of 2019) has recently been involved in resolving a question of balancing between, on the one hand, the right to privacy and, on the other hand, the freedom of access of all citizens to the data held by public authorities. In particular, the Court found that Italian laws were in contrast with Constitution and European Law in the light of the principle of proportionality where it established the obligation for all public managers to publish on the internet complete information on their incomes.

#### *Discrimination and legal certainty*

In the *Dansk Industri case*<sup>25</sup>, the Danish judge was called to balance the principle prohibiting discrimination on the grounds of age, such as interpreted by the well-established case law of the ECJ, with the principles of legal certainty and protection of legitimate expectations, both recognised also by the domestic legal order, with an autonomous and potentially conflicting meaning, as fundamental constitutional values. The ECJ ruled that a national court cannot rely on the latter principles in order to continue to apply a rule of national law that is at odds with the general principle prohibiting discrimination on grounds of age.

As emerging from the cases discussed above, the balancing test is something that A.I. cannot replicate because it does not involve syllogistic reasoning, subjected to the non-contradiction principle. European Courts assess different interests considering every single circumstance of the case and evaluating them in a not deductive way. The rule included in the minor premise is just one, and the A.I. tools cannot decide which one has to prevail; they can solely draw conclusions when the rule applicable is known. Starting from two premises that are clearly in contrast between them, the machine cannot determine the most relevant interest. This kind of mental operation is something that the actual technology still cannot reproduce. It involves the interpretation of the whole context of the case, that goes beyond the deductive reasoning based on the analysis of previous cases – the only things that a computer can consider – and keep the judicial system evolving according to the social change.

---

<sup>25</sup> C-441/14, *Rasmussen v. Dansk Industri*. Mr Rasmussen had been employed by Ajos since 1 June 1984, he was, in principle, entitled to a severance allowance equal to three months' salary under Paragraph 2a(1) of the Law on salaried employees. However, since he had reached the age of 60 by the date of his departure and was entitled to an old-age pension payable by the employer under a scheme which he had joined before reaching the age of 50, Paragraph 2a(3) of that law, as interpreted in consistent national case-law, barred his entitlement to the severance allowance, even though he remained on the employment market after his departure from Ajos.

### 2.3 The limits of A.I. compared to H.I.

The judge always has to balance rights and to evaluate the factual circumstances and a computer cannot do this activity. The Italian Constitutional Court<sup>26</sup>, dealing with recidivism, stated that absolute presumptions violate the principle of equality, if they are arbitrary and irrational, that is, if they do not respond to generalized experience data, summarized in the formula of the *id quod plerumque accidit*. The provision of an obligatory increase of penalty only linked to the crime title, without any “ascertainment of the concrete significance of the new crime episode ‘under the profile of the more accentuated guilt and the greater dangerousness of the offender’” also violates the article 27, third paragraph of the Constitution, which implies “a constant ‘principle of proportion’ between quality and quantity of the sanction, on the one hand, and offence, on the other”.

The risk of using an algorithm of predictive justice can be racial prejudice, a criminal law based on the author and on his characteristics more than on the facts. That’s what happened in the US. None of the sentencing instruments uses race as a variable, yet many variables included in the models play the role of “proxies” for race, in that they strongly correlate with race and reflect racial bias. For example, considering a defendant’s place of residence (e.g. zip codes) can end up targeting neighbourhoods where residents are predominantly Low-income African-Americans. These group-based features are then incorporated into the algorithms, which may mean that racial minorities face longer sentences for the same crimes as similarly adjudicated non-minority defendants. A ProPublica<sup>27</sup> study found that COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), used in *State v. Loomis* case, predicts black defendants will have higher risks of recidivism than they actually do, while white defendants are predicted to have lower rates than they actually do.

People have the right to be treated – and sentenced – as individuals and not on account of “risky” characteristics of a group to which they belong<sup>28</sup>. This is exactly what the Italian Constitutional Court considers illegitimate, an automatic sanction based exclusively on some characteristics of the defendant or on the kind of crime committed.

Considering all the characteristics of the human intelligence that have been underlined until now, other limits of AI can be summarised as follows<sup>29</sup>. The first one is the difficulty of translating normative or factual propositions into a language that allows computation. This difficulty is declined, in particular: a) in the complexity to translate in a logically treatable form the propositions of everyday life and therefore also the juridical ones; b) in the difficulty of considering all the variables that come into play; c) in the difficulty of dominating ambiguous concepts; in other words algorithms lack the human ability to individualize. Two other limits are: d) an inferential engine based on deductive logic

---

<sup>26</sup> Italian Constitutional Court, case n. 28/2015

<sup>27</sup> Available at <https://www.propublica.org/article/how-we-analyzed-the-compass-recidivism-algorithm>.

<sup>28</sup> Angèle Christin, Alex Rosenblat, and Danah Boyd, *Courts and Predictive Algorithms*, in [www.datacivilrights.org](http://www.datacivilrights.org).

<sup>29</sup> See, R. Rumiati, C. Bona, Dalla testimonianza alla sentenza, *Il giudizio tra mente e cervello*, Bologna, Il mulino, 2019, p. 249.

allows to carry out the only reasonings dominated by it; e) the decision is the result of inferential mechanisms that follow their own logic. Finally, the computers are intelligent but not wise, they learn by imitation, so their rationality depends on the reference model.

## **Chapter 3: Dealing with Ethics and AI in the justice field**

### **3.1 The Ethical charter on the Use of Artificial Intelligence in Judicial Systems**

AI in the justice field, as it is already shown above, is not about the future, but it is already applied in different ways; so the question is no longer whether or not we should be in favour of AI development but, more realistically, it is important to open a debate on how the judicial system will be able to rely on technology without being overwhelmed and overturned by them.

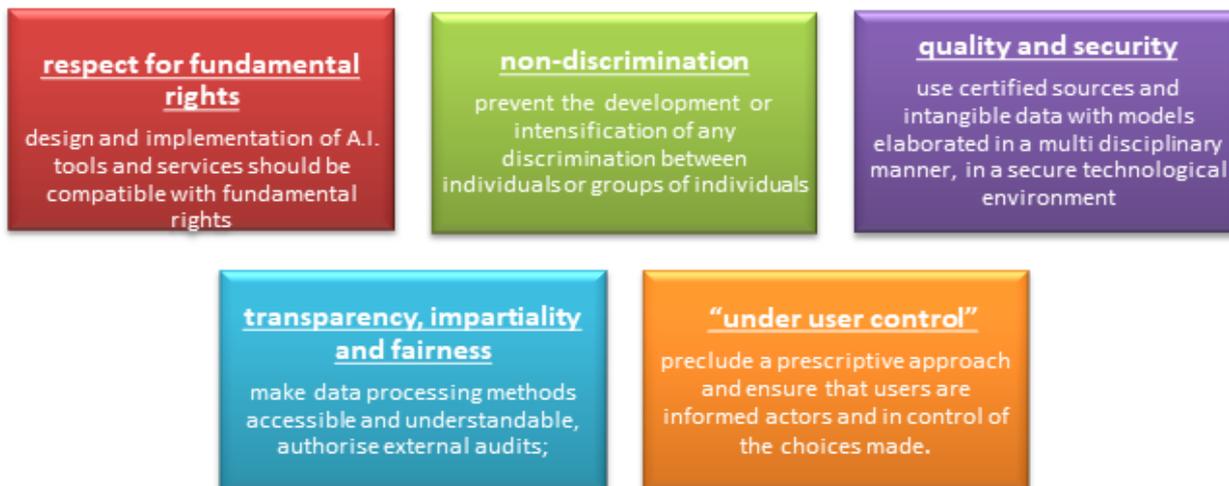
Not only scholars and private researchers' associations but even public institutions in Europe are contributing to the debate about using AI technical innovation to improve the justice system, paying attention to the respect of human rights, adopting acts of soft law that should be able to orient judges, legislators and lawyers towards a rights-oriented approach to AI.

A key role between the soft law documents on AI application on justice is now assumed by the "European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment", approved by the European Commission for the efficiency of justice (CEPEJ), which is an innovative body for improving the quality and efficiency of the European judicial systems and strengthening the court users' confidence in such systems, established by the Committee of Ministers of the Council of Europe.

The Charter is having a significant influence on the debate about AI and justice, because it is adopted by an international – and influential – organism and it is indistinctly directed to all kind of audience of public or private actors involved in the construction and utilisation of AI tools applied to justice.

The growing importance of AI-based application in the justice systems pushed the CEPEJ to seek the state-of-the-art in that field, with special regard to the application of AI for improving the efficiency of justice and to the responsibility that may arise from those applications.

The Charter aims to describe how AI tools are used in the different European countries and which issues may be involved looking forward to further development of these technologies. As a result of this analysis and reflection, CEPEJ fixed five principles, which are the core content of the Charter, to guide public and private actors involved in the judicial systems to approach to the challenge of the AI. In particular, those principles are:



### 3.2 Perspectives and concerns about using AI tool

The main task of the CEPEJ - as it can be read on its website - is to offer public policy makers important methods to better understand all the innovations proposed and perceive their potential, as well as their limitations.

As it emerges from the brief description of the principles set up by the CEPEJ, AI tools are not regarded as a substitute of the judge but as a support for those who are working in the justice system, to improve accuracy and efficiency of justice. Nonetheless, this perspective shows how there still are some issues, which all subjects involved in the judicial field need to face.

First, the Charter underlines the importance of the technical design of AI tools, to ensure respect of human rights and coherence to the legal framework: it is necessary to involve ethical rules directly in designing and programming phase, realising wide cooperation between judges, lawyers and software engineers.

In designing software suitable to the Charter's principles, engineers should adopt an *ethical-by-design* or *human rights-by-design* approach. Algorithms – which are at the base of most AI tools applied to justice – are always exposed to the risk of bias, that is particularly dangerous in the justice field (as is shown by the American *Loomis* case discussed above).

One way to prevent this consequence – as the CEPEJ suggests – may be watching over designers and make sure that what they have in mind is fully compliant with the Charter's principles. Basically computers are built on inputs inserted by the human designer, even if they are able to generate outputs that seem "intelligent" because they are unpredictable or based on learning by experience; hence, an AI is only as good as the programming that goes into it, and the leverage of data inputs and programming is the most effective to really enhance control on this kind of technical instruments.

An AI system is not - by itself - able to explain judicial reasoning, but it can only express a decision which is the reliable output of correlation and logic steps at the base of the archive of decisions that it previously analyses.

This attitude produces a lack of motivation, which is a serious attempt to the respect of the right to a fair trial; so the Charter - in particularly reading the Appendix III - invites to assure the best possible transparency in the whole A.I. influenced decision-making process: there should not be any lack of transparency in algorithm construction processes, often made by private proprietary companies which own intellectual property rights. Software producers should be accountable to the public, all the more if they are part of steps taken by state authorities to make data available to the public.

Applying the principles set by the Charter, the CEPEJ has drawn four categories of uses of AI tool in the justice field, divided on a scale from “uses to be encouraged” to “uses to be considered with the most extreme reservations”. This catalogue is intended to show the right way to develop AI in an ethical way, and it clearly demonstrate that now, analysing AI tools actually used in Europe and in the United States, most of the uses of decision-making algorithms (from online dispute resolution in civil matters to judge profiling and anticipating court decisions) are not really reliable to substitute the judge and always pose some ethical issues.

### **3.3 Which risks arise from implementing a robot-judge?**

Is the world - considering the actual state of technological evolution - ready to fully implement AI decision in the justice field, without any ethical concerns and in full respect of human rights? Reading the Charter, the answer seems to be no.

Algorithms, even if built transparently, are often designed by private corporations, not directly accountable to anyone but their shareholders and out from any democratic control. Therefore, a real concern emerges about the compatibility of implementing algorithms in the judicial decision-making process, because software designers are not linked to any political decision, expressed by the people through free elections, and they are even out of the political control of public administration. So algorithms are far from any democratic legitimacy for the power they would be eventually appointed<sup>30</sup>; there is an evident lack of democratic legitimacy in judicial decision delegated to A.I. tools – created by private companies – and that would be a wound to the democratic principles which rule our societies.

Conclusions of the CEPEJ should be accepted as a general method to introduce ethics in designing AI tools and to address the question of AI not underestimating human rights issues. Nevertheless, full control on the algorithm - *ex ante* with human rights-by-design and *ex post* with independent experts

---

<sup>30</sup> See Nemitz P., *Constitutional democracy and technology in the age of artificial intelligence*, in *Governing artificial intelligence: ethical, legal and technical opportunities and challenges*, Royal Society Publishing, 2018, who stresses that challenges of AI need to be addressed by rules which are enforceable and encompass the legitimacy of the democratic process. On the issue of democratic legitimacy of AI tool see even Scherer M. U., *Regulating Artificial Intelligence Systems: Risks, Challenges, Competences, and strategies*, in *Harvard Journal of law and technology*, 29, 2016 and Danaher J. *The Threat of Algocracy: Reality Resistance and Accomodation*, in *Philosophy and Technology*, 2016.

evaluating technical suitability of the software - could not be enough to fulfil the principle of a fair trial.

This principle requires a fully detailed motivation on the appreciation of each proof or fact, especially as it is shaped in some countries. This control cannot be done only by verifying the technical adequacy of the algorithm; in order to assure a full control from the addressees of the decisions and from public opinion, it has to be expressed by more complex advice, to avoid the uncertain reliability of the algorithmic evaluation altogether.

Anyways, the party concerned in a trial should have access to and be able to challenge the scientific validity of an algorithm, the weighting given to its various elements and any erroneous conclusions it comes to. Not only the party but even the judge of the appeal should be able to question the logic and functioning of the algorithm. The Italian Council of State requires this kind of control<sup>31</sup> - the last instance administrative court - exercising his supervision on “digital” administrative decisions. The Council of State set an equation between algorithm and motivation, assuming that the judge should have the same kind of control of every logic step made by the algorithm, annulling the administrative decision if the motivation is not coherent or sufficient.

This type of control is the same foreseen by the CEPEJ, but it cannot work in all cases: the power of algorithms is the capacity of doing a number of logical steps that goes beyond human capacities: that is the real advantage using AI tools, but it can prevent humans to really control this enormous amount of data. Considering these limits, human *ex post* control on the logical coherence of an AI decision could be not trustworthy.

Data scientists’ studies<sup>32</sup> showed that algorithms are not really fair and objective, because they are directly connected to the definition of “success” given by the designer; they can ignore fundamental variables, incorporate prejudices and, if they are wrong, they do not easily consent any appeal. An algorithm is an “opinion embedded in math”, so it can not be really fair and neutral and neither it can be used as a substitute of the judge.

Hence, there is another risk arising from predictive justice, that is not only linked to the incomplete development of technical instruments or to the unreliability of the data output<sup>33</sup>. The judge could be influenced by the results of the algorithm-based output, and become more indifferent to the specificities of the concrete case. This kind of psychological risk is called - by the French philosophers Garapon and Lassegue<sup>34</sup> - *effet moutonnier*, and it brings the AI outcomes to rise at a normative status, depriving justice of the human value and evolutive chances. Judges should evolve

---

<sup>31</sup> Italian Council of State, decision n. 2270/2019.

<sup>32</sup> O’Neil C., *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Books, 2016.

<sup>33</sup> On this topic, for more details see the Annex III to the Ethical Charter, § 81 and § 105, where is discussed the online dispute resolution mechanism and the study conducted by the UCL on the ECHR jurisprudence.

<sup>34</sup> Garapon and Lassegue, *Justice digitale. Révolution graphique et rupture anthropologique*, Paris, 2018, pp. 279

and adequate to new social-economic contexts: too much stability in the case-law – empowered by data analysis – may lead to deciding issues without questioning about social changes.

This can also constitute a breach of the principle of judge's independence, even from the previous case-law, overall in those countries who have a civil law tradition, where *stare decisis* principle is not mandatory as in the common law countries. As the jurisprudence of the ECHR also confirmed, the issue of legal certainty should be balanced with the vitality in judicial interpretation (ECHR, *Greek Catholic parish Lupeni and Others v. Romania*, 29/11/2016, § 116).

### **3.4 European Commission approach on implementing AI**

Even another institution is trying to manage an ethical approach to all questions arising from AI tools at the European level, focusing on some of the topics linked to the justice field.

The European Commission established a group of independent experts to draft ethical guidelines for Artificial Intelligence, sharing this document with industry, research institutes and public authorities. The so-called “guidelines for a trustworthy AI” establish three fundamental elements for building trust in a human-centric AI. In particular, trustworthy AI should be: 1) lawful - respecting all applicable laws and regulations; 2) ethical - respecting ethical principles and values; 3) robust - both from a technical perspective while taking into account its social environment.

In order to assure those tasks, the Expert group put forward a set of 7 key requirements and a specific assessment list aims to help verify the application of each of the key requirements. Looking at those requirements is possible to note they are comparable to the principles of the Ethical Charter discussed above: for example, the European Commission stressed the importance of a human rights oriented approach, of complete transparency and non-discrimination and fairness.

In addition, the guidelines underline the importance of accountability, to ensure responsibility for AI systems and their outcomes, and of auditability, which enables the assessment of algorithms, data and design processes and plays a key role therein, especially in critical applications, like of course are the possible uses in the justice field.

These rules and guidelines are more than ever useful indications for those countries who want to implement AI tools to improve efficiency and quality of justice. Between them, the most ambitious project comes from Estonia, where the Ministry of Justice asked to design a “robot judge” that could decide small claims disputes. Even if the project is in its early phases, it suggests how AI tools are spreading faster in the EU administration of justice.

### **3.5 To the future and beyond**

The reference of all the soft law acts is the current technology and the actual possibilities to apply those innovations to justice. But what if we look forward to the evolution and goals that technology could reach in the future on AI - even if not realized yet?

A lot of interest in the AI field is now arising from studies about quantistic computer, which represents a technology which uses the rule of relativity to create “bits” which are not conditioned to a

single status but which can assume two different status at the same time. Nowadays, the quantistic computer already exists - the IT company IBM had already built one model - but is not seen as a very revolutionary innovation, because even if it is faster than a traditional computer, it uses the same scheme for reasoning, and it is not really mimetic of the human mind.

But some scientists<sup>35</sup> suggested that this technology could be integrated into deep learning machine - which today is commonly used - to create an AI tool able to explore an infinite number of variables.

This technology could lead to a computer able to manage “quantistic category”, not subjected to the non-contradiction principle; a single thing could be, at the same time, black and white, true and false. This kind of AI will generate results and solutions that the human mind simply can control neither understand.

Can this kind of future AI be the basis for a robot judge, without the limits of the human being but able to emulate even the balancing reasoning?

Technically it could, but we think that it should not.

As Stéphane Leyenberger, Executive Secretary of the CEPEJ, said: “*justice remains a moral principle of social life based on the recognition and respect of the order and the others’ rights*”.

Justice should remain, in its core, fully human, even considering the social role of the judge, which is more than a technician of the law. The trial and hearings are a moment in which the citizen meets the State, and find legitimacy for an institution and for exercising the powers linked to justice. An automated decision, even if right, could not be really fair if it is deprived of its human side, which can even consist only in a certain symbolic significance and a sympathetic ear for the more vulnerable people in our society<sup>36</sup>.

The real concern, thinking of a higher degree of automation, is that people will distrust the judiciary, starting seeing the law no more like a verdict handed down from society, but rather as the result of an incomprehensible equation. Justice is not only to apply the law, but its core is to reach - between a complex, ritualized and even exhausting process - the solution of a dispute which could be socially shared and accepted, said Jean Lassègue.

In conclusion, for all the fear of AI doling out the wrong sentences, perhaps the real concern comes from what happens if they deliver the right ones.

---

<sup>35</sup> Prati E., *Mente Artificiale*, Milano, 2017.

<sup>36</sup> As Burkhard Schafer, Professor of Computational Legal Theory at the University of Edinburgh, said, quoted by McMullan T., *A.I. Judges: The Future of Justice Hangs in the Balance*, in [www.medium.com](http://www.medium.com); Schafer is going to publish the results of his studies in the book *Future law: Emerging Technology, Ethics and Regulation*, that will be released on 1 February 2020 by Edinburgh University Press.