SOME CONSIDERATIONS ON THE RELATIONSHIP BETWEEN HUMANS AND ARTIFICIAL INTELLIGENCE

THE "HUMAN RESERVE" AND HUMAN SUPERVISION

Juli Ponce Solé

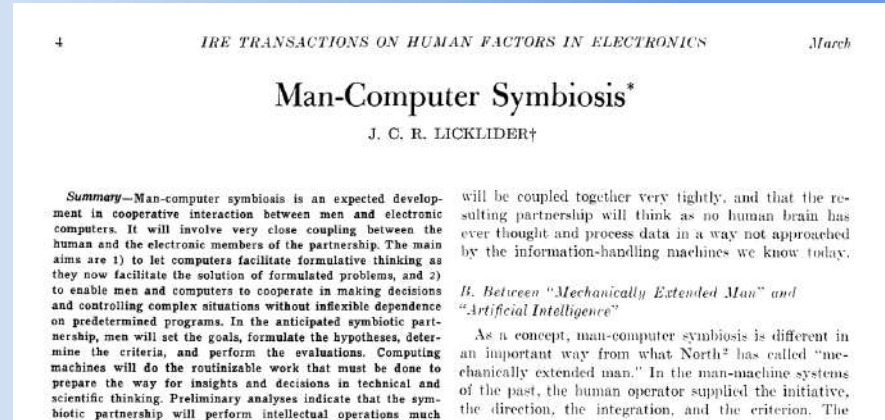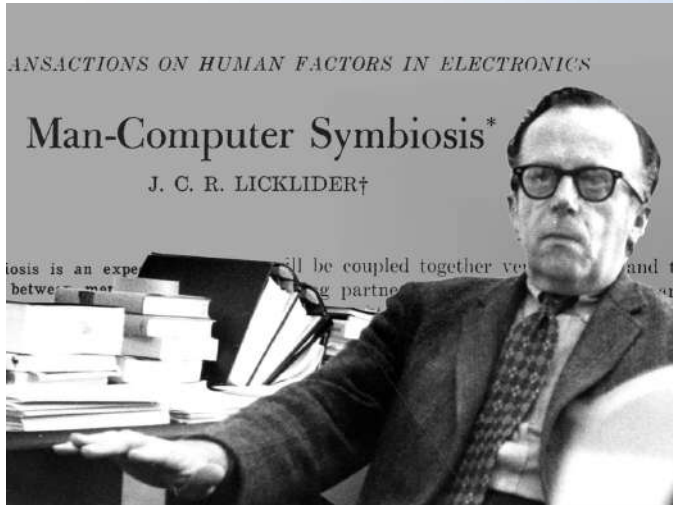jponce@ub.edu

1

INTRODUCTION

- HUMAN - ARTIFICIAL INTELLIGENCE INTERACTION: 1960

- DOCUMENTARY: The Man Who Saved the World

This Man Single Handedly Stopped All Out Nuclear War - The Man Who Sa... ...ntary - Bing video,

- LAW AND LIMITS AND SAFEGUARDS TO THE USE OF ARTIFICIAL INTELLIGENCE IN TAKING FORMAL ADMINISTRATIVE DECISIONS (RULE-MAKING, ACTS, CONTRACT, PLAN)

ADVANTAGES AND RISKS OF AI:

- ADVANTAGES OF AI

    – GENERATION NEW EMPLOYMENT
    – GREATER PROCESSING CAPACITY
    – PREDICTIONS THROUGH CORRELATIONS
    – AVOIDING COGNITIVE BIASES AND NOISE

- Italian Consiglio di Stato, decisión 13/12/201913, No 8472
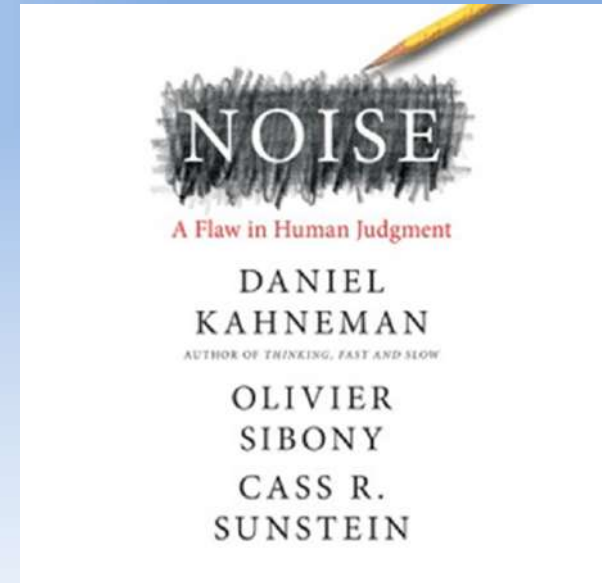
- "7.1 In general terms, it should be stressed that public administration must also be able to exploit the considerable potential of the digital revolution.In this context, the use of computer algorithms for decision-making in the public and private spheres is based on the fear of efficiency and neutrality. In many fields, **algorithms promise to become the tool through which to correct the distortions and imperfections that typically characterise the cognitive processes and choices made by human beings, highlighted especially in recent years by an impressive literature on behavioural economics and cognitive psychology.** In this context, the decisions made by the algorithm thus assume an aura of neutrality, the result of aseptic rational calculations based on data".

4

– GREATER EFFECTIVENESS, EFFICIENCY: BETTER ADMINISTRATION

Italian Consiglio di Stato, decision 4.2.2020, Nº 881

"As already highlighted in the previous section, the usefulness of this operational method of public interest management is particularly evident with regard to procedures, such as the one at issue in this litigation, which are serialised or standardised, involve the processing of large numbers of applications and are characterised by the acquisition of certain and objectively verifiable data and the absence of any discretionary appreciation.

The full admissibility of these instruments is in line with the canons of efficiency and cost-effectiveness of administrative action (Article 1 of Law No. 241 of 1990), which, in accordance with the constitutional principle of good administration ("buon andamento") (Article 97 of the Constitution), require the administration to achieve its objectives with the least expenditure of means and resources and by streamlining and accelerating the procedure".

RISKS: POSSIBLE BIG SCALE IMPACTS

UNEMPLOYMENT

Revista Vasca de Gestión de Personas y Organizaciones Públicas
Núm. especial - Special issue 3/2019. Págs. 34-51

EL FUTURO DEL TRABAJO EN LA ADMINISTRACIÓN PÚBLICA. ¿ESTAMOS PREPARADOS?
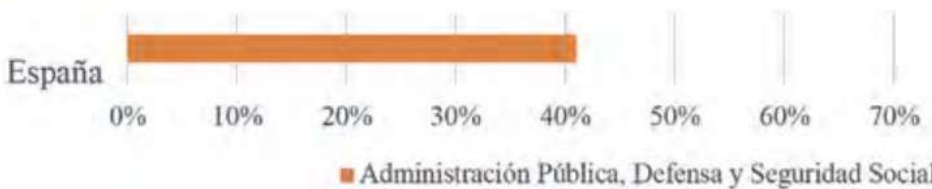FUTURE OF WORK IN THE PUBLIC SECTOR. ARE WE READY?

Andrés Pastor Bermúdez
Gerente adjunto
Gerencia de Informática de la Seguridad Social
andres.pastor@seg-social.es

Pedro Nogales Fuentes
Jefe de la unidad de nuevas tecnologías
Gerencia de Informática de la Seguridad Social
pedro.nogales@seg-social.es

Laburpena: Hurrengo urteetan, industria-iraultzaren industriako lanaren eraldaketarik handienera joan gara. Urtean 2035. urterako aurrerapen robotiko edo kognitibo teknologiko aurreratuek aurrera egiteko aukera ugari automatizatzeko aukera emango dute sektore guztietan eta lanbideetan, eta horrek esan nahi du milioika lanpostu baino gehiago desagertzea Espainiako administrazio publikoetan, aldi berean agertzen diren lanbideak ez dira existitzen. Egiteko une egokia da eta digitalen profilak sortzea, oso moldagarriak dira oraindik ere definitzeko. Denboran aurrera egin ahala, paradigma-aldaketari aurre egiteko proaktibotasun-aukera leihoa murrizten da, bai erakunde publikoengatik, bai langileentzat.
Gako-hitzak: digitalizazioa, automatizazioa, adimen artifiziala, robotak, datuen azterketak, lana, enplegua, talentua.

Gráfico 4

Probabilidad de automatización en función del nivel de estudios del trabajador (%)

Fuente: elaboración propia. a partir de datos de Arntz, Gregory, and Zierahn (2016). OCDE.

Gráfico 6

Riesgo de automatización en la AAPP España

■ Administración Pública, Defensa y Seguridad Social

Fuente: elaboración propia a partir de datos de Nedelkoska et al., 2018.
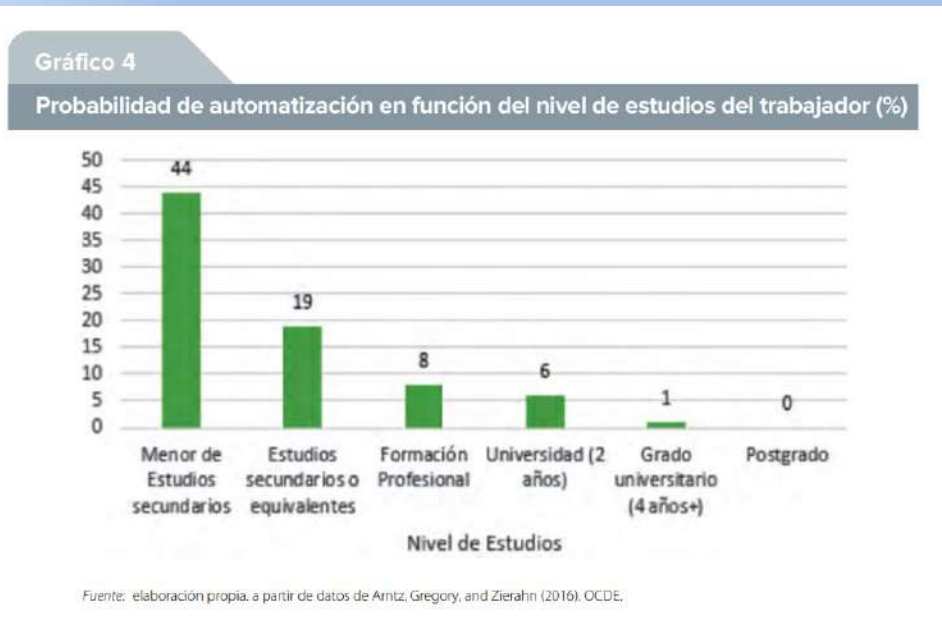
RISKS: POSSIBLE BIG SCALE IMPACTS

– **CORRELATIONS** INSTEAD OF CAUSALITY: STATISTICAL HALLUCINATIONS AND CONSERVATIVE TENDENCY

– **ERRORS (BUGS)**

– **BIASES**: PEOPLE (AUTOMATION BIAS), PROGRAMMERS, DATA AND STATISTICS

– **OPACITY**: BLACK BOXES... WHICH ARE CONTROLLED IN HUMANS BY MEANS OF THE OBLIGATION OF DUE DILLIGENCE OR DUE CARE LINKED TO THE GOOD ADMINISTRATION

Mailing Date: 3/26/08

Dear SOPHIE STIPES,

MA D 01 (MI)

Your MEDICAID benefits will be discontinued effective APRIL 30, 2008 due to the following reason(s):

-FAILURE TO COOPERATE IN ESTABLISHING ELIGI-BILITY

-FAILURE TO COOPERATE IN VERIFYING INCOME SUPPORTING LAW(S) OR REGULATION(S) :

470IAC2.1-1-2

Important : If you believe you may be eligible for Medicaid benefits under another category and have more information about your

8

UNIVERSITAT DE BARCELONA

I will focus specially on full automation of discretionary administrative decisions

Human-Centered AI
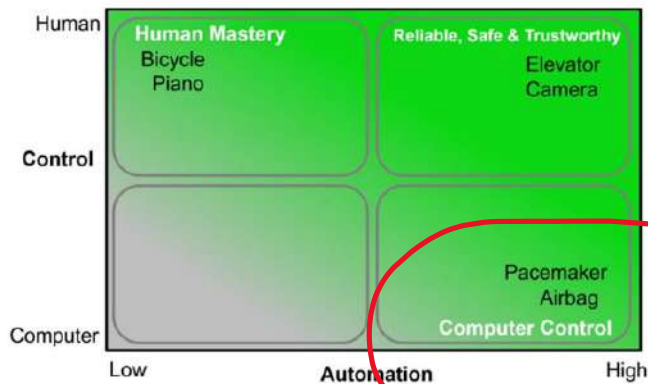
Figure 3: Regions requiring rapid action (high automation, low human control) and human mastery (high human control, low automation).

Figure 3. **Typologies of automated decision-making and human intervention**

Image source: Binns y Veale 2021.

DISTINCTION

- 2 TYPES OF AI USE

  - AUTOMATED DECISIONS
  - SEMI-AUTOMATED DECISIONS, SUPPORT FOR HUMAN ADMINISTRATIVE DECISION

- 2 TYPES OF AI

  - SYMBOLIC, GOOD OLD FASHIONED AI (GOFAI), SIMPLE, DETERMINISTIC: develops deductive inferences
  - NON-SYMBOLIC, STATISTICAL, COMPLEX, PREDICTIVE (MACHINE LEARNING, DEEP LEARNING): develops inductive inferences

- 2 TYPES OF ADMINISTRATIVE POWERS

  - DISCRETIONARY
  - NONDISCRETIONARY, BIND POWER

HUMANS VS. AI

RISKS OF AI: GOING BEYOND

POSSIBLE VIOLATION OF HUMAN DIGNITY AND RIGHTS (DRAFT EU REGULATION ON AI)

WHY AND HOW AN ALGORITHMIC SYSTEM CAN AFFECT DIGNITY OR RIGHTS IN A SPECIFIC WAY THAT A HUMAN CANNOT?

In relation to fully automated decisions in the field of discretionary powers:

1. *RULE-BASED SYSTEMS OF AI (SIMPLE/DETERMINISTIC) THAT DEVELOP DEDUCTIVE INFERENCES*: WHEN USING AI IT IS NOT POSSIBLE ELIMINATE DISCRETIONARY POWERS IF EXISTING: PROHIBITION OF *FETTERING* BY THE RIGHT TO GOOD ADMINISTRATION

2. *AI USING STATISTICS (COMPLEX/PREDICTIVE) THAT DEVELOP INDUCTIVE INFERENCES (E.G. MACHINE LEARNING, DEEP LEARNING):*

– **AI HAS NO EMPATHY**
– **AI CAN NOT MAKE ABDUCTIONS**   VIOLATION RIGHT TO GOOD ADMINISTRATION

3. IN BOTH CASES, *AUDI ALTERAM PARTE*: NEED OF RESPECTING THE RIGHT TO GOOD ADMINISTRATION

UNIVERSITAT DE BARCELONA

1. RULE-BASED SYSTEMS (DEDUCTIVE INFERENCES):WHEN REGULATING THE USE OF AI IT IS NOT POSSIBLE ELIMINATE DISCRETIONARY POWERS: PROHIBITION OF FETTERING

- The problem in rule-based systems

  – Legal prohibition of <mark>fettering</mark>: duty of due care or due diligence linked to the right to good administration

2.STATISTICAL AI (INDUCTIVE INFERENCES):

1. IT HAS NO EMPATHY
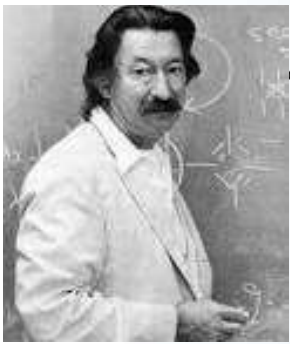
- LACK OF **EMPATHY** (DIFFERENT FROM SIMPATHY) OF AI

- BEEING HUMAN IS HAVING THE CAPACITY OF A FEELING OF IDENTIFICATION WITH OTHER HUMANS (EXCEPT PSYCHOS)

- AI CAN IMITATE EMPATHY…LIKE **PSYCHOS**

https://www.youtube.com/watch?v=2MG7gGQvaG0

THIS ODDLY CHEERFUL AND OPTIMISTIC AI IS A CREEPY COMBINATION OF **NORMAN BATES** (FROM THE 1960 ALFRED HITCHCOCK MOVIE PSYCHO) AND A **ROBOT**

EMPATHY AND WISDOM IS OUT OF AI

JOSEPH WEIZENBAUM

COMPUTER POWER AND HUMAN REASON

FROM JUDGMENT TO CALCULATION
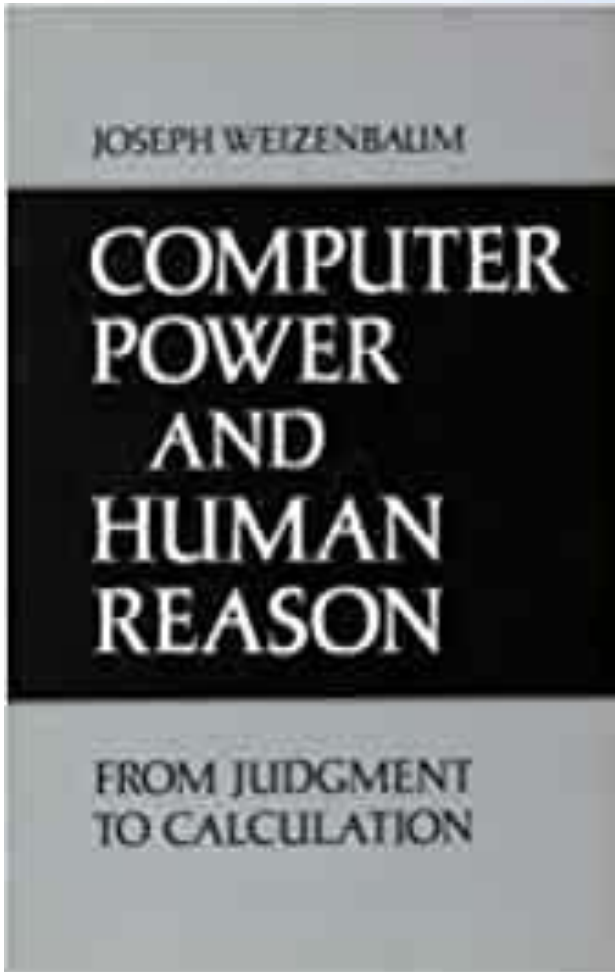
Weizenbaum (a MIT scientist) called in **1976** for societal consensus that machines not replace humans in work that benefits from wisdom and empathy—a state that a computer would be unable to have (but can mimic emotions like psychos). Drawing on the work of other contemporaries looking at specific instances in which A.I. would be inappropriate, Weizenbaum named:

-customer service representatives,
-therapists,
-eldercare workers,
-soldiers,
-judges,
-and police officers

as roles that ought to be fulfilled only by humans.

Public employees?

15

- EMPATHY, PUBLIC SERVICES AND PRINCIPLE OF EQUITY
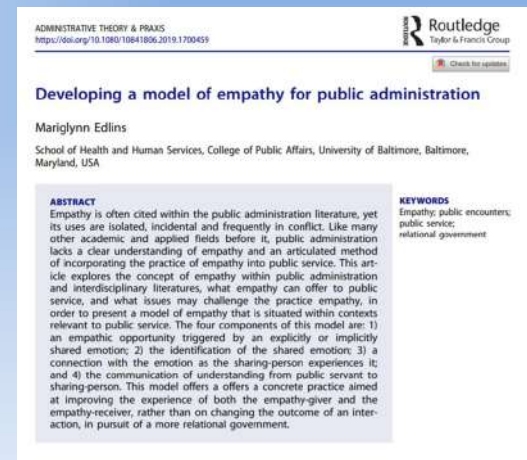
- ANTONIO DAMASIO AND THE DESCARTE´S ERROR: ENLIGHTMENT, COLD RATIONALITY AND EMOTIONS.

- RIGHT TO GOOD ADMINISTRATION: TAKING INTO CONSIDERATION ALL RELEVANT FACTORES BEFORE DECIDING, INCLUDING THE SPECIFIC SITUATION OF HUMANS IN LEGAL RELATIONS WITH PUBLIC ADMINISTRATION

2.STATISTICAL AI (INDUCTIVE INFERENCES):

2.IT CAN NOT DEVELOP ABDUCTIVE REASONING

- "a form of reasoning where assumptions are made to explain observations. For example, if an agent were to observe that some light was not working, it can hypothesize what is happening in the world to explain why the light was not working. An intelligent tutoring system could try to explain why a student gives some answer in terms of what the student understands and does not understand.

- The term **abduction** was coined by Peirce (1839-1914) to differentiate this type of reasoning from **deduction**, which involves determining what logically follows from a set of axioms, and **induction**, which involves inferring general relationships from examples": Artificial Intelligence - foundations of computational agents -- 5.6 Abduction (artint.info).

AI CAN NOT DEVELOP ABDUCTIVE REASONING

UNIVERSITAT DE BARCELONA

Abductive policy making. Draft section of report from research fellowship – Researching design for policy (wordpress.com)

Inductive vs. Deductive vs. Abductive Reasoning | Merriam-Webster

Deduction	Induction	Abduction
Compliance of the datum to the theory (rule)	Compliance of the theory (rule) to the datum	Circularity output and search for other theories
Rule - All the balls in this box are white	Case - These balls are from this box	Rule - All the balls in this box are white
Case - These balls are from this box	Result - These balls are not white	Result - These balls are not white. What could explain this?
Result - These balls are white or more or less white	Rule - Not all the balls from this box are white	Case - These balls are from another box (so where would they be?)

Source. Adapted from Peirce (1970 /1878).

18

3. IN GENERAL, *AUDI ALTERAM PARTE*

- Human intervention is imposed by law during administrative procedures, avoiding completely automated decisions

- Existence of the right to be heard (*audi alteram parte*), a component of the right to good administration established by article 41 of the European Charter of Fundamental Rights and a part of the constitutional traditions of the member states, in connection with the right of every person to have access to his or her file, while respecting the legitimate interests of confidentiality and of professional and business secrecy.

NO FETTERING, LACK OF EMPATHY, NO ABDUCTION AND *AUDI ALTERM PARTE*: NO AI FOR FULLY AUTOMATED DISCRETIONARY DECISIONS=HUMAN RESERVE

- SOCIAL PRECAUTIONARY PRINCIPLE AND LAW:

  – "RESERVE OF HUMANITY"

HUMANITY RESERVE AND "ARTIFICIAL DISCRETION"

- EXCLUSION OF FULLY AUTOMATED DISCRETIONARY DECISIONS, TAKEN ONLY BY AI, IN CERTAIN AREAS, NOT DUE TO TECHNICAL REASONS, BUT TO LEGAL CONSIDERATIONS

EXAMPLES OF "RESERVE OF HUMANITY" BY LAW

- GERMANY, CATALONIA, SPAIN, EU...

German Administrative Procedures Act (VwVfG)
Section 35a Fully automated issuing of an administrative act

"An administrative act may be adopted entirely by automatic bodies, provided that this is permitted by law and that there is neither a discretion nor a margin of appreciation."

UNIVERSITAT DE BARCELONA

- Article 44. Automated administrative action.

- 1. The Catalan public administrations may carry out automated actions to ascertain that the requirements established by law have been met, to declare the expected consequences, to adopt resolutions and to communicate or certify the data, acts, resolutions or agreements contained in their information systems, by means of the use of the electronic signature system that they determine.

- **2. Only those acts that may be adopted with programming based on objective criteria and parameters shall be subject to automated administrative action.**

- 3. Automated administrative action does not affect the ownership of the competence of the administrative bodies or the competences attributed for the resolution of administrative appeals.

Spanish Charter of Digital Rights

- XVIII.6
- Efforts shall be made to promote citizens' rights as regards artificial intelligence recognized in this Charter in the framework of administrative action, recognizing in all cases the rights to:

decision-making being reserved to persons, in the absence of legislation providing for the adoption of automated decisions with the necessary guarantees

RATIONALE

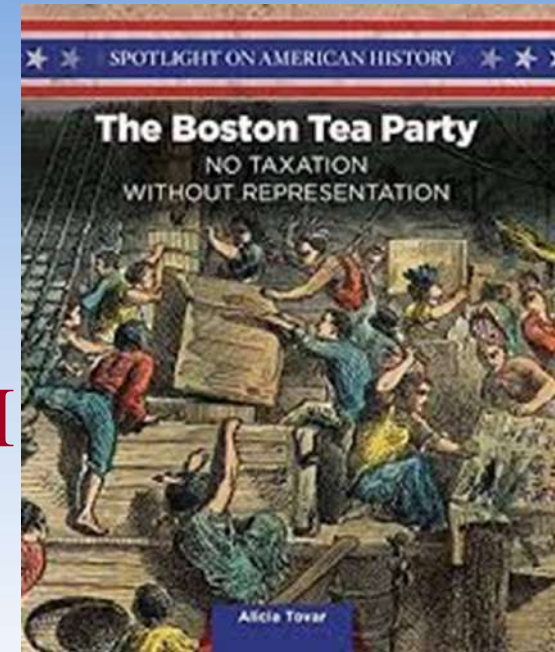- EXISTENCE OF DISCRETIONARY POWERS: NEED FOR ABDUCTIONS AND EMPATHY FOR GOOD ADMINISTRATION

- ALL KINDS OF DISCRETION OUT OF AI?

SPANISH CASE: CAN BE EXCEPTIONS, BUT…

- *NO TAXATION WITHOUT REPRESENTATION…*

NO AUTOMATION WITHOUT REPRESENTATION

- **SPAIN: BY LAW PASSED BY PARLIAMENT (ART.18 SPANISH CONSTITUTION).**

- BUT CONSIDER DANISH POLITICAL AGREEMENT ON DIGITALLY READY LEGISLATION AND SIDE EFFECTS…

EU?

UNIVERSITAT DE BARCELONA

Art. 22 GDPR:

Art. 22 GDPR Automated individual decision-making, including profiling

1.The data subject shall **have the right not to be subject to a decision based solely on automated processing**, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

2.Paragraph 1 **shall not apply if** the decision:

- A) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

- B) **is authorised by Union or Member State law** to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

- C) is based on the data subject's explicit consent.

3.In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least **the right to obtain human intervention** on the part of the controller, to express his or her point of view and to contest the decision.

4.Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

- EU DRAFT REGULATION ON AI PROHIBITING IT USE (UNACCEPTABLE RISK): ART. 5

- NEED OF PROHIBITING USE OF AI IN RELATION TO FULLY AUTOMATED DISCRETIONARY ADMINISTRATIVE DECISIONS? SHOULD BE IT AN UNACCEPTABLE RISK?

UNIVERSITAT DE BARCELONA

- IF AI CAN BE USED (NON DISCRETIONARY DECISIONS IN RULE-BASED AI AND STATISTICAL AI AS A SUPPORT FOR HUMAN DECISION), THEN...HUMAN SUPERVISION

- E.G. ART. 22 GDPR (IN CASE OF EXCEPTION TO THE RESERVE OF HUMANITY), ART. 41 SPANISH 40/2015 ACT, ART. 96 SPANISH GENERAL TAX LAW, SPANISH CHARTER OF DIGITAL RIGHTS, ART. XXV.2 AND 3 (when the "reserve of humankind" does not apply, i.e. in regulated powers or in discretionary powers where there is a specific rule allowing the use of AI).)

- ART. 14 EU DRAFT REGULATION ON AI AND AMENDMENTS 2023

UNIVERSITAT DE BARCELONA

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS	Council's Common Position on Artificial Intelligence Act 2022	Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))(1)
COM/2021/206 final *Article 14* *Human oversight* 1.High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.		High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they be effectively overseen by natural persons *as proportionate to the risks associated with those systems. Natural persons in charge of ensuring human oversight shall have sufficient level of AI literacy in accordance with Article 4b and the necessary support and authority to exercise that function,* during the period in which the AI system is in use *and to allow for thorough investigation after an incident*.

Amendment 214
Proposal for a regulation
Article 4 b (new)
Article 4 b
AI literacy

- 1.   *When implementing this Regulation, the Union and the Member States shall promote measures for the development of a sufficient level of AI literacy, across sectors and taking into account the different needs of groups of providers, deployers and affected persons concerned, including through education and training, skilling and reskilling programmes and while ensuring proper gender and age balance, in view of allowing a democratic control of AI systems*

- 2.   *Providers and deployers of AI systems shall take measures to ensure a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, and considering the persons or groups of persons on which the AI systems are to be used.*

- **3.   *Such literacy measures shall consist, in particular, of the teaching of basic notions and skills about AI systems and their functioning, including the different types of products and uses, their risks and benefits.***

- 4.   *A sufficient level of AI literacy is one that contributes, as necessary, to the ability of providers and deployers to ensure compliance and enforcement of this Regulation.*

UNIVERSITAT DE BARCELONA

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS	Council's Common Position on Artificial Intelligence Act 2022	Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))(1)
COM/2021/206 final 2.Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter.		2. Human oversight shall aim at preventing or minimising the risks to health, safety**,** fundamental rights *or environment* that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter *and where decisions based solely on automated processing by AI systems produce legal or otherwise significant effects on the persons or groups of persons on which the system is to be used*.

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS	Council's Common Position on Artificial Intelligence Act 2022	Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))(1)
COM/2021/206 final		
3.Human oversight shall be ensured through either one or all of the following measures: (a)Mesures identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service; (b) Mesures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user.		3. Human oversight *shall take into account the specific risks, the level of automation, and context of the AI system and* shall be ensured through either one or all of the following *types of* measures:

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS	Council's Common Position on Artificial Intelligence Act 2022	Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))(1)
COM/2021/206 final 4.The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances: (a)fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible; (b)remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias'), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons; (c)be able to correctly interpret the high-risk AI system's output, taking into account in particular the characteristics of the system and the interpretation tools and methods available; (d)be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system; (e)be able to intervene on the operation of the high-risk AI system or interrupt the system through a "stop" button or a similar procedure.	**(a) to understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation;** **(b) to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias')** **(c) to correctly interpret the high-risk AI system's output, taking into account for example the interpretation tools and methods available**	4. *For the purpose of implementing paragraphs 1 to 3, the high-risk AI system* shall *be provided to the user in such a way that natural persons* to whom human oversight is assigned *are enabled*, as appropriate *and proportionate* to the circumstances: (a) *be aware of and sufficiently* understand the *relevant* capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible; (e) be able to intervene on the operation of the high-risk AI system or interrupt, the system through a "stop" button or a similar procedure *that allows the system to come to a halt in a safe state, except if the human interference increases the risks or would negatively impact the performance in consideration of generally acknowledged state-of-the-art*.

UNIVERSITAT DE BARCELONA

ART. 14: HUMAN IN THE LOOP IN CASE OF HIGH RISK SYSTEMS

| Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

COM/2021/206 final

5.For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons. | Council's Common Position on Artificial Intelligence Act 2022

For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been separately verified and confirmed by at least two natural persons. **The requirement for a separate verification by at least two natural persons shall not apply to high risk AI systems used for the purpose of law enforcement, migration, border control or asylum, in cases where Union or national law considers the application of this requirement to be disproportionate.** | Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))(1)

5. For high-risk AI systems referred to in point1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons *with the necessary competence, training and authority*. |

37

CONCLUSIONS AND QUESTIONS

- THE FACT THAT AI CAN BE AVAILABLE TO BE APPLIED IN ANY AREA OF ADMINISTRATIVE ACTIVITY DOES NOT MEAN IT IS ALWAYS LEGALLY POSSIBLE AND NECESSARY TO USE IT

- DOES THE NEED OF

  – AVOIDING OPACITY (BLACK BOXES),
  – HAVING JUSTIFICATION AND EXPLANATION OF AI,
  AND
  – HAVING EMPATHY AND
  – ABDUCTIVE INFERENCES

  INTERACT TO JUSTIFY REDUCING THE USE OF NON-SYMBOLICAL AI IN PUBLIC ADMINISTRATION TO JUST SUPPORTING HUMAN DECISIONS?

-NEED FOR PREVIOUS IMPACT ASSESMENTS ABOUT **POSSIBILITY**, NEED, SUITABILITY AND EFFICIENCY OF AI, SPECIALLY IN THE CASE OF DISCRETIONARY POWERS:

- *DEFAULT OPTION, OPT IN*: NO POSSIBLE USE AI TO TAKE FULLY AUTOMATED DISCRETIONARY DECISIONS (DUE TO REASONS LINKED TO RIGHT TO GOOD ADMINISTRATION: FETTERING, LACK OF EMPATHY, LACK OF ABDUCTIVE INFERENCES AND *AUDI ALTERAM PARTE*) UNLESS A LAW PASSED BY PARLAMENT SAYS YES (AND GUARANTEES GOOD ADMINISTRATION).

- PARADOXAL POSSIBLE SIDE EFFECT OF DIGITALIZATION AND HUMAN RESERVE: ELIMINATION OF DISCRETIONARY POWERS BY REGULATION (BOTH LEGISLATION AND ADMINISTRTRATIVE REGULATIONS: E.G. DENMARK). PROBLEMS.

– IF USE OF AI IS POSSIBLE, THEN A SECOND NEW SPECIFIC ASSESMENT OF NEED, SUITABILITY AND EFFICIENCY IS NEEDED CASE BY CASE

- IF AI IS GOING TO BE USED IN A CASE, THERE IS A NEED TO ESTABLISH EFFECTIVE HUMAN SUPERVISION.

  - WHEN? Is the commitment to put a human in the loop is achieved by simply ensuring subsequent human intervention to override the decision, in the case of administrative decisions by means of an administrative appeal, for example?

  - In our opinion, the simple possibility of overriding a decision a posteriori cannot be enough: good administration implies administrative due care and due diligence before the decision is taken through an administrative procedure

WHO?

NOT JUST A HUMAN IN THE LOOP: SHE MUST BE THE RIGTH HUMAN IN THE LOOP: KNOWLEDGE, SKILLS AND DEBIASING.

E.G. **US. Public Law No: 117-207 (10/17/2022) Artificial Intelligence Training for the Acquisition Workforce Act or the AI Training Act**

This bill requires the Office of Management and Budget (OMB) to establish or otherwise provide an artificial intelligence (AI) training program for the acquisition workforce of executive agencies (e.g., those responsible for program management or logistics)

A PUBLIC EMPLOYEE? IS IT A FUNCTION RESERVED TO PUBLIC EMPLOYEES?

- HOW? COMPLEXITY, SPEED AND HUMAN LIMITATIONS: AVOIDING THE *PLACEBO* EFFECT…BUT, IS IT POSSIBLE IN THE CASE OF STATISTICAL AI?

UNIVERSITAT DE BARCELONA

- IS THERE A CONTRADICTION BETWEEN ART. 22 GDPR AND 14 EU DRAFT?

- DO EU MEMBER STATES NEED MORE AND BETTER REGULATION ABOUT THOSE ISSUES?
  - IF YES, CAN A MEMBER STATE GO BEYOND THE FUTURE EU REGULATION ON AI?

- SHOULD CITIZENS HAVE A SPECIFIC LEGAL RIGHT TO HUMAN SUPERVISION IN THEIR RELATIONS WITH PUBLIC ADMINISTRATION?

- WHAT ROLE FOR THE NEW SPANISH AGENCY FOR THE SUPERVISION OF AI?

Royal Decree 729/2023, of August 22, which approves the Statute of the Spanish Agency for the Supervision of Artificial Intelligence.

Article 4.   *Object and purposes*.

1.   The Agency is responsible for carrying out tasks of supervision, advice, awareness and training aimed at public and private law entities for the proper implementation of all national and European regulations regarding the proper use and development of artificial intelligence systems, more specifically, of algorithms. In addition, the Agency will have the function of inspection, verification, sanction and other functions attributed to it by the European regulations applicable to it and, in particular, in the field of

a)   Raising awareness, disseminating and promoting training, and the responsible, sustainable and reliable development and use of artificial intelligence.

b)   The definition of mechanisms to advise and assist society and other actors related to the development and use of artificial intelligence.

e)   The supervision of the implementation, use or commercialization of systems that include artificial intelligence and, especially, those that may pose significant risks to health, safety, equal treatment and non-discrimination, particularly between women and men, and to other fundamental rights.

FINAL: THE CHALLENGE

- WILL WE BE ABLE OF GETTING THE BETTER OF HUMANS AND THE BETTER OF AI IN THE PUBLIC SECTOR?