Università degli studi di Trento

Law and Ethics in Artificial Intelligence

Prof. Casonato

Academic Year 2024/2025

Contents

AI Definition by AI ACT
Human Oversight
Duties of Deployers Vs Duties of Providers

Chapter 1

Ethics of AI

1.1 What is ethics

Ethics answer this question as follow:

- Well-founded standards of right and wrong that prescribe what humans ought to do: rights, duties, consequences...
- Questioning, discovering and defending our values, principles and purpose. Finding out who we are.
- Systematizing, defending, and recommending concepts of right and wrong conduct
- System of moral principles.

Three basic ethical approaches:

1. Consequentialism: effects/impact. What is matter is the consequence of my actions.
2. Deontology: duty/virtues-based ethics. I have to respect my duties, regardless the consequences of my actions. Examples: Trolley dilemma, Conjoined twins, AI(?)
3. Principlism: non maleficence, beneficence, autonomy, justice

Asimov's 3 (+1) Laws of Robotics (1942-1986)

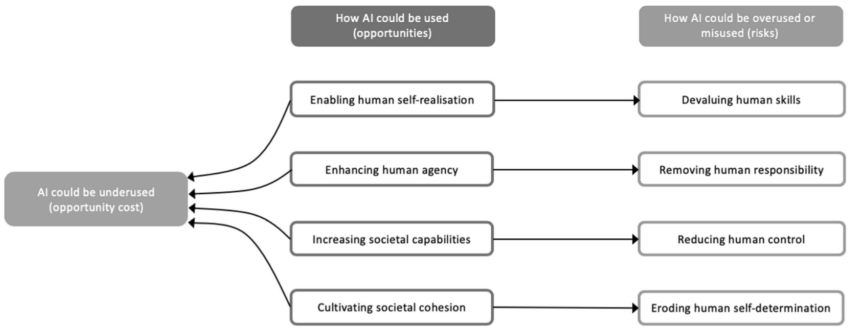
0. A robot may not injure humanity, or, by inaction, allow humanity to come to harm [non mal. - environment?].
   Law added later, it is the most important.
1. A robot may not injure a human being or, through inaction, allow a human being to come to harm [lethal aut. weapons?]
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law [autonomy?]
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws

LAWS: Lethal Autonomous Weapons Systems AI Act not covers the use of AI in military operations. This systems have no control (both on the use and on the control of these AI systems), no ethical rules, no human-mortality and no human-like intelligence.

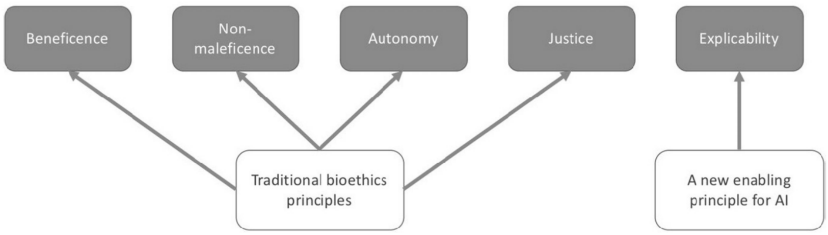
The target of ethical rules is the machine, but in this case there is another target: the human who have provided that machine. AI act focus more on this second target. A human should not allow a machine to kill another human.

Additionally, AI is too stupid to act in a complex environment like the military one.

AI uses



AI Principles (Luciano Floridi)



Beneficence Development of AI should ultimately promote the well-being of all sentient creatures. AI should be developed for the common good and the benefit of humanity.

Ensure that AI technologies benefit and empower as many people as possible.

Non-Maleficence Avoiding the misuse of AI. Prevention of infringements on personal privacy (linked to individuals' access to, and control over, how personal data is used). AI arms race (the autonomous power to hurt, destroy or deceive human beings should never be vested in AI).

The people developing AI, or the technology itself, should be encouraged not to do harm? (Frankenstein or his monster?).

Autonomy Machine have the right to make decisions for themselves but humans have to decide the limit of this right. Striking a balance between the decision: making power we retain for ourselves and that which we delegate to AI.

The autonomy of humans be promoted, but also the autonomy of machines should be restricted and made intrinsically reversible (where to draw the line?). Protecting the intrinsic value of human choice (significant decisions); containing the risk of delegating too much to AI.

Justice Promoting Prosperity and preserving Solidarity. Using AI to correct past wrongs such as eliminating unfair discrimination. Ensuring that the use of AI creates benefits that are shared (or at least shareable) and preventing the creation of new harms, such as the undermining of existing social structures.

Explicability Machine have to have understandable and interpretable intelligibility. Accountability: who is responsible for the way it works?
A small fraction of humanity is currently engaged in the design and development of a set of technologies that are already transforming the everyday lives of just about everyone else. Transparent responsibility in decisionmaking process (both tech and legal): interdisciplinarity.

Principles for robotic engineering EPSRC and AHRC, 2011

- Robots should not be designed solely or primarily to kill or harm humans. [**non human maleficence**]
- Humans, not robots, are responsible agents. Robots are tools designed to achieve human goals. [**autonomy**]
- Robots should be designed in ways that assure their safety and security. [**non AI maleficence**]
- Robots should not be designed to exploit vulnerable users by evoking an emotional response or dependency. It should always be possible to tell a robot from a human. [**justice, autonomy**]
- It should always be possible to find out who is legally responsible for a robot. [**justice, autonomy**]

Chapter 2

Digital Constitutionalism

Any society in which no provision is made for guaranteeing rights or for the separation of powers, has no Constitution.
- Art. 16, Déclaration des droits de l'homme et du citoyen, 1789

In order we lives in a constitutional we have to guarantee rights and have a limited power.

2.1 Limitation of power

Private vs Public Public dependent on AI (private): in decision-making, in essential public function.
Private chooses values/disvalues, exercising a public/normative/moral function.

Remedies:

- Macro: explicability/accountability
- Micro: IRBs/audits, reflective designing?

Algorithmic democracy, this means that people that develop AI systems who influence our life should be elected by people. This is the power that must be controlled.

2.2 AI & human rights

2.2.1 Old Rights

AI can generate images that create confusion because it's difficult to distinguish if they are real or fake. Most of the AI systems this days can pass the Turing Test. The most important right is the transparency: i must know my interlocutor's nature. This right is included in AI Act:

- Article 50: persons concerned are informed that they are interacting with an AI system

- Article 86: the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure right to explanation)

Another topic covered by AI Act is **explicability** (i have to understand the reason of the output):

- Article 13: High-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately

It is important to respect the **non-discriminatory principle**. To do this datasets must be relevant, sufficiently representative, free of errors and complete.

2.2.2 New Rights

AI Act provide an important principle: **Human in the Loop**, Human Oversight. This means that a human have the possibility to understand the output and possibles bias on it.

cerca

Automation Bias The propensity for humans to favour suggestions from automated decision-making systems and to ignore contradictory information made without automation, even if it is correct.

The second right is the **probabilistic approach**. I don't have the possibility to confront my point of view with one that is strongly different.

The third right is **King Midas issue**. Transformative effect on the environment: "Enveloping" the world?
The right to a human environment. Living environments built on a human scale.

Fourth right is the right to AI. The right to be judged, transported, examinated by AI.

Chapter 3

The law and/of Artificial Intelligence

Technology as social complex phenomenon needed for regulation (tailored, innovative, targeted). Special characteristics of regulatory object: ethical sensitiveness; social impact; scientific complexity and unpredictability; uncertainty; unexplicability (AI). Need to balancing individual rights and public interests at stake. We need new regulatory mechanisms.

A definition of law: Law like social product *"Where there is a society, there is the law"* Society is a group of individuals, a system of institutions and a set of legal rules. A legal system is a system of rules, created and enforced through social institutions, aiming to govern human behaviour.
How are rules produced?
Specific rules (procedures) which must be respected (validity)
Different channels and levels of production?
The sources of law (mechanisms to produce legal rules) – legal and institutional pluralism (national and supranational).
How do rules/sources interact? hierarchy (Constitution – parliamentary law) – scope of competences (State – EU).

Who produces the law?
- THE NATURE OF THE RULE-MAKER:
Public (expression of power) – private (expression of private autonomy)
- THE LEVEL OF THE PUBLIC LAW-MAKER
State (parliament, executive, courts, regions)
Supranational (EU, Council of Europe)
Global (United Nations)

Which kind of law?
- THE NATURE OF RULES/ACT
Binding – non binding (hard law-soft law)
- THE SOURCE OF PRODUCTION Constitution
European and international law
law of he Parliament (...)

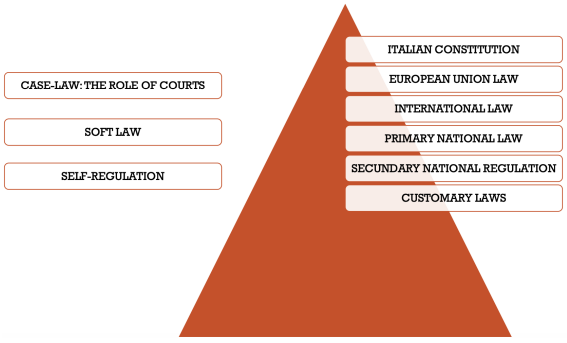
Example: META

As its community grew to more than two billion people, it became increasingly clear to the Facebook company that it shouldn't be making so many decisions about speech and online safety on its own. The Oversight Board was created to help Facebook answer some of the most difficult questions around freedom of expression online: what to take down, what to leave up and why.

The board uses its independent judgment to support people's right to free expression and ensure that those rights are being adequately respected. The board's decisions to uphold or reverse Facebook's content decisions will be binding, meaning that Facebook will have to implement them, unless doing so could violate the law.

The purpose of the board is to promote free expression by making principled, independent decisions regarding content on Facebook and Instagram and by issuing recommendations on the relevant Facebook Company Content Policy. When fully staffed, the board will consist of 40 members from around the world that represent a diverse set of disciplines and backgrounds. These members will be empowered to select content cases for review and to uphold or reverse Facebook's content decisions. The board is not designed to be a simple extension of Facebook's existing content review process. Rather, it will review a selected number of highly emblematic cases and determine if decisions were made in accordance with Facebook's stated values and policies.

There are different kind of laws: constitutional law, public law, private law, European law, international law,...



3.1 The national level: the law of the parliament

3.1.1 France: judges' profiling

"The identity data of magistrates and members of the judiciary cannot be reused with the purpose or effect of evaluating, analysing, comparing or predicting their actual or alleged professional practices."

Any healthcare professional who decides to use, for an act of prevention, diagnosis or treatment, a medical device involving algorithmic data processing learned from

massive data shall ensure that the person concerned has been informed and, where appropriate, warned of the resulting interpretation.

A better knowledge, through this means, of case law would promote equality between litigants. This would result in a breach of the principle of equality before the law and of the right to a fair trial.

To prevent, through the processing of personal data, profiling of legal professionals on the basis of the decisions rendered (pressure or strategies for choosing a jurisdiction likely to alter the functioning of justice). These provisions do not establish any unjustified distinction between litigants and do not infringe the right to a fair and equitable procedure guaranteeing the balance of the rights of the parties.

3.1.2 The Italian bill on AI

The Council of Ministers, `n. 78 of April 23, 2024`, approved a bill that regulates the use of artificial intelligence in the sectors entrusted by the Regulation to the normative autonomy of the Member State. The bill was presented to the Senate for approval on May 20 (AS 1146). The objective of the bill is to promote *"a correct, transparent and responsible use, in an anthropocentric dimension, of artificial intelligence, aimed at seizing the opportunities"* (Article 1) to improve the living conditions of citizens and social cohesion. `Article 3` contains the fundamental principles and stipulates that the life cycle of AI systems and models must be based on the respect of fundamental rights and freedoms of the Italian and European legal systems, as well as the principles of transparency, proportionality, security, personal data protection, confidentiality, robustness, accuracy, non-discrimination, and sustainability.

A Proposal for a Law on AI (Government). The Parliament is discussing the Proposal that must be approved by both the Chamber and the Senate. Once approved, the Act will coexist with the AI Act. *Which Act will prevail in case of clash?*

3.2 International Actors: Council of Europe (soft law)

NOTE: soft law means that it's a non-binding tool. They provide principles.

Who produces the law? The European Union law.

European commission for the efficiency of justice European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment.

The Charter is intended for public and private stakeholders responsible for the design and deployment of AI tools and services that involve the processing of judicial decisions and data. It also concerns public decision-makers in charge of the legislative or regulatory framework, of the development, audit or use of such tools and services.

3.2.1 Legal sources of EU

They are binding acts

- Regulations as the most important EU legal source (AI Act Proposal):
  - to ensure uniform application of the EU law in all Member States
  - They are directly enforceable in national legal systems
  - They prevail on national law in case of conflict (national law does not apply; duty to respect national fundamental principles)
- Directives do not have direct applicability:
  - Duty for States to achieve a result and a deadline for compliance (national law)

3.3 The EU Law On ETIAS

ETIAS, or the European Travel Information and Authorization System, is a system implemented by the European Union to enhance security and manage the entry of travelers from visa-exempt countries. Here are some key points regarding EU law on ETIAS:

Purpose: ETIAS aims to improve security within the Schengen Area by pre-screening travelers before they arrive. It is designed to identify potential security risks and facilitate border management.

Legal Framework: The legal basis for ETIAS is found in the EU regulations that govern the Schengen Area and border control. The system is part of the EU's broader strategy to enhance security and manage migration effectively.

Application Process: Travelers from visa-exempt countries will need to apply for ETIAS authorization online before their trip. The application will require personal information, travel details, and answers to security-related questions.

Data Protection: ETIAS is subject to EU data protection laws, ensuring that personal data is processed fairly and securely. Applicants will have rights regarding their data, including access and rectification.

Implementation Timeline: ETIAS is expected to be fully operational by 2024, with a phased implementation to ensure that all systems are in place for effective operation.

Impact on Travelers: Once ETIAS is implemented, travelers will need to obtain authorization before traveling to the Schengen Area, which may affect travel plans and preparations.

cerca profiling algorithms

3.3.1 ETIAS as profiling algorithm

ETIAS (European Travel Information and Authorization System) incorporates elements of profiling algorithms as part of its security and risk assessment processes. Here are some key points regarding ETIAS as a profiling algorithm:

Risk Assessment: ETIAS uses automated processing to assess the risk associated with travelers from visa-exempt countries. This involves comparing the information provided in the ETIAS application against various databases and predetermined criteria to identify potential security threats.

Data Utilization: The profiling algorithm analyzes data such as personal information, travel history, and responses to security questions. This data helps determine whether a traveler poses a risk to public security or is likely to overstay their visa.

Automated Processing: The system employs automated means to process applications, which allows for quick assessments. However, this also raises concerns about the accuracy and fairness of automated decisions, as they may rely on unverified data and predetermined criteria that could lead to false positives.

Human Oversight: If the automated system flags a traveler as a potential risk (a "hit"), the case will undergo a manual review by competent authorities. This two-step process aims to balance efficiency with the need for human judgment in assessing individual cases.

Legal and Ethical Considerations: The use of profiling algorithms in ETIAS raises important legal and ethical questions, particularly regarding privacy, discrimination, and the right to appeal decisions. The system must comply with EU data protection laws, ensuring that personal data is handled fairly and transparently.

Transparency and Accountability: It is crucial for ETIAS to maintain transparency about how profiling algorithms work and the criteria used for risk assessment. This helps build public trust and ensures accountability in the decision-making process.

In summary, while ETIAS employs profiling algorithms to enhance security and streamline border management, it also necessitates careful consideration of the implications for individual rights and data protection.

3.4 The charter of fundamentals right in EU

AI-related rights are:

1. Right to Privacy: Individuals have the right to privacy, which includes the protection of personal data collected and processed by AI systems. This right

is enshrined in various legal frameworks, including the General Data Protection Regulation (GDPR) in the EU.

2. Right to Non-Discrimination: AI systems must not discriminate against individuals based on protected characteristics such as race, gender, age, or disability. This right is essential to ensure fairness and equality in the deployment of AI technologies.

3. Right to fair Administrative Decisions: guarantees the right to fair administrative decisions by establishing principles of good administration, transparency, and accountability. These rights are essential for protecting individuals in their interactions with administrative authorities and ensuring that decisions are made in a just and equitable manner.

4. Right to fair Trail and independent Judge: right to a fair trial and the right to an independent and impartial tribunal.

Chapter 4

The AI Act

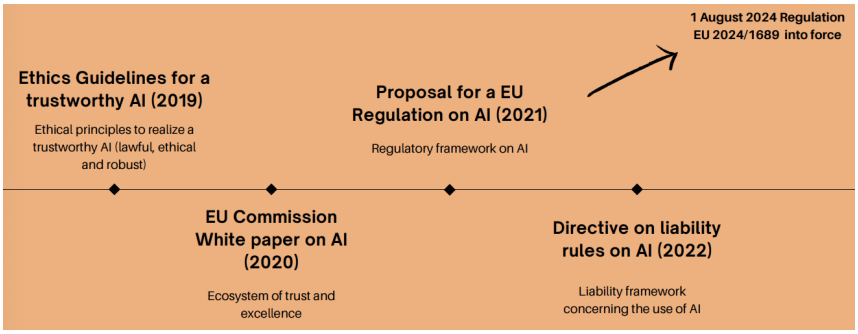


Figure 4.1: Timeline of AI Regulation in EU

AI Act is the important horizontal EU legal act laying down legal rules for AI. Introduce uniform rules across the EU on the development, production, marketing and use of AI systems within the EU single market in respect of the values protected and promoted by the Union.
Regulation: direct effect for all Member States and immediately binding.
New Legislative Framework (2008): creation and circulation of safe products.

The AI Act main goal

- Facilitate the creation of a single market for AI by removing barriers, both commercial and noncommercial, that could hinder the circulation of this technology within the EU (art. 114 TFEU);
- Ensure production and circulation of safe products, respectful of common values and fundamental rights protected by the EU;
- Promote innovation and development in this sector, ensuring legal certainty.

The subject of the AI Act

- Harmonised rules for the placing on the market, putting into service and use of AI systems in the EU;

- Bans certain AI practices;
- Specific requirements for high-risk AI systems and obligations for both providers and deployers of such systems;
- Harmonised transparency rules for specific AI systems;
- Harmonised rules to place on the market of general purpose AI models;
- Rules on market monitoring, governance of market surveillance;
- Measures to support innovation, with particular attention to SMEs, including start-ups;
- Penalties.

The structure of AI Act is based on two important frameworks:
Risk-based approach: rules are proportionally stricter on the basis of the prospected risk.
Conformity assessment and post-market control: Compliance with all the obligations provided for high-risk systems.

SHORTEN THE TRADE BARRIERS AND SIMPLIFY AI PRODUCTS CIRCULATION WITHIN THE EU MARKET BY ENSURING RESPECT FOR AND PROTECTION OF FUNDAMENTAL RIGHTS.

The scope of the AI Act

- Providers placing on the market or putting into service
- AI systems or placing on the market AI models for general purposes in the EU, regardless of whether they are established or located in the EU or in a third country;
- Deployers of AI systems that have their place of establishment or are located within the EU;
- AI providers and deployers that have their establishment or are located in a third country, where the AI output is used in the EU;
- Importers and distributors of AI systems;
- Manufacturers of products who place AI on the market or put it into service together with their product under their name and brand;
- People concerned who are in the EU

The Regulation does not apply to areas which do not fall within the scope of European Union law (EU competences). It does not affect the competences of the Member States in matters of national security;
`Example:` health protection in the strict sense (doctor-patient relationship).

There are fields of applications not covered by AI Act rules and obligations: AI systems developed and used only for military purposes, defence and national security; research activity on AI systems (except when the product is intended to be put into the market).

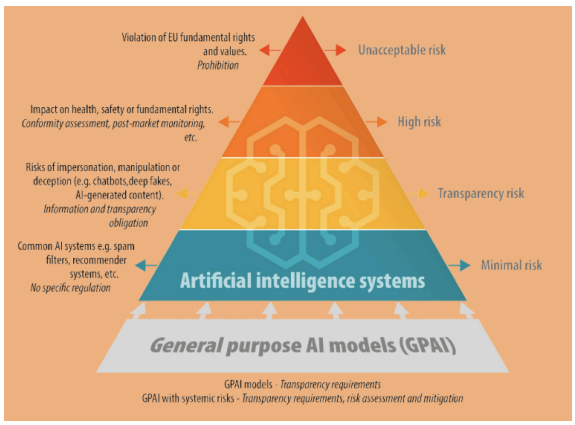
A future-proof approach foster innovation and keeping up with technological developments. Faster amendment procedures to modify some rules of the AI Act (High-risk systems categories, conformity assessment procedure, sandboxes).
Make the legal regulation more flexible and adaptable considering the nature of its object (AI technology).

The relevant definition within the AI Act 'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

'General-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market.

'Provider' means a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge;
'Deployer' means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal nonprofessional activity;

4.1 Risk-based Approach



An unacceptable risk is a clear and unacceptable threat to people health, safety

and fundamental rights. EU legislator decided to prohibit AI practices that could entail such kind of harms. It includes:

- Subliminal, manipulative techniques... distorting the behaviour of a person by impairing ability to make an informed decision, ... with significant harm;
- Exploitation of vulnerable persons (age, disability, socialeconomic situation), ... distorting the behaviour ... with significant harm (ex: dangerous games for children)
- Social scoring if:
  - detrimental treatment of persons in social contexts
  - unrelated to the contexts in which the data was originally collected disproportionate treatment
  - (E.g: parental reliability for placing children in the care of social services)
- Risk assessments: likelihood of committing a crime, based solely on the profiling (predictive policing). Not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity.
- Systems for creating facial recognition databases through the untargeted scraping from internet; Systems to infer emotions in workplace and education institutions. Except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons.
- Biometric categorisation systems to deduce race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation. Except for law enforcement reasons.
- The use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement, unless and in so far as such use is strictly necessary for one of the following objectives
  - the targeted search for specific victims, as well as the search for missing persons;
  - the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack;
  - the localisation or identification of a person suspected of having committed a criminal offence, for the purpose of conducting a criminal investigation or prosecution or executing a criminal penalty.

AI High-Risk systems within AI Act AI system is intended to be used as a safety component of a product, or is itself a product covered by the EU legislation listed in Annex I (medical devices). Product, whose safety component is the AI systems, or the AI system itself as a product, is required to undergo a third party conformity assessment (health and safety risks) under the relevant sectorial EU

legislation listed in Annex I.
AI systems falling under one or more of the critical areas and use cases referred to in Annex III have to be classified as high-risk ones.

- Biometric identification (exception art. 5)
- Critical infrastructure (traffic, digital infrastructure, supply of water, gas, heating and electricity)
- Education and vocational training (e.g. access to university; scoring of exams; monitor student's behaviours during exams)
- Employment, workers management and access to self-employment (ex. hiring and dismissal procedures)
- Access to and enjoyment of essential private and public services (healthcare; creditworthiness; life and health insurance)
- Law enforcement (evaluation of the reliability of evidence)
- Migration, asylum and border control management and predictions (automated evaluation of visa applicationa)
- Administration of justice and democratic processes (law application; election influence)

By derogation from paragraph 2, an AI system referred to in Annex III shall not be considered to be high-risk where it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making One of the following conditions must fulfilled:

- the AI system perform a narrow procedural task;
- the AI system improve the result of a previously completed human activity;
- the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review;
- the AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III.

4.1.1 Rules and requirements for AI High-Risk systems

Risk Management System Provision of a risk management system which has to consist in a continuous interative process concerning the entire lifecycle of the AI systems, with regular and systematic updating.

- Identification and analysis of known and foreseeable risks associated with the AI system
- Estimation and evaluation of risks that may emerge during the application of the systems for the intended purposes and under conditions of reasonably foreseeable misuse

- Evaluation of other possibly arising risks taking into consideration the data from post- market monitoring
- Adoption of suitable risk management measures (elimination or reduction of risks through adequate design and development, mitigation and control measures, provision of information)

Data and Data Governance AI systems shall be developed on the basis of training, validation and testing data sets, which has to be subject to appropriate data governance and management concerning:

- the relevant design choices;
- data collection
- relevant data preparation and processing operations (annotation, labelling, cleaning, enrichment and aggregation)
- the formulation of relevant assumptions a prior assessment of the availability, quantity and suitability of the data sets that are needed
- examination in view of possible biases and measures to detect, prevent and mitigate them
- identification of any possible data gaps or shortcomings

Nella miglior misura possibile

Training, validation and testing data sets shall be relevant, sufficiently representative, and to the best extent possible, free of errors and complete in view of the intended purpose. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used.

Data sets shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, contextual, behavioural or functional setting within which the high-risk AI system is intended to be used. Compliance with EU data law (e.g. privacy and use of personal data). Exceptional use of personal data to mitigate bias and discrimnation.

Technical Documentation and record-keeping Draw up technical documentation concerning the highrisk system before is placed into the market and keep it up-to date.

Technical documentation in order to demonstrate compliance with AI Act requirements.

High-risk AI systems shall technically allow for the automatic recording of events (logs) over the lifetime of the system in order to ensure a level of traceability of the functioning of a high-risk AI system, logging capabilities shall enable the recording of events relevant for identifying situations that may result in the high-risk AI systems and in facilitating post-market monitoring.

Trasparency Design and development of AI systems in a way to ensure that their operations are sufficiently transparent, in order to make AI users able to interpret the system's outputs and use it in a proper way.

The providers must give instructions for use that include clear, concise, complete and correct information.

The information has to be relevant, accessible and comprehensible to users.

The information concerning the AI systems must specify:

- identity and contact details of the provider
- characteristics, capabilities and limitations of performance of the AI systems
- all the changes to the AI system and its performance which have been predetermined by the provider
- the applicable human oversight measures and the technical measures put in place to facilitate the interpretation of AI outputs by the users
- computational and hardware resources needed, expected lifetime of the AI system and any necessary maintenance and care measures needed for a proper use and functioning
- a description of the mechanisms included within the highrisk AI system that allows deployers to properly collect, store and interpret the logs

Human Oversight Providers have to design and develop AI systems in order to ensure that AI functioning and operation can be effectively overseen by natural persons during their application.

Human intervention and oversight: preventing and minimising the risks to health, safety and fundamental rights which could emerge with the application of an high-risk AI.

The oversight measures shall be commensurate with the risks, level of autonomy and context of use of the high-risk AI system.

Human oversight measures have to be identified and built, when and if it's technically possible, into the AI system by the provider before the entrance in the market or its concrete application.

Accuracy, Robustness and Cybersecurity AI systems must have an appropriate level of accuracy, robustness and cybersecurity during all their lifecycle. Levels of accuracy shall be declared in the instructions for use.

AI systems shall be resilient to errors, faults or inconsistencies that may occur within the systems or the environment in which it operates or because of third parties attacks.

The robustness may be achieved through technical solutions (backup or fail-safe plans). Continuous learning systems has to be developed in a way to ensure that possibly biased outputs (due to biases in input data) are duly addressed with appropriate mitigation measures.

Providers obligations for High-Risk AI

- Ensure AI system compliance with AI Act requirements
- Ensure the quality of AI (quality system under art. 17)
- Keep relevant documentation and logs
- Ensure the systems undergoes the relevant conformity assessment procedure stated by the AI Act

Responsibilities along AI value chain: any distributor, importer, deployer or other third-party shall be considered to be a provider of a high-risk AI system if

- put their name or trademark on AI system
- make substantial modification of high-risk system
- modify the intended purposes of an AI system, also GPAI, not classified as high-risk
- not apply to third parties making accessible the AI systems under a free and open-source licence (also see art. 2 except art. 5 and art. 50)

Deployers obligation for High-Risk AI

- Take proper technical and organizational measures
- Assign human oversight tasks to natural persons
- If possible, control data input
- Monitoring AI functioning and reporting significant and serious events to competent authorities
- Inform natural persons that they are subject to the use of high-risk system for decision-making purposes

Fundamental rights impact assessment (art. 27): deployers that are bodies governed by public law, or are private entities providing public services, and deployers of high-risk AI systems evaluating access to financial loans or life and healthcare insurance, shall perform an assessment of the impact on fundamental rights that the use of such system may produce.

Conformity Assessment If provider has applied harmonized standards or common specifications, the provider shall opt for one of the following conformity procedures.

- Internal control (provider) on the basis of Annex VI;
- External control (notified body)

If substantial modifications are made, the AI must undergo a new conformity assessment.

4.2 AI systems Posing transparency risks within AI Act

Specific Transparency Obligations Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system, unless this is obvious from the point of view of a natural person who is reasonably well informed, observant and circumspect, taking into account the circumstances and the context of use.

Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated.

4.3 AI systems Posing Minimal Risks within AI Act

L'ufficio IA e gli Stati membri incoraggiano la stesura di un codice di condotta per favorire l'applicazione volontaria a sistemi IA diversi dai sistemi IA ad alto rischio di alcuni o tutti i requisiti previsti per quelli ad alto rischio.

Codes of conduct AI office and MS shall encourage the drawing up of code of conduct to foster the voluntary application to AI systems, other than high-risk AI systems, of some or all of the requirements set out for high-risk ones.

- EU ethical guidelines for trustworthy AI
- Assessing and minimising the impact of AI systems on environmental sustainability
- Promoting AI literacy
- Inclusive and diverse design
- Assessing and preventing the negative impact of AI systems on vulnerable persons or groups of vulnerable persons (accessibility for persons with a disability; gender equality)

Regulatory Sandboxes AI regulatory sandboxes provide for a controlled environment that fosters innovation and facilitates the development, training, testing and validation of innovative AI systems for a limited time before their being placed on the market or put into service.

Competent authorities shall provide guidance, supervision and support within the AI regulatory sandbox with a view to identifying risks, in particular to fundamental rights, health and safety, testing, mitigation measures, and their effectiveness in relation to the obligations and requirements of this Regulation and, where relevant, other Union and national law supervised within the sandbox.

Governance EU Level

- AI Office: compliance (64)
- AI Board: MS coordination and implementation (65)

- Advisory forum (67)
- Scientific panel of independent experts (68)

MS Level (70)

- Notifying authority
- Market surveillance authority
- (GPAI: Comm)

Ensure enforcememnt and implementation of the present Regulation.

AI Act Timeline 24 months from the date of entry into force; But...

- General provisions and prohibited practices: 6 months from entry into force;
- Notifying authorities and notified bodies, IA for general purposes, governance, sanctions: 12 months from entry into force;
- High-risk systems: 36 months

Chapter 5

General purpose AI Models Regulation

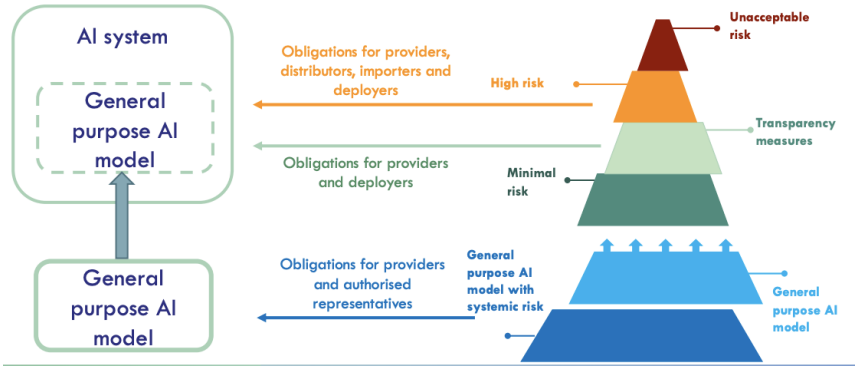


Figure 5.1: Integrating AI Models into the Risk Pyramid

AI Model Including where such an AI model is trained with a large amount of data using self-supervision at scale displays significant generality is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market can be integrated into a variety of downstream systems or applications.

Scope and Exceptions *"Placing on the market"* means the first making available of an AI system or a general-purpose AI model on the Union market.
"Making available on the market" means the supply of an AI system or a general-purpose AI model for distribution or use on the Union market in the course of a commercial activity, whether in return for payment or free of charge.
Exception: AI models used for research, development or prototyping before being released on the market.

Identification of General Purpose AI Models for with Systemic Risk The Commission publishes and updates a list of systemic risk models. A model is so classified if it has a high impact capabilities on the basis of appropriate technical

tools and methodologies that will be determined by the Commission. It presents equivalent capabilities or impact on the basis of a decision of the Commission.

5.1 General Purpose AI Models

5.1.1 Requirements

1. Drafting and updating the technical documentation which will be forwarded to the authorities.
2. Drafting and updating of technical documentation for downstream suppliers of AI systems wishing to integrate the mode.
3. Implementation of policies and procedures (including automated ones) to comply with union copyright law.
4. Drafting and making available to the public a sufficiently detailed summary of the content used for training.

Requirements 1 and 2 do not apply for AI models released under a free and open source licence that permits access, use, modification and distribution of the model and whose parameters, including weights, model architecture information and model usage information, are made publicly available (partial exemption). Requirements 3 and 4 regarding copyright still apply. The exception does not extend to models with systemic risk.

Systemic Risk: risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain.

5.1.2 General Purpose AI Models with Systematic Risk

1. Obligation for general purpose models
2. Evaluation of models using state-of-art protocols and tools
3. Assessment and mitigation of possible systemic risks at union level
4. Tracking and documenting relevant information on serious incidents and possible corrective measures
5. Adequate level of cybersecurity

Chapter 6

AI, biases and discrimination

6.1 Technology

6.1.1 Why is technology not neutral?

Unequal distribution of power and privileges brings to forms of oppressions: economy inequalities, sexism, racism, homophobia,...

AI poses a potential risk of perpetuating biases and forms of discriminations.

AI is often regarded as a neutral tool. Two concept will help us answering this question: duality of technology and politics of technology.

Duality of technology Technology is both physically constructed within a social context and socially constructed through the meanings attributed to it by various actors.
It entails:

- Technology is the product of human actions: outcome of human in design, development, modification, and use.
- Technology is the medium of human actions: it facilitates, but also it constrains human actions.
- Institutional condition of interactions with technology: condition that influence humans in the interaction with technology.
- Institutional consequences of interaction with technology: transformation of organizational structure in terms of dominance.

Politics of technology Two ways in which technological artifacts contain political properties (Winner, 1980):

1. They denote a certain kind of arrangements of power and authority within society: the design or specific properties of artifacts may cater to certain kind of social interests, while excluding the others.
2. They may encompass (COMPRENDE) more rigid forms of politics whereby they either require or are compatible with certain kind of power arrangements. Such technolo-

gies are often called inherently political. Think of AI system used by public authorities to guarantee security or to take public welfare decision.

Transformative effect of social and political factors on AI Social construction of technology + politics of technology =

- The concept of "social construction of technology" and the idea of "politics of technology" remind us of the transformative effect of social and political factors on AI.
- As our society and culture implicitly harbor biases, these biases permeate the design, use, and understanding of AI technologies, thus embedding them with the same biases, prejudices and power imbalances leading to discriminatory practices.

6.1.2 Why does AI discriminate?

AI-driven discriminations can occur due to biases at multiple stages. Several typologies have been suggested to describe how AI can be biased.

Kleinberg identifys risks of discrimination in the: choice of outcome, input information, training procedure.

Barocas and Selbst identify risks of discrimination in the: target variables, training data, relevant features, proxies, and intentional discriminations.

It can be said, also, that biases derive from:

- Biased choice of outcome and targets;
- Poor quality and lack of representativiness on the training data.
- Lack of diversity in the discipline → (cfr slide 26 on D&I policy)

Those sources of biases can be categorized in two types of inequality: symbolic inequality (inequal representation of groups in society) and distributive inequality (nequal distribution of social goods among individuals due to their belonging to a social group).

6.1.3 cases of AI discrimination

In order to better understand how biases can be injected into AI tools we will look at three cases dealing with: ethnicity, sexuality and gender.

Predictive policing tools and the race effect: PredPol Predictive policing software predict: where a crime might occur in a given space and time: place-based software; who might be involved in committing crime: person-based software. The police uses these tools to guide its law enforcement activities (e.g. patrolling areas).

A place-based software in use to several US Police Departments. It relies on historical data about where and when previous crimes have occurred by using police records data.

The use of such tool may exacerbate issues of discrimination against ethnic minorities. Even though ethnicity is not a data taken into consideration by the model, the software can have asymmetric effects on different population groups.
RACE EFFECT: Even though the types of crimes considered occurred in areas with both predominantly caucasian and predominantly african-americans or hispanics, the algorithm suggested patrolling in the latter areas at twice the rate of the former. The bias in the input data thus reproduced itself in the output, confirming the statistical rule: "garbage in - garbage out".
RACE-IN-THE-LOOP EFFECT: The implementation of such tools generated a loop effect, whereby patrols were repeatedly sent to the same neighborhood perpetuating and reinforcing the distortion (bias) of the input data. This was due to the fact that in our place-based model recent criminal events weight more than past events.

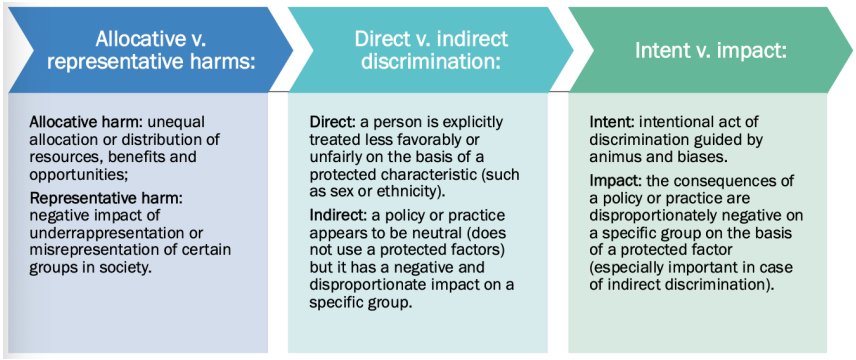
Algorithmic content moderation and LGBTQ+ speech: PerspectiveAPI PerspectiveAPI is a software used to identify, classify and eventually remove textual content posted on social network on the basis of their level of "Toxicity". The research shows that the software considers the drag queens account to be more toxic than the suprematists ones. This has a negative and discriminatory impact on LGBTQ+ community speech.

AI-Hiring systems and the gender gap: Amazon hiring tool In 2018 Amazon developed an AI system to automate the evaluation of job applicants' resumes. This system was to be trained on a large dataset of resumes submitted to Amazon over a decade, with the goal of creating a program capable of reviewing resumes and identifying top candidates. The AI system developed a bias against women. The training data used by the algorithm contained a significant number of resumes from male applicants. As a result, the algorithm learned to associate certain keywords, phrases, and experiences with male candidates, leading to a bias against female applicants. The bias occurred because the algorithm essentially replicated the patterns it observed in the data it was trained on. Since there was a gender imbalance in the historical data, with fewer resumes from women, the algorithm failed to effectively account for the skills and qualifications of female candidates.

6.2 Technology and the Law

6.2.1 What is discrimination in legal theory?

From a legal perspective, discrimination is generally understood to be a violation of the principle of equal treatment: less beneficial or disadvantageous treatment given to some individuals on the basis of a specific protected ground, such as race, religion, sex, gender, sexuality.



6.2.2 Why does the principle of nondiscrimination fail in confronting AI biases?

Three problems emerge when we try to apply the principle of nondiscrimination to AI biases:

- The Problem with Representative Harms:
  - Allocative harm means the unequal distribution of resources. This means that a biased algorithm unfairly distribute resources among individual based on a specific factor (in the Amazon case the factor would be gender): easier to prove and legally actionable.
  - Representative harm occurs when systems reinforce subordination of a group and it is cause by sterotyping, denigration, underapresentation of that specific group. This harm is hard to detect, formalize, and prove (this is the case of LLM program used for content moderation) → As a matter of fact, this type of harm would generally be not legally actionable, leaving discriminated people without legal protection.
- The Problems with Direct and Indirect Discrimination:
  - Direct discrimination is rare in AI as it involves explicit unfavorable treatment based on protected characteristics, which AI models typically avoid. However, black-box algorithms make it challenging to determine whether a protected factors has been used.
  - Indirect discrimination is more common in AI but has limitations, including difficulties in proving it due to complex intersectional categories, exception clauses justifying AI use, and assigning responsibility for the discrimination.
- The problem with the use of proxies
  - Proxy discrimination occurs when AI systems use variables correlated with protected characteristics (e.g., ZIP codes for racial discrimination).
  - AI can indirectly discriminate through proxies that closely relate to those characteristic.

6.2.3 How do the EU AI Act solve the issue of AI discrimination? (And how it does not?)

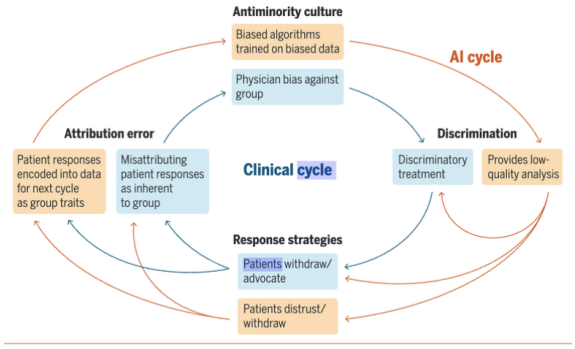
When AI system is qualified as a high-risk because of its effect on fundamental rights, some obligations provided by the AI Act towards providers shall be implemented. First of all, there are some compulsory– although generic - requirements indicated as mandatory in order to mitigate the risks for fundamental rights violations: "quality of data sets used" (art. 10); "technical documentation" (art. 11) and record-keeping (art. 12); transparency and provision of information to users (Art. 13), human oversight (Art. 14), accuracy (art. 15). In order to comply with these requirements, the AI Act contains both ex-ante and ex-post obligations. With regard to the risks of discriminations three obligations seem to advance the idea of biases mitigation as a mean to reach nondiscrimination.

Chapter 7

AI and medicine

7.1 Case Study: GP at Hand (eMEd)

Patients use an app called Babylon to seek help. They give all their data and in this way they pass from public health system to this.
In this way the patient can have a h24 assistance, a timely examination and enhanced relationship (facial recognition). This method could cut off the most vulnerable part of patients (only 6% have an age>45). This is AI biased and can lead to biased results, with both overestimation and underestimations of risk.
The bias includes race-based errors: more attention or resources to white patients than to members of racial and ethnic minorities.



AI Act (art.10) "Training, validation and testing data sets shall be relevant, sufficiently representative, and to the best extent possible, free of errors and complete in view of the intended purpose. Appropriate statistical properties".

AI Act (art.13) Operation: sufficiently transparent to enable deployers to interpret a system's output and use it appropriately.
Human oversight: understand, interpret, reverse (automation bias).

AI Act (art.86: hish-risk) Right to explanation of individual decision-making: the right to obtain from the deployer clear and meaningful explanations of the role

of the AI system in the decision-making procedure and the main elements of the decision.

Loi de bioéthique, art.17 A health care provider who decides to use ... a medical device that includes algorithmic processing ... must ensure that the person concerned has been informed and that is, where appropriate, informed of the resulting interpretation.

The pro for the doctor are schedule flexibility, work location and more time for other patients. In the other hand it can be lead to a de-skilling and an underestimation of his academic preparation.

Chapter 8

AI and Public Administration

8.1 Public administrations' tasks and powers

The use of AI by public administrations as a heterogeneous and cross-cutting phenomenon: variety of sectors, wide spectrum of powers and types of acts-decisions, different types of AI employed (or experimented). Examples:

- Monitoring compliance with regulatory requirements (instrumental to enforcement activities, including sanctions) (assessing frauds, singling out potential violators, etc.)
- Public security and law enforcement functions (biometric, mainly facial, recognition; verification and identification; border control, police, risk prediction)
- Urban planning
- Adjudicatory activities
- E-public procurement
- Enhancing public engagement and participation
- AI-assisted and semi-automated public services

A distinction particularly relevant for the use of AI systems by public authorities:

- AI used within the proceeding for preparatory functions
- AI issuing the final decision (decision-making function)

8.2 Legal coordinates

Public law regime, public values and guarantees:

- Provide reasons, ensure transparency vs "black box"
- Participation, due process, right to be heard
- Anchoring effect vs power to decide entrusted to the public official

Expert systems and model-based AI (*if-then* inference based on hard rules)	Machine learning applications (substantive digitalization of the public administration) and general purpose AI
(formal digitalization - automation of the public administration) Examples: • speed cameras-autovelox • Recruitment proceedings	○Supervised and unsupervised ○Data available to the public sector ○Different data as input: •Only text •Images •Audio •Structured?

Figure 8.1: Distinctions in the public sector uses of AI-systems

- Privacy vs anonymization and "blinding" strategies
- Access and transparency vs intellectual property
- Liability-related issues

8.3 Which kind of regulation for AI in the public sector?

Different approaches on the regulation of AI around the world, in search for a balance between: protection of rights and public values and promotion of AI and its advantages. General diffusion of a risk-based approach to regulation, but different levels of accepted risks. Soft-law (National strategies on AI, Guidelines, ethical principles).

8.4 The AI Act and the public sector

AI is regulated by the European Union as a risky product: risk-based approach. General scope of application of the AI Act is that there is no legal definition of AI-assisted administrative decisionmaking.

Most AI systems used in the public sector fall within article 3 definition of artificial intelligence: machine-based system, designed to operate with varying levels of autonomy, may exhibit adaptiveness after deployment for explicit or implicit objectives, infers, from the input it receives, how to generate outputs.
The AI Act applies to several uses of AI within the public sector.

Prohibited AI-systems Article 5 list examples relevant for the public sector:

- Real-time biometric identification in public spaces for law enforcement activities (with possible exceptions – targeted searches)
- Unsupervised image-scraping (see Clearview case)

- Predictive policing solely based on profiling an individual
- Social scoring

Low risk systems and systems featuring transparency risks art. 50: "Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant, and circumspect, taking into account the circumstances and the context of use. This obligation shall not apply to AI systems authorized by law to detect, prevent, investigate or prosecute criminal offenses, subject to appropriate safeguards for the rights and freedoms of third parties unless those systems are available for the public to report a criminal offense"
e.g.: chatbots and virtual assistants often used by public administrations in their communications with citizens.

Annex III high-risk AI systems

- Remote biometric identification systems in public spaces, where permitted by law.
- AI systems as safety components in managing critical infrastructure like digital networks, road traffic, or utilities such as water, gas, heating, or electricity.
- AI systems in education and vocational training, particularly those used for determining access, admission, or assignment to institutions.
- AI in employment and workforce management, including systems used for recruitment, job advertising, application filtering, and candidate evaluation.
- Access to essential services: AI used by public authorities to assess eligibility for public services (e.g., healthcare, benefits).
- Law enforcement: AI used for assessing the risk of a natural person offending or re-offending not solely on the basis of the profiling of natural persons, or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups
- Migration and border control: AI used for assessing risks related to security, migration, or health posed by individuals entering the EU.
  - polygraphs or similar tools;
  - AI systems intended to be used to assess a risk, including a security risk, a risk of irregular migration, or a health risk, posed by a natural person who intends to enter or who has entered into the territory of a Member State
- Administration of justice and democratic processes

Many tasks exercised by public administrations involve high-risk AI systems, however, the use of AI in a sector falling within Annex III does not automatically classify the system as high-risk: exceptions where a system does not play a determinative role in decision-making and does not pose a "significant risk of harm to the health, safety, or fundamental rights of natural persons". (Article 6(3)):

- It performs a narrow procedural task
- It improves a result previously determined by human activity.
- It detects patterns or deviations but does not influence the final decision without proper human review.
- It performs a preparatory task related to the use cases listed in Annex III.

The distinction between a system that substantially influences a decision and one that plays a limited procedural role is not always easy to establish (see automation bias – anchoring effect).

The European Commission is expected to issue guidelines (February 2025) to clarify these and assist with the practical interpretation of these provisions.

8.4.1 Requirements for high-risk AI systems

- Risk Management System (Article 9): continuous identification, evaluation, and mitigation of risks
- Retention of Logs (Article 12): continuous monitoring and review of the system
- Technical Documentation (Article 11): ensuring that the system can be assessed for compliance with the regulation's requirements
- Accuracy, Robustness, and Cybersecurity (Article 15)
- Data Governance (Article 10)
- Transparency (Article 13): providers must ensure high-risk AI systems are designed and developed in such a way that their operations are sufficiently transparent. This transparency is then essential for deployers — especially public administrations — so that they can understand, appropriately interpret, and explain the system's outputs. High-risk AI systems must come with clear instructions for use (concise, clear, complete, correct, accurately reflecting how the system works)
- (article 26(1)), the deployer must actively take steps to ensure that the system is used in line with the provided instructions. Ongoing monitoring of the system's performance.
- Human oversight (Article 14): oversight measures proportionate to the risks, level of autonomy, and context of use of the system. Measures can be incorporated into the system itself, if technically feasible, or implemented by the deployer (the public officer). The public officer tasked with overseeing the

system must: properly understand the relevant capacities and limitations of the high-risk AI system and monitor its operation, including detecting and addressing anomalies, dysfunctions, and unexpected performance; remain aware of the potential tendency to rely excessively on the output (automation bias), especially for systems used to provide information or recommendations; correctly interpret the output of the system; decide not to use the high-risk AI system or to disregard, override, or reverse its output; intervene in the operation of the high-risk AI system or stop the system.
There are some issues concernign human oversight by public officers: the ability to supervise requires a level of competence (AI literacy); the actual implementation of human oversight mechanisms also depends on how the responsibility/liability of the public officer is regulated in case of harmful events (both if he disregards the outputs, and if he follows the system's recommendation, resulting in a damage or loss).

8.4.2 AI impact assessment

An assessment of the impact on fundamental rights of high-risk AI systems is required exclusively for uses in the public sectors (deployers that are bodies governed by public law or private entities providing public services)

- identifying any potential risk to fundamental rights that may arise from the AI system's deployment;
- evaluating the system's impact on the protection of fundamental rights (privacy, non-discrimination, equal treatment);
- mitigating potential risks by taking measures to prevent, reduce, or eliminate harmful impacts on fundamental rights.

once the assessment has been performed, the deployer shall notify the market surveillance authority of its results, submitting a filled-out pre-established template.

8.4.3 The right to explanation of individual decision-making (art. 86)

"Any affected person subject to a decision which is taken by the deployer on the basis of the output from a high-risk AI system ... which produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights shall have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken".

No specific remedy in the AI Act to contest an administrative decision issued with the aid of an AI system.

8.5 Principles emerging from case-law

Few written rules currently address the use of AI by public authorities. Judicial review of AI-assisted administrative decisions poses different problems depending on the algorithm used. Difficulties with applying traditional mechanisms designed to ensure public law principles and guarantees (due process, participation, duty to give reasons, transparency etc.) in an "algorithmic environment". Not simply apply/translating traditional rules.

The "new paradigm" of AI-assisted administrative decision-making

- explainability, algorithmic transparency: provide information about the employed AI systems.
- non-exclusivity, human oversight: ensure that final decisions are traced back to the administration and to a specific civil servant (formal imputation)
- Algorithmic non-discrimination

8.5.1 The first legislative codification of the "new paradigm"

To improve efficiency, contracting authorities shall, where possible, automate their activities using technological solutions, including artificial intelligence.
In purchasing or developing the solutions contracting authorities:

- ensure the availability of the source code, the related documentation, as well as any other element useful for understanding the operating logic;
- introduce into the documents calling for tenders, clauses aimed at ensuring the assistance and maintenance services necessary to correct errors and unwanted effects deriving from automation.

Decisions taken through automation must respect the principles of:

- knowability and comprehensibility, whereby every economic operator has the right to know the existence of automated decision-making processes that concern him and, in this case, to receive significant information on the logic used;
- non-exclusivity of the algorithmic decision, so that in the decision-making process there is in any case a human contribution capable of controlling, validating or denying the automated decision;
- algorithmic non-discrimination, whereby the owner implements adequate technical and organizational measures in order to prevent discriminatory effects towards economic operators.

Contracting authorities shall adopt all technical and organizational measures to ensure that the factors leading to inaccuracies in the data are rectified and the risk of errors is minimised, as well as to prevent discriminatory effects against natural persons on the basis of nationality, ethnic origin, political opinions, religion, personal beliefs, trade union membership, somatic characteristics, genetic status, state

of health, gender or sexual orientation.

Public administrations publish on the institutional website, in the "Transparent Administration" section, the list of technological solutions referred to in paragraph 1 used for the purpose of carrying out their activities.

Chapter 9

AI and Sustainability

AI is a technology that fits present and future society.
We need an approach that goes further than the present day and also looks to the future society/ environment/ generations/...
The possible impacts include energy and water.

Why is it relevant for a lawyer? to regulate these aspects (quantitative data is always the key).

Bad side: it's difficult to estimate the energy consumption at the moment. The problem is the uncertainty of data. But important to be accurate because data is the aspect that could stop AI development.
Generative models consume a lot more energy than others.
AI needs water to cool down.
To regulate the building of new data centers.

Good side: Use AI to detect future climate changes.
Mapping deforestation.
Previsions on natural disasters (climate forecasting).
Detect iceberg melting.

Constitutional paradigm shift: New rights (example the right to water) didn't exist because it was irrelevant. In the past was a common interest to have an healthy environment, now it has became a right.

If something is a right there's a way to move it effective before a court. These rights are only at international level and not directly related to health.

New generations are essentially the new rights' holders. The problem is: how can someone that still not exist be a right-holder?
We're trying to create duties of protection for future generations.
The goal is to respect alive people's rights, and future generations' ones.
There is a reference of environment protection in the AI Act.

AI and Global Social Justice: exploit of poorer countries by global north. Data protection (es. META steals sensible data from countries with a few laws for data protection).

Chapter 10

RESPONSABILITÁ

AI and Liability

10.1 Some introductory remarks

Artificial Intelligence implies a collection of data and materials and/or rules for training. The existence of decisions involving legal effects on the natural person.

Possible scenarios for liability are: facial-vocal recognition systems, generative AI systems, medical examination, self-driving cars and other transport systems,...

10.2 Liability for violation of personal data

Concepts of "personal data" and "unlawful processing".
There are different rules in different legal frameworks, e.g. United States: fragmented approach, federal and national rules + tort law; European Union (EU): Regulation 2016/679 (GDPR) + Directive 2002/58 e-privacy.
In the European Union GDPR applies in the contexts of its territorial and material scopes. The first problem are MIXED Dataset (personal + non personal data) and the possibility to separate or not the set of information. When the processing concerns personal data, data protection rules apply. Liability regulations are unveiled in artt. 77-84 of the GDPR. Rights related to the violation of personal data are:

- Right to lodge a complaint with a supervisory authority + Right to an effective judicial remedy against a supervisory authority → High administrative sanctions
- Right to an effective judicial remedy against a controller or processor (e.g. the AI developer, the platform using AI, the entity that processes personal data...)
- Right to compensation and liability

Right to compensation and liability Existence of an infrigment of any provision of the GDPR.
Existence of material or non-material damages caused by the data processing.
Causal link between infringement and damages. A controller or processor shall be exempt from liability if it proves that it is not in any way responsible for the event giving rise to the damage.

Liability for violation of personal data in Italy Reclamo o segnalazione to the Garante. Civil Action: tribunale, alternative jurisdiction between residence of data subject and the establishment of the data controller.
Burden of proof:

- For the data subject: existence of damage and causal link between unlawful processing and damage
- For the data controller: proving that it is not in any way responsible for the event giving rise to the damage

Examples of violations Application of the principles of lawfulness (e.g. absence of legal basis for input, output and training, "web scraping"), transparency and fairness (e.g. lack of disclosure of information in the privacy policy) - artt. 5, 6-9, 13-14 GDPR.

10.3 Liability for infringement of intellectual property

Access to assets protected by forms of intellectual property by the AI, e.g. for training (e.g. generative AI). Liability for infringement of various rights, including copyright on images, text, music, etc., hence risk of plagiarism; patents; trademarks. Liability rules vary depending on the legal system and the right infringed. Damages and burden of proof are important issues: *is it possible to prove that the AI used the materials (e.g. for training)? Are there any exceptions to the IP rights for use of works to train AI?*

Cases in the US: Andersen vs. Stability AI Class action of visual artists for the use in training of protected works (5 billions images), without a licence, but obtained by web scraping. AI replicating the "artist's style" of the artists to create "new" images". Action for copyright infringement and damages. 12 August 2024 order of the court to dismiss, but still pending. Problem: proving the infringement.

Fair Use Section 107 of the Copyright Act: purposes such as criticism, comment, news reporting, teaching, scholarship, and research + 4 Factors for the test:

1. Purpose and character of the use
2. Nature of the copyrighted work
3. Amount and substantiality of the portion used in relation to the copyrighted work as a whole
4. Effect of the use upon the potential market for or value of the copyrighted work

Liability in the EU

- Exception to the rights for text and data mining carried out by research organizations and cultural heritage institutions for scientific research.
- Exception to the rights for data mining carried out when the data which was "mined" was accessed lawfully and the copyright owner has not expressly prohibited the use of the text/work for the purpose of data mining.
- Providers of AI generative models will be required to provide a detailed summary of the content used for the training, in a comprehensive way that will allow copyright or parties with legitimate interests to exercise and enforce their rights under EU law.

10.4 Liability for defective products

Damage generated during the use of a defective product with AI to natural persons. Exonerating proof by the producer: the defect did not exist when the product was placed on the market or was not recognizable on the basis of the technical and scientific knowledge. When the claimant faces excessive difficulties, due to technical or scientific complexity, in the burden of the proof, there might be presumptions or duty to disclosure for the producer.

10.5 Other forms of liability

10.5.1 Non-contractual liability

How to prove fault on the part of the AI?

- Fault-based liability: injured party has to prove that the defendant caused the damage intentionally or negligently
- Strict-based liability: injured party only needs to prove that a risk materialized
- Product-based liability: victims can claim for a defect present at the time the product was placed into market

10.5.2 Self-driving cars

- Different levels of autonomy
- Vehicle liability is governed by Article 2054 of the Civil Code
- For highly automated vehicles there is the need to an overall rethinking of the liability system + Product liability for manufacturing defects remains in place

s