Domande law and ethics

1. The most problematic issue we had to address was on Transparency and how it relates to the black box model.

The transparency principle is a core idea of trustworthy AI, and it exists to protect people from misuse of faults from AI systems. Transparent AI systems are easier to analyze, and it is also easier to find errors and biases within the model, therefore begin easier to correct.

In the Art. 13 of the proposal of regulation on AI (also known as AIA) describes transparency as a properly related to High-Risk AI Systems (HRAI).
Every HRAI should be transparent to enable users to:

a. Interpret system's output (Explicability)
b. Use it appropriately
   To be transparent it should:
c. Come with a set of instructions
d. Specify the contact and identity of the producer
e. Specify limitations and risks relate to its usage
f. Specify the oversight measure to be implemented (Right to be subjected to human decisions)

Many Data Mining and Machine learning techniques rely on a black box model, whose internals are either unknown to the observer or they are known but uninterpretable by humans.
Because of this black box, which is not understandable, we could unintentionally (or intentionally) create harmful AI that for example discriminates against vulnerable groups of people. Because the decision of the AI is not interpretable by humans, we would have had time understanding that the AI system is biased and unfair.
In our case we found a fundamental clash with the requirements for transparency and some AI models. The regulation fails to consider that for high-risk AI based in complex machine learning techniques that actively do use a black box model, are nigh impossible to explain, even with techniques like saliency maps, and it would stop all research based on black boxes for sensitive AI. On top of that many businesses are entirely based on secrecy of their algorithm and the AI Act would require a change in business model.

Ai it is right now, the AI Act proposes we ban all non-transparent high-risk AI even if collectively beneficial, but we don't know if other country would do the same.

2. The 5 principles declared by AI4People by Floridi derive from mostly from bio-ethic. These guidelines are the result of an analysis of several whitepapers written in the last years and by many different entities.

The 5 principals are:

1. Benevolence
2. Non-maleficence
3. Autonomy
4. Justice
5. Explicability

I personally think that autonomy is the most important ethical principle.

As Ai systems keep growing more and more ubiquitous in our life, and with their increasing accuracy and performance, they are slowly but surely superseding humans. However, the autonomy of humans should be empowered by AI systems, not reduced. A human will always have the right to decide whether to delegate his decision to a machine and therefore he should always be able to take control of the decision-making process.

This principal is also known as human in the loop (HITL) and we already have some regulations on the principal, as both in the GDPR and the AI act try to address in some way this key issue.
For example the AI act states that human in the loop is required for all high-risk AI systems because of the inherent risk they pose to our society and the consequences they have in case of errors or faults; we are not ready yet to delegate these responsibilities to heartless machines, particularly if we do not know how they came to their decision.

Floridi's approach to AI:

- Ethical principles,

The 5 ethical principles of AI are:

1. Beneficence
   a. Creating AI technology that is beneficial to humanity from Montreal and IEEE well-being
2. Non-maleficence
   a. Do no harm which is different from only beneficence as for example in privacy
3. Autonomy also called the power to decide whether to decide
   a. Not only should the autonomy of humans be promoted, but also the autonomy of machines should be restricted and made intrinsically reversible
   b. Meta autonomy or decide to delegate
4. Justice
   a. No unfair discrimination
5. Explicability
   a. To understand and hold to account the decision-making process of AI

- Opportunities and Risks for Society,

AI can be used (opportunities) to foster human nature at its potentialities creating opportunities, it can be underused thus creating opportunity cost or overused (risk) and misused. This could be caused by heavy handed or misconceived regulation, under-investment, or a public backlash (example genetically modified crops)

- Dual Advantage of an Ethical Approach,

With an analogy is the difference between playing according to the rules, and playing well, so that one may win the game. Applying an ethical approach to AI confers a dual advantage. On one side ethics enables organizations to take advantages of the social value that AI enables. This are the ability to identify and leverage new opportunities that are socially acceptable or preferrable. On the other side ethics enables organizations to anticipate and avoid or at least minimize costly mistakes. This also lowers the opportunity costs of choices not made or options not grabbed for fear of mistakes. Only work in socially acceptable risk to benefit situations

- Recommendations and Action points for a Good AI Society

AI should be designed and developed in ways that decrease inequality and further social empowerment, with respect for human autonomy, and increase benefits that are shared by all, equitably. We also believe that creating a Good AI Society requires a multistakeholder approach, which is the most effective way to ensure that AI will serve the needs of society, by enabling developers, users, and rule-makers to be on board and collaborating from the outset.

Points:

1. Assessment (valutare)
    - Capacity of existing institutions
    - What cannot be delegated to ai
    - Current legislations
2. Development (sviluppare)
    - Explicability
    - Secure algorithms for court
    - System for unwanted behaviors
    - Mechanism for remedy or compensate grievances
    - Guidelines
    - Liability insurance mechanism
    - Agreed metrics of trust
    - EU oversight agency
    - Legal instruments and contractual templates
3. Incentivization (incentivare)
    - Money for ai
    - Money for research
    - Cooperation cross sector
    - Inclusion of ethics in AI
    - Perception and understanding
    -
4. Support (supportare)
    - Self-regulatory code
    - Take responsibility for ethics
    - Education

The AI act (and introductory papers):

- **Definitions**

Definition of ai in the ai act: Software that is developed with one or more of the techniques and approaches listed in Annex I (machine learning, logic and knowledge based, Statistical approaches) and can generate outputs such as content, predictions, recommendations, or decision influencing the environment they interact with.

- **Risk-based approach**

It poses a risk of harm to the health and safety, or a risk of adverse impact on fundamental rights

- **Prohibited systems**

Social scoring, Facial recognition dark pattern ai, manipulation

1. Subliminal techniques
2. Exploitation of vulnerabilities of a specific group of persons
3. Social Scoring

- **Requirements for high-risk systems**

1. Transparency
2. Instruction for use
3. Human oversight
4. Consistency and High level of accuracy
5. Resilient
6. Quality management system
7. Conformity assessment (internal or external)

- **Classification of AI systems as High-Risk**

Ai system as safety component of a product or product such as:

1. Biomedical identification of natural person
2. Management and operations of critical infrastructure
3. Education and vocational training
4. Employment, workers management and access to self-employment
5. Access to an enjoyment of essential private services and public services and benefits
6. Law enforcement
7. Migration asylum and border control management
8. Administration of justice and democratic process

- **Risk management system**

Continuous development of a risk management system:

- Identification and analysis of risk
- Estimation and evaluation
- Estimation of possible arising risks
- Risk management measures

        - Elimination or reduction of risks as much as possible through adequate design and development
    - Testing for identifying the risks and the best management measures
    - Ocho ai bambini

- **Human oversight**

To prevent and minimize the health risks and the safety of fundamental rights, high-risk system requires human machine interface tools so that a human can effectively oversee during the period of use of the AI. The person needs to:

1. Understand capabilities and limits of the system and monitor its operations
2. Be aware of automation biases
3. Being able to override refuse and ignore the results of the system
4. Correctly interpret the output
5. Interrupt the system
6. For biometric identification needs at least 2 people for confirmation

- **Transparency obligations**

Article 52 certain AI system such as chatbots or deepfakes need to inform the user that that they are interacting with an AI system or that they are exposed to an emotion recognition system or a biometric categorization. That the video Is created by an AI

- **Obligations of providers of high-risk AI systems**

    - Compliant with requirements
    - Quality management system
    - Technical documentation
    - Keep logs automatically generated
    - Conformity assessment procedure
    - Registration obligation
    - Act if not conform the system
    - Inform the national competent authorities of the non-compliance and of any corrective action taken
    - CE marking
    - Upon request or national competent authorities demonstrate the conformity of the system

- **Obligations of users of high-risk AI systems**

User shall use the system in accordance with the instructions of use.

- Ensure input data is relevant
- Monitor the operation based on instruction of use
- Keep the logs

- **Conformity assessment and standards**

Certificate expires or substantial mods to the system need to apply for new conformity assessment

Art 43

- **Regulatory sandboxes**

Controlled environment for development, testing and validation for innovative AI systems as measures to support innovation

- **critical issues (according to Raposo and Edwards)**

EDWARDS

Biometrics

An area of controversy in the Act is whether it should include an effective total ban on use of facial recognition (or other types of biometric identification) and if so by what actors: law enforcement or private actors. At present under Article 5(1)(d) a ban exists but only in relation to:

- the targeted search for specific potential victims of crime, including missing children.
- the prevention of a specific, substantial, and imminent threat to the life or physical safety of natural persons or of a terrorist attack.
- the detection, localization, identification, or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA62 and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State.'

This is a very limited ban. Only '*real-time*' biometric identification systems are banned, i.e., those that identify individuals at a distance by comparing the biometrics of the observed subject with a biometric database without 'significant delay' (Article 3(37)).

Publicly accessible spaces '*does not cover places that are private in nature and normally not freely accessible for third parties, including law enforcement authorities, such as homes, private clubs, offices, warehouses and factories*. Online spaces are also not included.[12]

The restriction to law enforcement purposes excludes private security even though it may represent similar threats to fundamental rights. National security uses will also be excluded by virtue of Union law.[13]

Categorisation systems are only classed as ‘limited risk’ despite the heading of Annex III, containing the possibility that categorisation systems might in some future time be added to ‘high risk’. The distinction between identity and categorisation seems to have been drawn for consistency with the definition of biometric data in GDPR Article 4(14), which requires that biometric data be data processed with intent to identify a data subject. However, the question here is not what the *purpose* was for which personal data was processed, but whether the risk of *impact* on fundamental rights is high – hence this distinction seems irrelevant, when the real issue is that biometric categorisation has been shown to create severe impact on the rights of surveilled groups.

Similarly, the processing of facial or other data (temperature, sweat, eye movements, etc.) to establish emotional states (‘emotion recognition systems’) is also only in principle regulated by Title IV (‘limited risk’), though it may also fall into ‘high risk’ under Annex III, where used by law enforcement in limited circumstances (Annex II, 6(b)(‘polygraphs and similar tools’).

This means private-sector emotion recognition systems are in principle not classified as ‘high risk’ by the Act, although uses in employment or education *might* be caught by Annex III (3) and (4).

Raposo

Loopholes in the accountability dimension

The Draft Act is silent on two critical topics related to accountability. First, it does not establish a right to take legal action against suppliers or users of AI systems for non- compliance with its rules[64] (although non-compliance with these rules can be invoked in a civil liability proceeding). This limitation has been repeatedly pointed out both by European authorities (such as the EDPB and the EDPS)[65] and by civil rights groups.[66]

Additional accountability mechanisms should include civil liability for defaulters; however, specific compensation mechanisms are absent from this proposal. Producer liability is currently regulated in Directive No 85/374/EEC[67] on liability arising from defective products, which is currently in the process of being revised to adapt it to new challenges of the digital world.[68] It is expected that suitable solutions for the use of AI in various products will appear in the revisions. Other relevant documents are the 2019 Report on Liability for Artificial Intelligence and Other Emerging Digital Technologies[69] and the European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for AI (2020/ 2014(INL))[70]; however, these are mere recommendations/guidelines and not en- forceable regulations.

There is an additional problem not addressed in the Draft Act that is unrelated to stakeholder liability but instead concerns potential liability of the AI system itself. This has been increasingly discussed in the literature and is based on the premise that AI systems are legal persons, also a widely discussed topic.[71] The fact that the Draft Act remains silent on the issue of AI personhood can be taken as a stand in it- self (that AI systems are not legal persons nor can they be held accountable) or sim- ply as the postponement of a resolution of an extremely complex question.

Limitations on transparency

The Draft Act recognizes that in the field of AI, transparency obligations may face limitations arising from the protection of intellectual property rights. Therefore, in- formation obligations are 'limited only to the minimum necessary information for individuals to exercise their right to an effective remedy and to the necessary transparency towards supervision and enforcement authorities, in line with their man- dates' (Exploratorium Memorandum).[76]

In addition, under Article 52 of the Draft Act, transparency obligations do not apply to AI systems authorized by law to detect, prevent, investigate, or prosecute crimes, unless they are emotion recognition systems, in which case disclosure of their use is always mandatory (for instance, when police forces use facial recognition in undercover operations or when police or courts interrogate suspects).[77]

Moreover, it is difficult to comply with transparency obligations with respect to AI, intrinsic characteristics of which are opacity and extreme complexity (hence its characterization as a 'black box').[78] One might argue that most individuals affected by AI are unable to comprehend information about it. Thus, the requirement of transparency cannot be understood as requiring users of AI systems to have a precise understanding of how they work. Rather, this requirement should be limited to

providing an understanding of the main limitations of AI systems and identifying

their shortcomings.[79]

CONFLICTS BETWEEN NORMS

With several European norms regulating the same matter, it is only natural that conflicts will arise between rules. The Draft Act is part of a much broader project that aims to bring the EU into the digital world. For this purpose, several legal drafts are currently in the pipeline, namely the regulation on digital services,[84] the regulation on digital markets,[85] the regulation on data governance[86] and the Machinery Regulation, released together with the Draft Act.[87] In addition, a lengthy list of existing legislation is equally relevant to the field of AI, including the GDPR, the LED, the EUDPR,[88] the Directive on privacy and electronic communications,[89] the Regulation on medical devices,[90] the Regulation on in vitro diagnostic devices[91] and the Defective Products Directive (soon to be revised), just to name a few.

It will be almost impossible to avoid overlaps between all of these legal documents. Even if the norms do not conflict, a challenge remains in that it will be almost impossible to take them all into account when acting or making decisions with re- spect to AI. As a concrete example, consider that AI systems work with data, most of which are personal data; thus, it will be necessary to articulate the GDPR with the AI

regulation, which is no simple task, as both include very detailed rules, namely very detailed assessments (the DPIA under the GDPR and the conformity assessment under the AI Regulation).[92]

PRELIMINARY CONCERNS ABOUT THE DRAFT ACT

Despite being an important step for the EU in the technological domain, all is not 'rainbows and butterflies' in the AI Draft Act. First, in some regards, it is not enough. It is intended to be a comprehensive regulation on AI; however, it falls short of that goal. A detailed, comprehensive regulation would certainly be impossible to create, as AI is such an extensive and complex domain. However, there are glaring omissions in the act, among them its treatment of liability for damages.

Second, in other respects, it is too much, as certain matters are hyper-regulated by the Draft Act. European AI developers and manufacturers are asked to comply with an overwhelming number of requirements, some of them so demanding that they may be impossible to comply with. For example, Article 10/3 stipulates that 'training, validation and testing datasets must be relevant, representative, error-free and complete'. Experts point out that the idea of a completely error-free dataset is utopian.[97] Another difficulty of complying with this demand lies in its ambiguity.[98] Does the 'error' in question refer to the set of data, its classification, the way the intended behavior is represented, all of these aspects or other aspects? Furthermore, who will assess data quality and using what criteria?

Finally, the absence of a substantial boost to innovation in the Draft Act is a significant concern. The aim to bring about innovation in AI technologies is restricted to three articles, which, although introducing interesting provisions, fall short of what is required for a truly digital single market. Furthermore, there are only two measures in the Draft Act promoting innovation. AI investment in the EU may be hampered by such an 'innovation hole', which could advantage other leading players.[99] A definition of standards to be adopted by the rest of the world—an expression of the so-called Brussels effects[100]—may not be appropriate for the AI regulation, as most countries would prefer a model that is more balanced between fundamental rights and technological development. Ultimately, if the current proposal indeed becomes the new regulation, Europe may be inescapably relegated to the tail end of the digital revolution.