TRIAL Workshop – September, 3

AI and Ethics : An EU point of wiew

Yves Poullet

Prof Universities of Namur and Lille
NADI Co-chairman (University of Namur)
Chairman of The IFAP WG on Info-Erhics (Unesco)

UNIVERSITÉ DE NAMUR

Ethics and AI: everywhere…

OECD: 'Recommendation on artificial intelligence', Council of Ministers, May 22 2019

Council of Europe: Ad hoc Committee on Artificial Intelligence (CAHAI), Sept. 11, 2019

UNESCO: Ad Hoc Expert Group on Ethics and Artificial Intelligence (Nov. 2019)

EU – The « third Way »

- **EUROPEAN COMMISSION :** "Ethical Guidelines for trustworthy AI" (HLGE on AI Report – April 2019)

  + White Paper on AI – "Towards Excellence and **Trust**" (Feb. 2020).

- **EUROPEAN PARLIAMENT:** "PROPOSITION containing recommendations to the Commission on a framework of ethical aspects in artificial intelligence, robotics and related technologies (2020/2012 (INL))" **(**April, 21, 2020**)**

The Ethical question

- IF Digital might be considered as “the” building block of our societies (social, economic, political), our lives, our identities, how can we, as human and democratic societies, keep control of such a fast, complex, protean, ubiquitous innovation with unpredictable applications? : *“Human in the loop, human on the loop”*

The HLGE conclusions

- According to the Guidelines, trustworthy AI should be:
- (1) lawful - respecting all applicable laws and regulations
- (2) ethical - respecting ethical principles and values
- (3) robust - both from a technical perspective while taking into account its social environment

The HLGE conclusions (2)

The Guidelines put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy. A specific assessment list aims to help verify the application of each of the key requirements:

- **Human agency and oversight**: AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches
- **Technical Robustness and safety**: AI systems need to be resilient and secure. They need to be safe, ensuring a fall back plan in case something goes wrong, as well as being accurate, reliable and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.
- **Privacy and data governance**: besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, taking into account the quality and integrity of the data, and ensuring legitimised access to data.
- **Transparency:** the data, system and AI business models should be transparent. Traceability mechanisms can help achieving this. Moreover, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.

The HLGE conclusions (3)

- **Diversity, non-discrimination and fairness**: **Unfair bias** must be avoided, as it could could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.
- **Societal and environmental well-being**: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment, including other living beings, and their social and societal impact should be carefully considered.
- **Accountability:** Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. Moreover, adequate an accessible redress should be ensured.

Need for an enlargment - Nature of the risks

- Which risks?
  - Traditional risks – individual risks: Civil liability and D. P. rules moreless adequate
    - Physical injuries
    - Financial and economic losses
    - Moral individual damages
  - Collective risks to groups or to the society itself taking into account the societal AI impacts: How to face them?
    - Group's privacy and **discrimination**
    - **Environmental damages**
    - Distortion as regards market **Competition**
    - **Regulatory insidious changes** (see for instance: one-to one insurance; AI and preserving children welfare)
    - **Democracy**

The absolute need to enlarge the reflection – A global ethico legal approach -

'Parliament concludes, following previous discussions on the ethical aspects of artificial intelligence, robotics and related technologies (IoT)*, whereas this ethical dimension must be defined as a series of principles leading to a Union-wide legal framework, supervised by the competent national bodies, coordinated and reinforced by a European artificial intelligence agency and duly respected and certified within the internal market.'*

- The aim: "*Artificial Intelligence Based on Man (Human in the loop) and Developed by Man (Human on the loop)*". '*Any artificial intelligence, robotics and related technologies, including software, algorithms and data used or produced by such technologies, shall be developed, deployed and used in a* ***human-centric manner*** *with the aim of contributing to the existence of a democratic, pluralistic and equitable society by safeguarding human autonomy and decision-making and ensuring human agency.*' … "*The development, deployment and use of development, deployment and use of artificial intelligence, robotics and related technologies, including software, algorithms and data used or produced by such technologies, shall be carried out in the best interest of citizens. In particular, the potential of such technologies and the opportunities that they provide shall be taken into consideration having regard at all times to the need to protect and foster the social, environmental and economic well-being of society.' (art.5.3.)*

The Governance of AI systems: *DO GOOD DO NOT HARM: . Ethics by design - Risk assessment (multistakeholder and multidisciplinary participation)*

- ***Distinction between high risks AI systems and others…*** *Beyond GDPR : « high risk of breaching the principles of safety, transparency, accountability, non-bias or non-discrimination, social responsibility and gender balance, environmental friendliness and sustainability, privacy and governance, should be considered high-risk from a compliance with ethical principles perspective where that is the conclusion of an impartial, regulated and external risk assessment by the national supervisory authority.*
- ***Need to take into account all actors (developers, deployers and users)****: Beyond GDPR "Developers, deployers and users are responsible for compliance with safety, transparency, and accountability principles to the extent of their involvement with the artificial intelligence, robotics and related technologies concerned, including the software, algorithms and data used or produced by such technologies. Developers should ensure that the technologies concerned are designed and built in line with safety features, whereas deployers and users should deploy and use the concerned technologies in full observance of those features.' EXPLAINABILITY, SAFETY and ACCOUNTABILITY'*
- ***Need for actors to take into account general interest and their social responsibility*** : *'Any artificial intelligence, robotics and related technologies, including software, algorithms and data used or produced by such technologies, shall be developed, deployed and used in the Union in compliance with the relevant Union law, principles and values, in a manner that ensures optimal social, environmental and economic outcomes and that does not result in injury or harm of any kind to being caused to individuals or society … (and) in a socially responsible manner (art. 10.1 and 2)*

The Governance of AI systems

- ***Beyond Privacy Assessment, an a priori control by a multidisciplinary and multistakeholders independent body:***
  - At a national level: ' *National supervisory authorities should ensure the gathering of a maximum number of stakeholders such as industry, businesses, social partners, researchers, consumers and civil society organizations, and provide a pluralistic forum for reflection and exchange of views so as to achieve comprehensible and accurate conclusions for the purpose of guiding how governance is regulated.*' (see already, the Danish, German and UK Data Ethics Commisisons)
  - **At a EU level: «** EU Agency for Artificial Intelligence - '*to supervise the application of the proposed Regulation; to issue guidance as regards the application of the proposed Regulation; to liaise with the "Supervisory Authority" in each Member State and coordinate their mandate and tasks; to develop a European certificate of compliance with ethical principles and finally to support regular exchanges with concerned stakeholders and the civil society.*'

Which Ethical values? (UNESCO Convention on bioethics, 2005)

"Convinced that it is necessary and that it is time for the international community to set out universal principles on which humanity can respond to the growing dilemmas and controversies that science and technology pose to humanity and the environment,"

1. *DIGNITY : No manipulation: "Human as a target and never as a tool"- Robot never a human – Human never reduced to his or her data*
2. *AUTONOMY : Education- Freedom of choice – need for transparency - Right to have explanation of automated decision by a human being*
3. *SOCIAL JUSTICE : No discrimination (bias) - Collective approach beyond individual protection (groups' privacy) - Need for an extension of the concept of the Information Universal service within a digital society*
4. *DO GOOD DO NOT HARM: . Ethics by design - Risk assessment (multistakeholder and multidisciplinary participation)*

If you want to continue the discussion, please do not hesitate to contact me:

Yves.poullet@unamur.be