# AI - TRUSTWORTHINESS - INSIGHTS INTO KEY ETHICAL AND LEGAL ISSUES IN LIGHT OF THE NEW EU AI ACT

Author: Konstantinos Ntzoufas, Lawyer, LLM in Civil Law, Msc in Banking and Finance, Legal Ethics Advisor on AI

# AI ACT ANTE PORTAS –A NEW REGULATORY ENVIROMENT

- **The new EU AI Act (endorsed by European Parliament on 13.3.2024,approval pending by European Council ) is bound to come into force till the end of Mai 2024.**

- ✓ **Trustworthy and Human-centric AI in the limelight as overarching charachteristics of AI systems.**

- Definition of AI Systems: *'AI system' means a machine-based system 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;''*

- Key features:

a) Autonomy (partial or total**)**

**b) Ability to infer from the input** .

The latest definition differentiates from the initial as it puts more emphasis on the ability of the AI system to infer which is akin to more sophisticated technologies.

# Risk based approach of AI Systems

- The EU AI act adopts a risk-based regulatory framework providing in essence different regulatory ‘‘sub-regimes’’ according to their category of risk .

- Three categories of risk:

**a. Unacceptable-prohibited AI systems (manipulative, subliminal, social scoring etc. or even some biometric systems): Totally forbidden!**

**Shift from previous draft to a hard stance on real-time biometric AI systems in publicly accessible places, only exceptionally permitted for searching of human traffic and sexual exploitation victims and terrorist attacks .**

**b. High Risk AI Systems (defined in article 6 of AI Act): -AI Systems that pose a significant risk of harm to the health, security or fundamental rights of a person. *They must meet a series of strict standards/requirements set out in Chapter II of AI Act.***

***Two categories of High-Risk AI systems:***
***i)Those in ANNEX II(***covered by specific EU legislation, such as civil aviation, vehicle security, marine equipment, toys, lifts, pressure equipment and personal protective equipment.)

# Risk based approach of AI Systems

*ii) High-risk AI Systems listed in ANNEX III ( e.g. r*emote biometric identification systems, AI used as a safety component in critical infrastructure, in education, employment, credit scoring, law enforcement etc).

*The latter can avoid the classification as High-Risk AI under certain circumstances.*

- *All High Risk-AI Systems must be registered in the relevant EU database.*

*c.* ***Minimal Risk AI Systems****: Not explicitly addressed, they create a low or minimal risk to individuals with regards to the the impact they have om he health, safety and fundamental rights in the EU.*

- **The distinct case of General Purpose AI Models**: General purpose AI (**GPAI**) models are specifically regulated and classified under the AI Act. Their main feature is the generality and the capability to competently perform a wide range of distinct tasks. These models are typically trained on large amounts of data, through various methods, such as self supervised, unsupervised or reinforcement learning.

- ✓ **Providers/Developers of GPAI must comply with the obligations set in article 53 of AI Act.( technical documentation etc.)**

# GENERAL PRINCIPLES ON ALL AI SYSTEMS TRUSTWORTHY AND ETHICAL AI

The initial article 4a of AI Act had introduced an array of principles applicable to all AI systems regardless of their classification, **however it was not inserted in the final draft**. In any case **recital 27 of EU AI Act** refers to those 7 principles established as guidelines by the High Expert Group on AI (AI HLEG) urging to embrace them and use in the ambit of drafting Voluntary Codes of Conduct :

a)Human agency and oversight( human-centric): AI must serve people, respect human dignity and personal autonomy, and that is functioning in a way that can be appropriately controlled and overseen by humans.

b)Technical robustness and safety: AI allows robustness in the case of problems and resilience against attempts to alter the use or performance of the AI system.

c) Privacy and data governance: compliance with GDPR and all relevant legislation.

d) Transparency: Traceable and Explainable AI

e) Diversity, non discrimination and fairness: avoiding biases and promoting equality and diversity

f) Social and environmental well-being : developing AI in a sustainable and environmental friendly manner.

# TITLE III-CHAPTER I OF AI ACT--APPLICABLE TO HIGH RISK AI SYSTEMS

**RISK MANAGEMENT SYSTEM AS CORE REQUIREMENT FOR HIGH RISK AI SYSTEMS**

- In the case of RMS( Risk Management System) its role consists in ensuring that the residual risk from the use of AI is acceptable.
- Article 9 and 17 of AI Act provides for establishing, implementing, documenting and maintaining a risk management system( RMS) and quality management system(QMS).
- Main goal of RMS : identification and analysis of the known and the reasonably foreseeable risks that the high-risk AI system can pose to the health, safety or fundamental rights when the high-risk AI system is used in accordance with its intended purpose and adoption of appropriate and targeted risk management measures designed to address the risks.

In addition **and after the development and placing in the market of AI,** the evaluation of other possibly arising risks based on the analysis of data gathered from the post-market monitoring system referred to in Article 61 of the Proposal for the EU AI Act .

- Main goal of QMS- the purpose of the QMS is to ensure that the provider is well prepared to manage the risks that AI can pose.

# DATA GOVERNANCE AND PRIVACY-HIGH RISK AI SYSTEMS

- (Article 10)-Training, validation and testing data used in AI sets shall be subject to appropriate data governance and management practices appropriate for the intended purpose of the AI system.
- The right to privacy and to protection of personal data must be guaranteed throughout the entire lifecycle of the AI system.
- The principles of data minimisation and data protection by design and by default, as set out in Union data protection law, are applicable when personal data are processed(see recital 45a)

# FAIR AI AND COMPATTING BIASES AND DICRIMINATION IN HIGH AI SYSTEMS

- AI systems can perpetuate biases in data leading to unfair and uneven outcomes and discrimination.

  *The data sets should also have the appropriate statistical properties, including as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used, with specific attention to the mitigation of possible biases in the data sets, that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination prohibited under Union law, especially where data outputs influence inputs for future operations (feedback loops).*

- Counter measures embedded in article 10 of AI Act to tackle this problem.

- Special provision of article 10 **in order to detect negative bias**: *Providers of such systems may exceptionally process special categories of personal data referred to in Article 9(1) of Regulation (EU) 2016/679, Article 10 of Directive (EU) 2016/680 and Article 10(1) of Regulation (EU) 2018/1725, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons, including technical limitations on the re-use and use of state-of-the-art security and privacy-preserving.*

# HIGH RISK AI SYSTEMS
# Human agency and oversight

- AI systems shall be developed and used as a tool that serves people, respects human dignity and personal autonomy, and that is functioning in a way **that can be appropriately controlled and overseen by humans.**
- **Human oversight helps ensuring that an AI system does not undermine human autonomy or causes other adverse effects. Oversight may be achieved through governance mechanisms such as a human-in-the loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC) approach. HITL refers to the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable.**
- **Especially for High –Risk AI Systems this principle is embedded in article 14 :''***High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use''.*

# ACCURACY, TECHNICAL ROBUSTNESS AND SAFETY OF AI SYSTEMS

- Article 15 of AI Act: High-risk AI systems shall be designed and developed in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity, and that they perform consistently in those respects throughout their lifecycle.

a. **Technical robusteness**

AI systems shall be developed and used in a way to minimize unintended and unexpected harm as well as being robust in case of unintended problems .

- ✓ Resilience against attempts to alter the use or performance of the AI system so as to allow unlawful use by malicious third parties.

- ➢ *E.g. Provision for'' fail-safe plans'' in case of detected anomalies. Also interplay with the (European) Cyber Resilience ACT (CRA) which is also bound to be enacted.*

# ACCURACY, TECHNICAL ROBUSTNESS AND SAFETY OF AI SYSTEMS

**b. Accuracy**:

*High-risk AI systems that continue to learn after being placed on the market or put into service shall be developed in such a way as to eliminate or reduce as far as possible the risk of possibly biased outputs influencing input for future operations ('feedback loops'), and as to ensure that any such feedback loops are duly addressed with appropriate mitigation measures.*

# TRANSPARENCY AND HIGH RISK AI SYSTEMS

- High-risk AI systems should be designed in a manner to enable deployers to understand how the **AI system works**, evaluate its functionality, and comprehend its strengths and limitations.

- **Appropriate information in the form of instructions of use.** Such information should include the characteristics, capabilities and limitations of performance of the AI system.

- ✓ Also addressing the problem of the so called ‘’black box’’ algorithms’ problem .

- **Special transparency requirements for AI systems (irrespective of their classification) that interact directly with natural persons( article 50)** : Natural persons concerned must be informed that they are interacting with an AI system.

# ADDITIONAL NOTEWORTHY REQUIREMENTS - OBLIGATIONS AND NOVELTIES IN AI ACT

- Obligation for Technical Documentation for providers/developers of High risk AI Systems according to article 11:The technical documentation of a high-risk AI system shall be drawn up **before that system is placed on the market or put into service and shall be kept up-to date.**

**Technical Documentation** in a clear and comprehensive form to assess the compliance of the AI system with the requirements of AI Act. Coupled with the obligation for record-keeping(keeping log files throughout the lifecycle of the AI system).

- **AI Literacy (for all AI Systems) -Article 4:**Providers and deployers of AI systems shall take measures to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, and considering the persons or groups of persons on whom the AI systems are to be used.

- **Obligation for certain categories of deployers of AI Systems( banks, police, medical aid, firebrigade) to draw up a FRIA( Fundamental Rights Impact Assessment): B**odies governed by public law, or are private entities providing public services, and deployers High-risk AI systems referred to in points 5 (b) and (c) of Annex III, shall perform an assessment of the impact on fundamental rights that the use of such system may produce.

# A NEW LANDSCAPE-INTERRELATIONSHIP OF AI ACT WITH GUIDELINES ON AI OF HIGH GROUPS OF EXPERTS AND THE ROLE OF EUROPEAN STANDARDS.

- The Ethic Guidelines for Trustworthy AI(issued by the High Level Expert Group on Artificial Intelligence, 2019) although not having a legal binding nature -soft law- .The still serve as a basic roadmap to the realm of the AI best ethical practices and principles.

- Encouraging the adoption of Codes of Conduct for Voluntary Application of Specific Requirements( Article 69 of AI Act)

- The role of European harmonized Standards for Trustworthy AI Act in articles 40-42 of AI Act to be issued **as a presumption of AI Act compliance**(articles 40-42 of AI Act).

A blueprint for the European Standards is already requested by CEN and CENELEC.

*The European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC) are two distinct private international non-profit organizations.*

*See for instance with regard to Risk Management :ISO /IEC 23894 : 2023.*

*ECJ judgement on European Standards (''measures implementing or applying an act of EU Law''), James Elliott Construction Limited vs. Irish Asphalt Limited, 27 October 2016.*

# EXTRATERRIORIALITY AND INTERNATIONAL PRIVATE LAW ISSUES OF AI ACT

- Broad scope of application of the AI Act:

  - Applies to producers or deployers that place the AI system in the market in EU regardless of where these providers are established or located.
  - Affected persons that are located in EU .
  - Providers and deployers of AI systems that have their place of establishment or who are located in a third country, where the output produced by the system is used in the Union.

*Significant Exception: Military and Defence AI systems do not fall into the scope of AI Act .*

*Regarding the transmission of Personal data ECJ jurisprudence on Schrems I and II must be followed.*

- *Applicable law : Between the developers and the deployers Regulation of Rome I on the applicable law on contracts will be critical.*

# AI ACT AND THE ‘’GREEN CHALLENGE’’

- Emphasis put on the ‘’Green AI’’: See Recital 3 of AI Act: ’’**To contribute to *reaching the carbon neutrality targets*,** European companies should seek to utilise all available technological advancements that can assist in realising this goal. Artificial Intelligence is a technology that has the potential of being used to process the ever-growing amount of data created during industrial, environmental, health and other processes. To facilitate investments in AI-based analysis and optimisation tools, this Regulation should provide a predictable and proportionate environment for *low-risk industrial solutions.*’’

- **Effect on High-risk AI design: ’’High-risk AI systems shall be designed and developed with, the logging capabilities enabling the recording of energy consumption, the measurement or calculation of resource use and environmental impact of the high-risk AI system during all phases of the system’s lifecycle.’’( article 12).**

# AI LIABILITY PRODUCT DIRECTIVE AS A REGULATORY COUNTERPART FOR AI ACT

- The use of AI is intrinsically coupled with the liability for damages stemming from its use.
- **AI Liability Directive (AILD-Proposal COM 2022,496 final)**:
  - Addresses the needs of ensuring effective compensation for victims of damage caused by AI systems.
  - Regulates non –contractual liability(torts,etc.)
  - Aims at easing the burden of proof: Rebuttable proof of evidence in favour of the victim.If the latter proves non-compliance with AI Act and that a causal link with the AI performance is reasonably likely, **the court can presume that this non-compliance caused the damage.**

# CONCLUSIONS

- AI Act lays a new common ground in EU  but still many technicalities are left open.
- European Standardization Organizations are expected to have a pivotal role by  providing standards which will constitute ‘’safe harbours’’ of compliance.
- AI Act  shall be interacting with other legislation, such as EU Charter, GDPR, CRA(Cyber Resilience Act), Data Governance Act, MDR( Medical Devices Regulation), etc.

**THANK YOU FOR YOUR ATTENTION!**