**Module 3: Legal interpretations of fairness and transparency (1 lesson)**

**PRE-READING: Why regulating AI?**

* Elon Musk: regulate AI to combat 'existential threat' before it's too late, The Guardian: <https://www.theguardian.com/technology/2017/jul/17/elon-musk-regulation-ai-combat-existential-threat-tesla-spacex-ceo>
* Cambridge Analytica scandal highlights need for AI regulation, The Guardian, April 2018: <https://www.theguardian.com/technology/2018/apr/16/cambridge-analytica-scandal-highlights-need-for-ai-regulation>
* Towards Intelligent Regulation of Artificial Intelligence, Cambridge University Press:<https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/towards-intelligent-regulation-of-artificial-intelligence/AF1AD1940B70DB88D2B24202EE933F1B>

**READING:** **Regulating AI**

# China and Europe are leading the push to regulate AI – one of them could set the global playbook, CNBC, May 2022: <https://www.cnbc.com/2022/05/26/china-and-europe-are-leading-the-push-to-regulate-ai.html>

# UK National AI Strategy – AI Action Plan, Gov.Uk, Office for AI, July, 2022: <https://www.gov.uk/government/publications/national-ai-strategy-ai-action-plan/national-ai-strategy-ai-action-plan>

# A European approach to artificial intelligence, European Commission: <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>

* First regulatory sandbox on Artificial Intelligence presented, European Commission, June 2022: <https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented>
* How to Train an AI with GDPR Limitations: <https://intellias.com/how-to-train-an-ai-with-gdpr-limitations/>

**CONTENT:**

***Table of contents***

1. **Introduction**

1.1. Worldwide efforts to regulate AI

1. **Transparecny and Fairness in the AI proposed Regulation**

2.1. Introduction

2.2. Risk Levels

2.3. Transparency obligations

2.4 Fairness

1. **Transparency and Fairness: other sources**

3.1. EU

3.2. UK

1. **Transparency and Fairness in the GDPR**

4.1. Introduction

4.2. GDPR principles

4.3. Fairness and transparency obligations

1. **Introduction** 
   1. **Different efforts to regulate AI**

**Slide 2 EU flag**

In recent years, we have seen several countries are joining efforts in an attempt to regulate AI. In the EU, in April 2021, the European Commission published a proposal for an AI Regulation that would directly apply to the 27 Member States. In September 2022, another proposal was published for an AI Liability Directive. In addition, several efforts have been made at policy level, such as the *Ethics Guidelines for Trustworthy”* AI issued by the High-Level Expert *Group* of European Commission in 2019.

Slide 3 US flag

In the US, there are different proposals pending at House and Senate that would regulate AI and its impacts on society. For instance, the “Justice Against Malicious Algorithms Act” would lift the immunity of online platforms which knowingly or recklessly use algorithms or other technology which produces harmful effects. The **“**Algorithmic Justice and Online Platform Transparency Act” that would prohibit algorithmic processes that discriminate on the basis of race, age or gender, and require online platforms to describe to users in plain language the types of algorithmic processes they employ and the information they collect.

Slide 4 China flag

In China, the “Provisions on the Administration of Algorithm Recommendation for Internet Information Services” came into force in March 2022, whereby, amongst others, service providers shall not set up any algorithm model to induce users to indulge themselves, excessively consume, or otherwise violate laws and regulations or ethics and shall establish and improve mechanisms to identify illegal and undesirable information.

1. **Transparency and fairness obligations set out in the proposed AI Regulation** 
   1. **Introduction to the AI proposed Regulation**

The proposed text of the Regulation is not final yet and it is being discussed within the Council, the Parliament and its preparatory bodies. It is likely that the current text will undergo changes during the legislative procedure. A date for its final approval and entry into force is not known yet. However, companies and public entities are looking at the provisions in the proposal as a benchmark for compliance.

The scope of the AI Act is meant to be extraterritorial, that means, that not only EU providers will be bound by the AI Regulation but also companies which are not EU based but intend to operate in the EU or use its outputs within the EU.

* 1. **Risk-based approach**

In order to elaborate on the transparency and fairness obligations set out in the AI proposed Regulation, we shall commence by explaining that the proposal contains a risk-based approach. The proposal defines 4 levels of risk in AI systems:

1. Unacceptable risk: AI posing unaceptable risks will be banned or allowed only under certain specific circumstanced and subject to rigorous obligations.
2. High-risk: AI systems are permitted on the European market subject to compliance with certain mandatory requirements and an ex-ante conformity assessment prior to its placement on the market or putting into service. The classification of an AI system as high-risk is based on the intended purpose of the AI system.
3. Limited risk: refers to AI systems with specific transparency obligations.

1. Minimal or no risk: free use of minimal-risk AI.

Depending on the purpose and its function the AI system will be classified on a different level. The proposal provides a list of some AI systems and its classification:

* Unacceptable risk: AI posing unacceptable risks will be banned or allowed under certain exceptions. (Title II of the proposal.)
* Subliminal techniques beyond a person’s consciousness in order to distort a person’s behavior in a manner that causes or is likely to cause that person physical or psychological harm.
* AI systems for the evaluation or classification of the trustworthiness of natural persons based on their social behaviour.
* The use of real-time remote biometric identification systems in public spaces for law enforcement purposes. Except for certain crimes and to prevent acts of terrorism and always subject to prior judicial authorization.
* High risk: List provided in Annex III: AI systems meant to be used in the following areas:
* Biometric identification and categorisation of natural persons
* AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water gas, heating and electricity
* Education and vocational training
* Employment, workers management and access to self-employment
* Access to and enjoyment of essential private services and public services and benefits (including public assistance benefits and creditworthiness)
* Law Enforcement
* Migration, asylum and border control management
* Administration of justice and democratic processes
* Limited risk: Limited risk refers to AI systems with specific transparency obligations. When using AI systems such as chatbots, users should be aware that they are interacting with a machine so they can take an informed decision to continue or step back.
* Minimal or no risk systems: The proposal allows the free use of minimal-risk AI. This includes applications such as AI-enabled video games, spam filters, customer and market segmentation systems, etc.
  1. **Transparency obligations**
* The proposed Regulation sets certain transparrency obligations for high-risk AI systems. (Article 13 of the proposed Regulation). No black box AI where the user cannot interpret the systems’ outputs shall be put in the market.

High-risk AI systems:

Shall be designed and developed ensuring transparency to enable the interpretation of the systems’ output and to use it appropriately.

* Shall include instructions for use.
* Shall include information on:
  + - the provider (name, contact details)
    - Its intended purpose
    - the level of accuracy, robustness and cybersecurity
    - any circumstance, including misuse that can lead to risks
    - Specifications in terms of training, validation and testing
* Human oversight measures
* the expected lifetime of the system including any necessary maintenance
* Transparency obligations for certain AI systems: Article 52 of the proposed Regulation poses certain transparency obligations for certain AI systems, in this regard:

Certain AI systems:

* Natural persons shall be informed that they are interacting with an AI system (e.g. **chatbots**) (except when used for law enforcement).
* Natural persons shall be informed of the use of any **emotion recognition system** or a **biometric categorization system** (except when used for law enforcement).
* **Deep fakes** shall disclose that the image, audio, or video has been created or manipulated artificially (except when used for law enforcement or necessary for the right to freedom of expression and the right to freedom of the arts).
* Transparency, IPRs and confidentiality: The EC emphasises that the increased transparency obligations will also not disproportionately affect the right to protection of intellectual since they will be 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 mandates. The transparency obligations towards end-users of the AI and natural persons will not require the disclose of trade secrets, know-how or information related to patents.
  1. **Fairness**

The proposal does not include the notion of fairness as such. However, the proposal complements existing Union law on non-discrimination with specific requirements that aim to minimise the risk of algorithmic discrimination, and therefore to render AI systems fair. In particular, the design and the quality of data sets used for the development of AI systems complemented with obligations for testing, risk management, documentation and human oversight throughout the AI systems’ lifecycle aims to design, develop and implement AI systems which are fair and do not discriminate or produce harm on natural persons. In this line, Tittle III of the proposal set out the following obligations for high-risk AI systems:

1. Risk management system: identification, monitoring and management of risks for natural persons (including discrimination) along with mitigation measures. Risks should always be acceptable and those risks that are not acceptable should not be run.
2. Data and data governance: training, validation and testing of AI models should be done with **high quality data sets**:

* fit for purpose
* prior assessment of the suitability, identification of gaps etc.
* ensuring bias monitoring
* personal data processing 🡪 in accordance with the GDPR or any applicable law 🡪 your DPO or a privacy expert should be involved at the early stages of the design process.

1. Technical documentation: to be drafted in such a way that demonstrate compliance with the obligations set out for high-risk AI systems and shall be ready before the system is placed on the market.
2. Record keeping: Automatic record of events (‘logs’), start and end date and time, which data has been checked, input data, natural persons involved, etc.
3. Transparency (as explained in the previous subsection)
4. Human oversight:

* Systems should be *overseen* by natural persons who need to be able to *understand* the capacities and limitations, to interpret the system’s outputs and to be able to decide not to use, *disregard, override or reverse* the systems’ *outputs*.
* For *biometric identification* systems, the outputs shall be priorly *verified* by at least *2 natural persons*.

1. **Transparency and fairness: other sources**
   1. **EU soft law ethical AI**

The EC has appointed 52 experts to a High-Level Expert Group on Artificial Intelligence (AI HLEG), comprising representatives from academia, civil society, as well as industry to make recommendations on future related policy development and ethical, legal and societal issues related to AI, including socio-economic challenges. In April 2019, the AI HLEG presented the Ethics Guidelines for Trustworthy AI, identifying 7 principles which have been used as a benchmark to design, implement and deploy ethical AI systems. These 7 principles have been somehow embedded into the AI proposed Regulation. These principles are applicable to the entire AI systems’ life cycle:

* 1. Human oversight and agency
  2. Technical robustness and safety
  3. Privacy and data governance
  4. Transparency
  5. Diversity, non-discrimination and fairness
  6. Societal and environmental well-being
  7. Accountability
  8. **UK Algorithmic Transparency Standard**

In 2021, UK Government published the “*Algorithmic Transparency Standard*”. The standard is to be piloted by public sector bodies whereby the UK government aims at setting an example on ethical deployment of AI.

The aim of the Algorithmic Transparency Standard is to (1) ensure citizens have trust and confidence on how the public sector uses AI and (2) strengthen UK’s data ethics leadership and AI governance.

The standard and template are designed to maintain information about how the algorithm operates and why it is used in a standardized way.

Completed standards are published online.

1. **Transparency and fairness in the GDPR:** 
   1. **Introduction**

The General Data Protection Regulation (GDPR) is the EU law that governs personal data processing. At the beginning of the design process, you should check whether the AI system will be processing personal data and specially whether the AI system uses personal data to produce outputs. In the affirmative, privacy laws will need to be checked and complied with. In the EU, the GDPR and national privacy laws will apply and, in the UK, the UKGDPR. However, it should be noted that the GDPR (same as the proposed AI regulation) has an extraterritorial scope.

Here are a couple of examples on the extraterritoriality of the GDPR:

1. When a company is established in a third-country but operates within the EU, will be bound by the GDPR.
2. When company is established in the EU and process personal data of its customers outside the EU, for instance, in Morocco: the GDPR is also applicable.

* When processing personal data, you should *always check with* your *privacy team* which privacy law applies and you should involve them at the early stages of the project.
  1. **The GDPR principles**

The GDPR follows a principle-based approach. Meaning that, data controllers are responsible (accountable) for demonstrating compliance, at all times and on every stage of the personal data processing process, with the following principles:

* 1. Lawfulness, fairness and transparency
  2. Purpose limitation (explicit and legitimate purpose)
  3. Data minimisation (only data processing limited to what is necessary)
  4. Accuracy (data kept accurate and up to date)
  5. Storage limitation (data stored for no longer than necessary)
  6. Integrity and confidentiality (appropriate security measures)
  7. Accountability (the controller is always responsible for GDPR compliance)

* ***Fairness*** means that the data processing should always be fair, and it shall not be processed in a way that is detrimental, unexpected or misleading to the individuals concerned.
* ***Transparency obligations*:** individuals should always be informed of when, how and for what purposed their personal data is being processed. Information to be provided in privacy notices includes (Article 13 of the GDPR), namely:
* Name and contact details of the entity processing the data
* Purpose of the personal data processing
* Applicable legal basis
* Recipients of the data
* International data transfers
* Retention period
* Data protection rights
* The existence of automated decion-making, including profilin
* Etc.

In addition, individuals whose date is being processed have, amongst others, the right to access to the personal data that is being processed.

🡪 ***Right of access*** to: which data is being processed, its purpose, , the recipients, the retention period, and whether they are being subject to automated-decision making or profiling, etc.