

NLP and the Law

Dr Lachlan Urquhart

12^h Feb 2025.

Things to Consider

- Further discussion on this in ethics workshop later in year (May 7th TBD).
- Where does this intersect with NLP?

EU AI Act 2024

- AIA defines obligations based on risk.
- Defines design obligations for systems and responsibilities for providers, deployers, importers, distributors, and third parties, especially for High-Risk AI (HRAIS).
- Across lifecycle and supply chain.
- Operationalising - turn to AI designers?

Scope of AI Systems Covered

- AIA defines AI systems “‘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**;” Art 3(1).
- Creates **tiers of risk** depending on AI application.




Prohibited AI (in the wild)

- Systems that pose *unacceptable risks* to health, safety, fundamental rights.
- Law enforcement live remote automated facial recognition (except – for serious crimes; finding victims; terror attacks)
- Subliminal manipulation/deception to distort behaviour of group
- Exploitation of any vulnerabilities based on age, disability, economic situation
- Social credit scoring
- Emotion recognition in education and workplaces.

High Risk AI Systems (HRAIS)

- Linked to safety applications (e.g., integrated into riskier products in Annex I – cars, medical device, toys etc.) or if is a standalone safety system.
- Also, if is on a *list* of high-risk contexts (Annex III) i.e.
 - AI in critical infrastructure like energy;
 - targeted job screening;
 - emotion recognition that is not work/education based;
 - profiling;
 - crime analytics;
 - border control;
 - polygraphs;
 - to research /interpret facts + law in courts.



Reqs for HRAIS - Data governance (Art 10)

- Ensure ***acquisition, training, validation and testing of data*** sets meet strict standards
- e.g., representative of target environment/population; ***correction of gaps/errors***; measure ***accuracy***; mitigate discrimination + bias ***harms***.

Technical documentation (Art 11) and Record Keeping (Art 12).

Technical documentation covering what is in Annex IV e.g. intended purpose, how it interacts with software/hardware other systems; APIs, description of the UI, instructions for use.

And **automated recording of logs** (lifespan of system) – recording situations where risk emerged, substantial modifications, enable post market monitoring.

Risk Management System (Art 9)

Eliminate risks to human health, safety or fundamental rights + put mitigation and control measures in.

Inform deployers of risks.

Automated data logging in use and analysis across system ***life cycle***.

Designing for human oversight (Art 14)

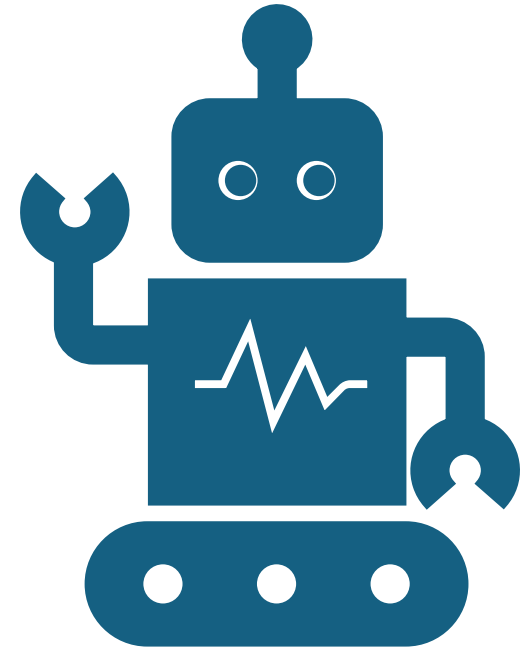
- Create Human-Machine Interface tools.
- Allow natural persons to exercise oversight and control during use.
- This includes pausing, reversing, overriding or stopping the system during anomalies, dysfunctions and unexpected actions.

Implications for Design

- HCI focused on designing and evaluating user experiences and interactions with technologies.
- **Human Oversight.** Ensure ‘system operation is transparent to users to understand output and use it’ (Art 13).
- Links to **Explainable AI?** Beyond this – how to design to ***support human action with actual*** AI use?

Providers of General Purpose AI

- What is Gen Purpose AI in the AIA?
- “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”. See also Recital 65.





GPAI (2)

- Art 53 – obligations for providers of GPAI differ if pose systemic risk or not (i.e. negative impacts on public health, safety, security, fundamental rights).
 - Tech documentation incl training/testing process/evaluation/energy consumption /floating point operations.
 - Keep documentation
 - Policy to comply with copyright/related rights
 - Summary of content used for training
 - Does not apply to models released under free / open source licence.

Draft EU Code GPAI

- Draft Code of Practice from European Commission (finalized April 2025)
- Commitments of Signatories - Copyright
 - Create and implement internal copyright policy (and publish a summary).
 - Assess compliance of third-party datasets
 - Lawful access to copyright protected content (as per TDM exemption)
 - Don't crawl websites with copyright infringing content
 - Respect robot exclusion protocol (robots.txt – IETF)
 - Respect rights reservations (that are expressed computationally)
 - Publish information on rights reservation compliance
 - Prohibit copyright infringing uses of the model.

When GPAI posing 'systemic risks'

Model evaluation

Size/quality of dataset/energy consumption

Assess and mitigate risks at scale

Notify serious incidents/corrective measures to AI office.

Adequate cybersecurity e.g. data poisoning

GPAI Purely for Research

- Recital 25
- AIA should ‘support innovation, should respect freedom of science, and should not undermine research and development activity’.
- Exclude
 - ‘AI systems and models specifically developed and put into service for the sole purpose of scientific research and development.’
 - Also during commercial R+D prior to products being put on market
 - Yet any R+D needs to be done “in accordance with recognised ethical and professional standards for scientific research and should be conducted in accordance with applicable Union law.”
 - Rules do apply when in real world testing and reg sandboxes, and once products put on market.

Open Source

- Recital 102
- Art 1(12) - “This Regulation does not apply to AI systems released under free and open-source licences, unless they are placed on the market or put into service as **high-risk AI systems** or as an AI system that falls under **Article 5 or 50.**” i.e. prohibited AI system or certain AI systems covered in Art 50 e.g. deep fakes.
- “General-purpose AI models released under free and open-source licences should be considered to ensure **high levels of transparency and openness if their parameters, including the weights, the information on the model architecture, and the information on model usage are made publicly available.** The licence should be considered to be free and open-source also when it allows **users to run, copy, distribute, study, change and improve software and data, including models under the condition that the original provider of the model is credited, the identical or comparable terms of distribution are respected.**”
 - so seems to be GPL/copyleft?

NLP and Copyright

- Breach of Copyright
- Use of copyrighted material - Licensing– permission of rightsholders? Attribution to authors when scraped at scale? Compensation. Court cases in US/UK –eg Getty v Stable AI.
- Training with royalty free/ public domain/creative commons licensed (or contractual agreements with publishers) – economic rights
- Moral Rights of Authors to control how works are used... (often waived)
- Copyright Exemptions for text and data mining – (s29A CDPA) for research, non-commercial, where there is lawful access to the work (i.e. so not just scraping – need some permission).
- Liability – if model trained on compromised data – how to indemnify further use? Contractually?
- UK Intellectual Property Office Report / DCMS / DSIT consultation.



NLP and Data Protection

- **Web Scraping** – lack of awareness by data subjects – ICO argues for greater transparency from developers around how data is being used to enable subjects to use their rights (e.g. right to erasure/portability/restriction/subject access rights etc.)
- Legality of data scraping for training data – what is the lawful basis? Unlikely to be consent... other grounds. Legitimate interests of controller?
- What is legitimate interest of controller being pursued – is data processing necessary – is legit interest being balanced against fundamental rights/freedoms of data subjects.
- Concerns around anonymity around a model trained on personal data – need to be interrogated to see how that can be the case – taking into account state of art methods used by developers of model.



Data Protection Principles Art 5 GDPR

- **Lawfulness**, fairness (unexpected uses?) and transparency
- Purpose limitation – multiple purposes of processing
- Data minimisation – often opposite with scale of scraping
- Accuracy
- Storage limitation
- Integrity and confidentiality (security)
- Accountability – demonstrate steps taken to data subjects/regulators

Thoughts/ Questions?

- Lachlan.Urquhart@ed.ac.uk