Artificial Intelligence: mitigating bias

Understanding the Legislative & Regulatory Contexts

Dr Elizabeth Lomas e.lomas@ucl.ac.uk

Denise Almeida denise.almeida.18@ucl.ac.uk

Overview

- Accessibility
- Limitations
- Learning outcomes
- Who we are
- What we will cover

Who we are

Dr Elizabeth Lomas
Associate Professor
Information Governance, UCL
e.lomas@ucl.ac.uk

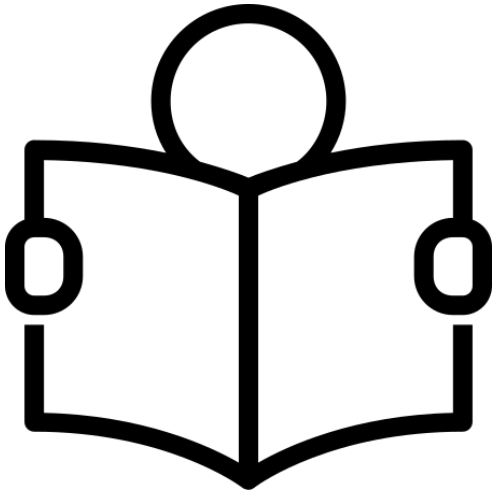
Denise Almeida
Post Graduate
UCL
denise.almeida.18@ucl.ac.uk

What we will cover

- Part 1: Legislation and Regulatory Contexts
- Part 2: Personal Data Concepts: the GDPR and AI
- Part 3: GDPR in practice: RoPA
- Part 4: Impact Assessments
- Part 5 Case-Study: AI-driven Recruitment
- Part 6: Future gazing: ePrivacy and AI regulation
- **Activities and Readings**

Readings and Activities Overview

- Readings and activities have been provided for the respective parts of each learning unit.
- These readings should form the point of departure for further reading.



Part 1

Legislation and Regulatory Contexts

Ethical and Compliant AI Innovation

- Developed
- Deployed
- User/end user
- Location
- Data/Data banks/lakes accrual
- Purpose
- Proportionality
- Models
- Accountability
- Explainability
- Challenge
- Legal Responsibilities

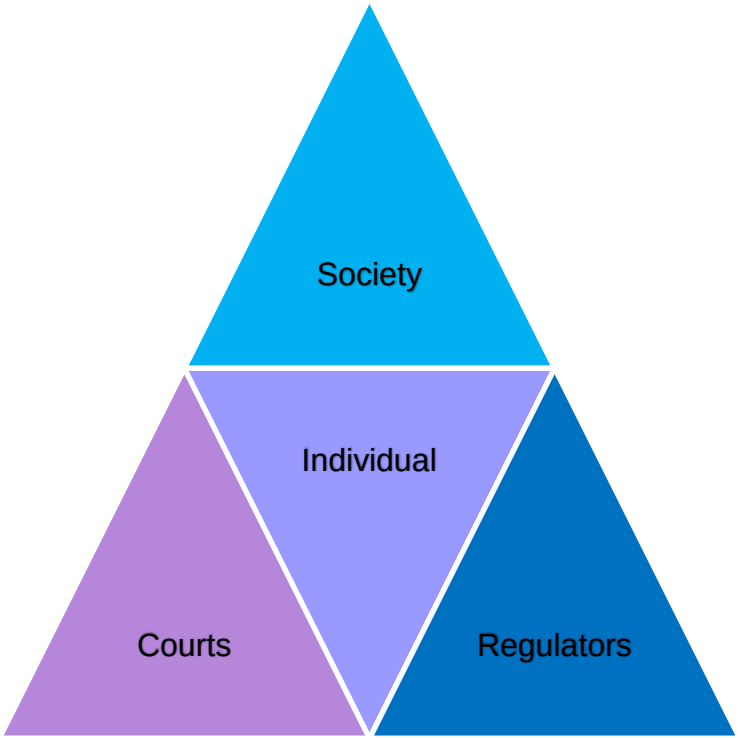
Ethical Innovation for Artificial Intelligence
https://ei4ai.wordpress.com/

Global Legislative Contexts

Multiple jurisdictions and perspectives add to the complexity of compliance:

- United Nations
- European Union
- UK
- Canada
- USA
- China
- India
- Local/State

Legislation and Societal Context



Intersection of:

- Legislation
- Case law
- Regulators
- Best practice

Public versus Private Law

➢ Differing evidential standards

Warfare

Risk Management

Legislation

- Prescriptive- thou shalt
- Balancing competing considerations

Avoiding negative impacts

- Impact of Breaches
- Financial
- Harm
- Damage

Providing for rights and freedoms

Human Rights as a Foundation

Universal Declaration of Human Rights - 1948

- Article 7 – all equal before the law
- Article 12 – right to privacy
- Article 17 – right to property
- Article 18 – right to freedom of thought
- Article 19 – right to freedom of expression
- Article 27 - right to the protection of the moral and material interests resulting from any scientific, literary or artistic production

Equality, Diversity and Inclusion

EU Charter of Fundamental Rights (Article 21)
UK Equality Act 2010
USA Civil Rights Act

➢Protected Characteristics

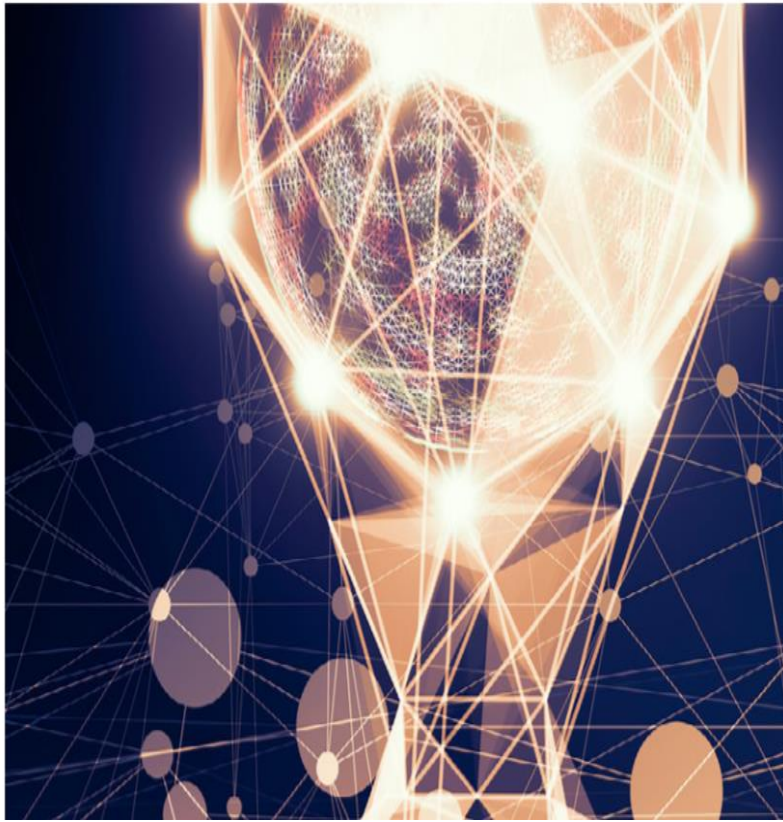
Age, Disability, Gender Reassignment, Marriage and Civil Partnership, Pregnancy and maternity, Race, Religion or Belief, Sex, Sexual orientation, Genetics, Language.

➢ Direct, Indirect, Associative, Perceptive Discrimination

Intellectual Property Rights

-Patents
-Industrial Design
-Trade secrets and ‘know how’
-Trade Marks
-Brand
-Copyright – needs human creator
-Digital Rights Management
-Software, Database Rights etcs

- Legislated for in different ways
- Last for different times
- Differ internationally



Intellectual Property Rights

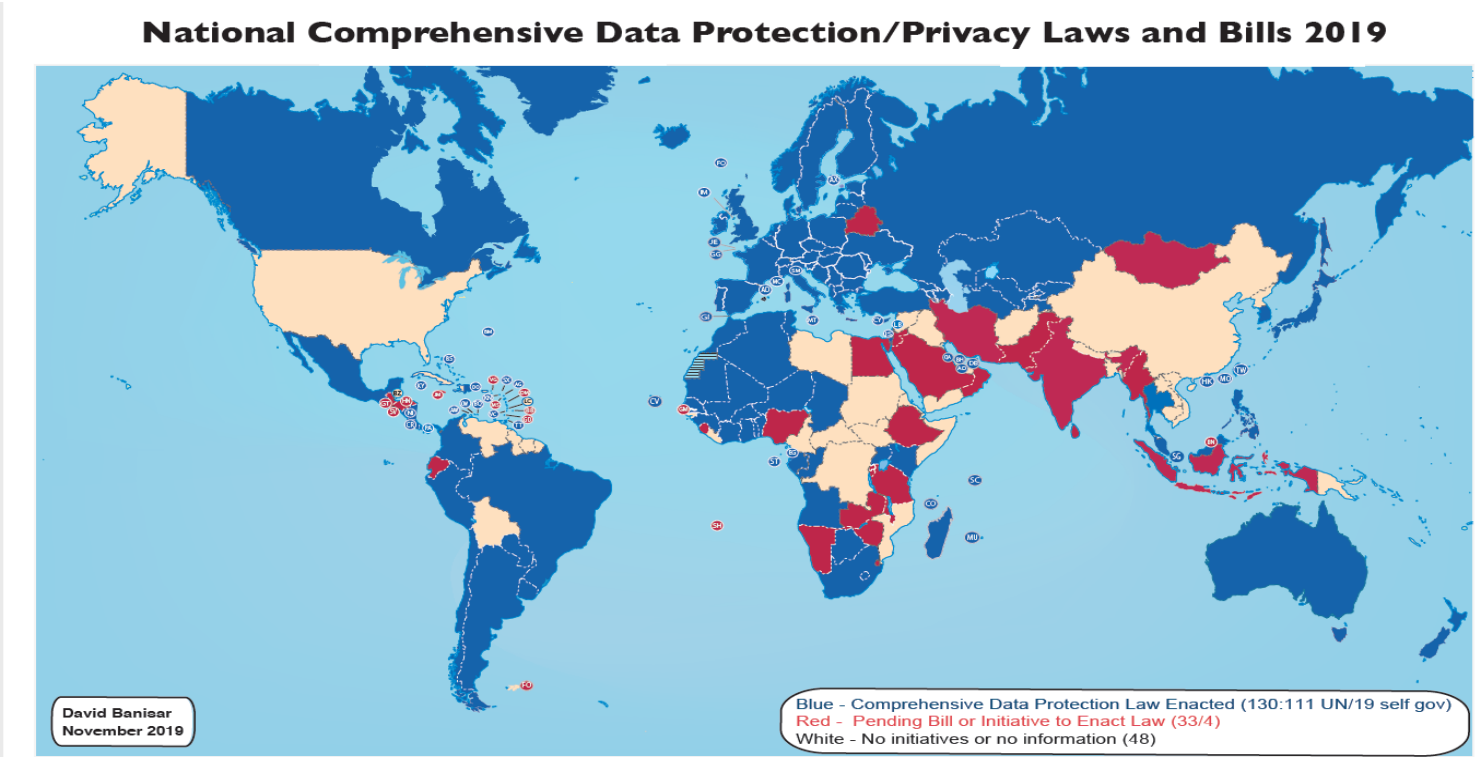
Provides for:
- Control
- Collaboration Frameworks
- Commoditization
- Monopolies
- Protection

- Software is protected by copyright law – provides paternity and moral rights
- Fair Dealing vs Fair Use
- Inventions related to software are protected by patents
- Trade secrets – contracts/NDAs

World Intellectual Property Organisation

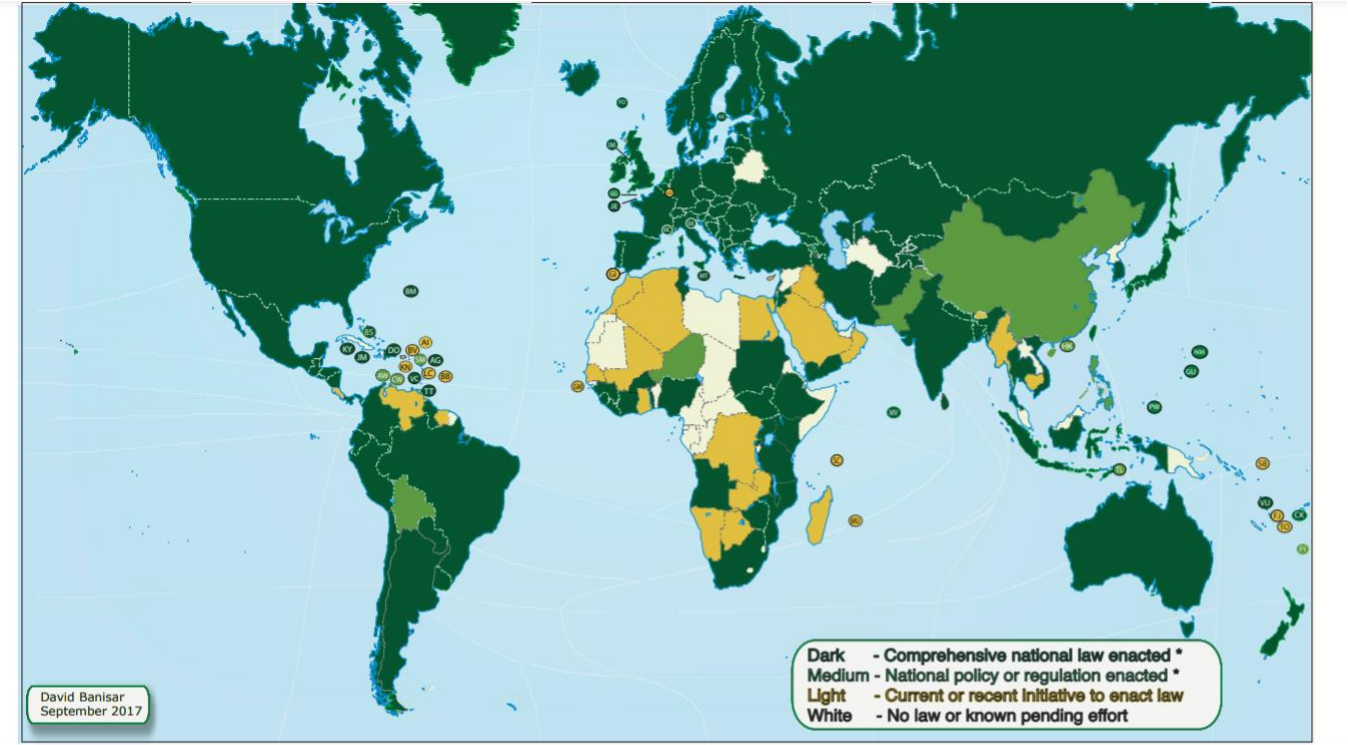
"Our mission is to lead the **development of a balanced** and **effective** international IP system that enables **innovation** and **creativity** for the **benefit of all**. Our mandate, governing bodies and procedures are set out in the WIPO Convention, which established WIPO in 1967."
https://www.wipo.int/about-wipo/en/

Privacy and Personal Data: Mapping of information privacy regimes



This image is produced with permission of David Banisar. See https://www.researchgate.net/publication/337060085_National_Comprehensive_Data_ProtectionPrivacy_Laws_and_Bills_-_November_2019

Mapping of information access regimes



This image is produced with permission of David Banisar. See https://www.researchgate.net/publication/256010224_National_Right_to_Information_Laws_Regulations_and_Initiatives_2017_Map

Activity: Mapping out the legislative and regulatory landscape

To map out the legislative context either for your organization or an organization you are familiar with. Consider:

- What is the broad legislative context?
- Who are the regulators?
- How/where/why might AI be developed/deployed?
- Who are the stakeholders?
- What could be the benefits of AI in this context?
- What might be the risks taking into account any issues around potential harm or damage or rights infringements?

Part 1 - Readings

Read **Chapter 3** on ‘Who bears responsibility for the threats, risks, harms and wrongs posed by advanced digital technologies?’, in Yeung, K. ed. (2019) *Responsibility and AI*: a study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Strasbourg: Council of Europe. Available at: https://rm.coe.int/responsability-and-ai-en/168097d9c5 (Accessed 1 July 2021)

Part 2

Personal Data Concepts: The GDPR and AI

GDPR Principles

General Data Protection Regulation: special category data

- racial or ethnic origin;
- political opinions;
- religious or philosophical beliefs;
- trade union membership;
- data concerning health or sex life and sexual orientation;
- genetic data and biometric data when processed to uniquely identify a person

All processing needs a legal basis

processing personal data...

- Consent
- Legitimate interests
- Vital interests
- Performance of a contract
- Compliance with a legal obligation
- Public interest

processing special category data...

- Explicit consent
- Legal claims or courts
- Vital interests
- Preventative or occupational medicine
- Employment law obligation
- Made public by data subject
- Archiving, research, statistics
- Substantial public interest

GDPR: children

- UK children under the age of 13 can never, themselves, give consent to the processing of their personal data in relation to online services.
- For UK children between the ages of 13 and 15 (inclusive), the general rule is that if an organisation seeks consent to process their personal data, then parental consent must be obtained, unless the relevant individual Member State legislates to reduce the age threshold – although the threshold can never drop below 13 years of age.
- Children aged 16 or older may give consent for the processing of their personal data themselves.

GDPR responsibilities

- Data Controller(s) –
- Data sharing agreement

- Data Processor -
- Data processing agreement

- Human intervention and decision making

General Data Protection Regulation: getting it wrong

Regulatory powers

Public reprimands and warnings

Order into compliance

Order to rectify, restrict or erase

Order to temporarily stop processing

Order a stop to international data sharing

ICO Fines

up to 20 Million Euro (or 4% Turnover)

Getting the principles wrong

Infringing the rights of an individual

Sharing personal data overseas without safeguards

up to 10 Million Euro (or 2% Turnover)

Not reporting breaches to the ICO

No privacy by design

No records of processing

General Data Protection Regulation: Privacy by Design

Privacy by Design systems to minimize DP risks, such as data minimisation, anonymisation, pseudonymisation

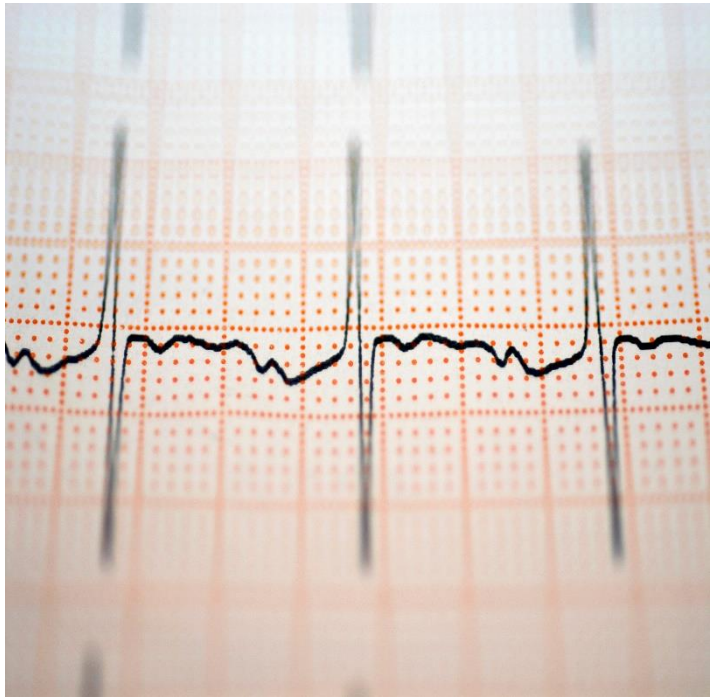
DP documentation of processes – Records of Processing

Data Protection Impact Assessments

Algorithmic Impact Assessments

Privacy Notices

General Data Protection Regulation: Privacy by Design



Data Protection Impact Assessments (Article 35)

Human Intervention and Automated Decision Making (Article 13, 15, 22)

Algorithmic Impact Assessments

See Article 29 Working Party/European Data Protection Board Guidelines

Wider Context

Automated Decision Systems Accountability Act – California

Fair Credit Reporting Act – Federal Law

Part 2 – Readings

Read the **ICO**'s 2020 discussions on **addressing risks of bias and discrimination in AI design** which are at https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/what-do-we-need-to-do-to-ensure-lawfulness-fairness-and-transparency-in-ai-systems/#howshouldweaddress . These sections are within its *Guidance on Data Protection and AI*. The full publication is available at: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/ (Accessed 1 July 2021). Whilst this is a guide from the UK's Information Commissioner, it provides considerations of how to build an AI framework which is compatible with data protection legislation and as a whole will provide you with many critical points for developing and deploying an AI system.

Part 2 – Activities

In addition, building on your Part 1 mapping of the regulatory and legislative landscape, you may wish to look at one of the following links, considering global differences and synergies in data protection law. The DLA Piper link provides an easy tool for comparisons.

Data Protection and Privacy Legislation Worldwide
https://unctad.org/page/data-protection-and-privacy-legislation-worldwide

DLA Piper's Data Protection Laws of the World
https://www.dlapiperdataprotection.com/

EU GDPR Law Comparison
https://www.activemind.legal/law/

Part 3

GDPR in Practice:
Records of Processing Activities (ROPA)

Records of Processing Activity (ROPA)

- Defined in art. 30 of the GDPR,
- Records detailing all of the processing activities of controllers or processors:
  - Details of controller and/or processor
  - Purpose of the processing
  - Description of the categories of data subjects and data recipients
  - Details on transfers to third-countries or international organisations
  - Description of technical and organisational security measures

What can the records be used for?

- Provide evidence of accountability;
- Assist in breach investigations and/or responding to individual rights requests;
- Highlighting areas of risk which require Data Protection Impact Assessments;
- Inform policy and procedure development;
- Facilitate the maintenance of retention schedules.

ROPA and AI

- Aid **transparency** by keeping detailed records of processes, procedures and data used for AI applications
- Highlight **high risk areas** which require further impact assessments and, potentially, mitigation strategies and changes to the product

Part 3 – Activity and Readings

Using the provided ICO Records of Processing Template, create records for an AI/ML driven process within your organization.

Read through the ICO guidelines on Records of Processing Activities and the ICO and The Alan Turing Institute's guidance on explainable AI, particularly **task 2 of part 2,** for further support in this activity.

Part 4

Impact Assessments

Impact Assessments

Impact assessment (IA) is a structured a process for considering the implications, for people and their environment, of proposed actions while there is still an opportunity to modify (or even, if appropriate, abandon) the proposals. It is applied at all levels of decision-making, from policies to specific projects.

Data Protection Impact Assessments (DPIA) are a legal requirement of the GDPR for all potentially high-risk activities, including “**new technologies**” and “**innovative processes**”

AI would fall under both requirements –
a DPIA is encouraged for all AI projects which involve personal data

Records Management Implications

ACLU About

NEWS & COMMENTARY

The Untold Number of People Implicated in Crimes They Didn't Commit Because of Face Recognition

The sheer scope of police face recognition use in this country means that people have almost certainly been — and will continue to be — misidentified, if not arrested and charged for crimes they didn't commit.

Source: Garvie, C. (2020) The Untold Number of People Implicated in Crimes They Didn't Commit Because of Face Recognition, American Civil Liberties Union. Available at: https://www.aclu.org/news/privacy-technology/the-untold-number-of-people-implicated-in-crimes-they-didnt-commit-because-of-face-recognition/

- Issues of inaccuracy, particularly exacerbated when the data relates to people belonging to minority groups (see **Fairness and AI module**);
- Lack of accountability and explainability: records not kept or accessible following a match, giving no recourse to individuals;
- Keeping track of lawful basis of processing;
- For archives management, specifically, the use of algorithms for the appraisal of archives also needs to be addressed from an ethical standpoint, as human biases can be built into the algorithms in question.

Part 4 - Readings

On issues of global governance:

Digital Platforms Require a Global Governance Framework (2019) *Centre for International Governance Innovation*. Available at: https://www.cigionline.org/articles/digital-platforms-require-global-governance-framework

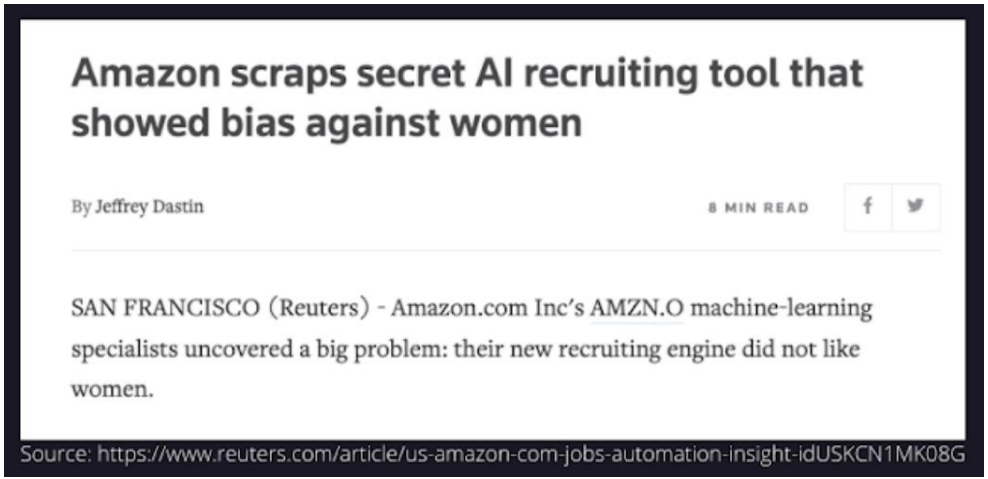
On impact assessments as a framework for auditing:

Kazim, E., Denny, D. M. T. and Koshiyama, A. (2021) 'AI auditing and impact assessment: according to the UK information commissioner's office', *AI and Ethics*. doi: 10.1007/s43681-021-00039-2.

Part 5

Case-Study: AI-driven Recruitment

Recruitment



Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Source: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

What went wrong?

- Special category data collected without enough consideration or legal basis for processing;
- No privacy assessment conducted to consider the privacy implications of the use of the algorithm;
- No considerations for algorithmic bias and bias associated with the input data.

Part 5 - Readings

Reuters article on the Amazon AI recruitment tool: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

On job seekers attitudes towards AI driven recruitment: Friedrich-Alexander-University, Schöller Endowed Chair for Information Systems, Erlangen-Nuremberg, Germany, Ochmann, J. and Laumer, S. (2020) 'AI Recruitment: Explaining job seekers' acceptance of automation in human resource management', in heine, M., Poustcchi, K., and Krasnova, H., *WI2020 Zentrale Tracks*. GITO Verlag, pp. 1633–1648. doi: 10.30844/wi_2020_q1-ochmann.

Part 6

Future gazing:
ePrivacy and AI regulation

Incoming regulations

The EU regulatory proposal aims to provide AI developers, deployers and users with clear requirements and obligations regarding specific uses of AI.

The EU's strategy on AI is focused on 3 goals:

- place Europe ahead of technological developments and encourage the uptake of AI by the public and private sectors
- prepare for socio-economic changes brought about by AI
- ensure an appropriate ethical and legal framework

The long-awaited ePrivacy regulation should also provide a helpful regulatory context to enforce on online privacy issues associated with AI, such as real-time bidding.

A new legislative paradigm?

Responsibilities?

1. How have you controlled the development, testing, rollout and then implementation in context of AI systems ensuring the management and challenge of bias?
2. What specific consents, notices and checks and balances are in place for fairness and transparency for these purposes?
3. On what basis have the data banks been built and used in relation to which purposes?
4. What specific consents, notices and checks and balances are in place for fairness and transparency for data bank accrual and use and what should not be allowable in terms of data scraping etc?

A new legislative paradigm?

5. What are the limitations of AI performance capabilities for different purposes taking into consideration the design context?

6. What accountability is in place for different use?

8. How can this accountability be explicitly exercised, explained and audited for a range of stakeholder needs?

9. How are complaint and challenge processes enabled and afforded to all?

10. Can counter-AI initiatives be conducted to challenge and test law enforcement and audit systems?

Lomas & Almeida 2021

Thank you!

Part 6 - Readings

European Union's Coordinated Plan on Artificial Intelligence 2021 (https://digital-strategy.ec.europa.eu/en/policies/plan-ai) alongside point 1 (Explanatory Context) of the proposed Artificial Intelligence Act: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

EU proposal for an ePrivacy Regulation: https://digital-strategy.ec.europa.eu/en/policies/eprivacy-regulation

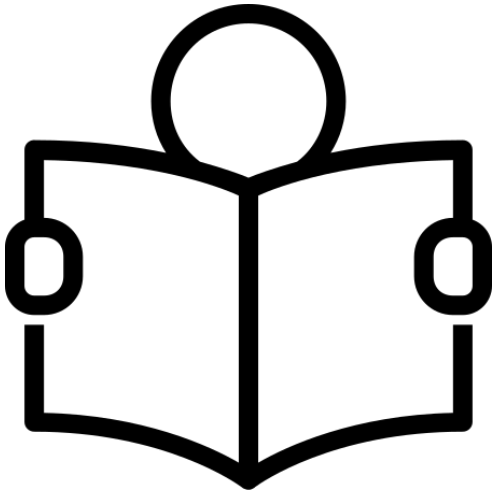
Veale, M. and Borgesius, F.Z., 2021. Adtech and Real-Time Bidding under European Data Protection Law. Available at: https://osf.io/preprints/socarxiv/wg8fq/

All Activities and Readings

with additional readings and links

Readings and Activities Overview

- Readings and activities have been provided for the respective parts of each learning unit.
- These readings should form the point of departure for further reading.



Part 1-Activity: Mapping out the legislative and regulatory landscape

To map out the legislative context either for your organization or an organization you are familiar with. Consider:

- What is the broad legislative context?
- Who are the regulators?
- How/where/why might AI be developed/deployed?
- Who are the stakeholders?
- What could be the benefits of AI in this context?
- What responsibilities exist?
- What might be the risks taking into account any issues around potential harm or damage or rights infringements?

Part 1 - Readings

Read **Chapter 3** on ‘Who bears responsibility for the threats, risks, harms and wrongs posed by advanced digital technologies?’, in Yeung, K. ed. (2019) *Responsibility and AI*: a study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Strasbourg: Council of Europe. Available at: https://rm.coe.int/responsability-and-ai-en/168097d9c5 (Accessed 1 July 2021)

Part 2 – Readings

Read the **ICO**'s 2020 discussions on **addressing risks of bias and discrimination in AI design** which are at https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/what-do-we-need-to-do-to-ensure-lawfulness-fairness-and-transparency-in-ai-systems/#howshouldweaddress . These sections are within its *Guidance on Data Protection and AI*. The full publication is available at: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/ (Accessed 1 July 2021). Whilst this is a guide from the UK's Information Commissioner, it provides considerations of how to build an AI framework which is compatible with data protection legislation and as a whole will provide you with many critical points for developing and deploying an AI system.

Part 2 – Activities

In addition, building on your Part 1 mapping of the regulatory and legislative landscape, you may wish to look at one of the following links, considering global differences and synergies in data protection law. The DLA Piper link provides an easy tool for comparisons.

Data Protection and Privacy Legislation Worldwide
https://unctad.org/page/data-protection-and-privacy-legislation-worldwide

DLA Piper's Data Protection Laws of the World
https://www.dlapiperdataprotection.com/

EU GDPR Law Comparison
https://www.activemind.legal/law/

Part 3 – Activity and Readings

Using the provided ICO Records of Processing Template, create records for an AI/ML driven process within your organization.

Read through the ICO guidelines on Records of Processing Activities and the ICO and The Alan Turing Institute's guidance on explainable AI, particularly **task 2 of part 2,** for further support in this activity.

Part 4 - Readings

On issues of global governance:

Digital Platforms Require a Global Governance Framework (2019) *Centre for International Governance Innovation*. Available at: https://www.cigionline.org/articles/digital-platforms-require-global-governance-framework

On impact assessments as a framework for auditing:

Kazim, E., Denny, D. M. T. and Koshiyama, A. (2021) 'AI auditing and impact assessment: according to the UK information commissioner's office', *AI and Ethics*. doi: 10.1007/s43681-021-00039-2.

Part 5 - Readings

Reuters article on the Amazon AI recruitment tool: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

On job seekers attitudes towards AI driven recruitment: Friedrich-Alexander-University, Schöller Endowed Chair for Information Systems, Erlangen-Nuremberg, Germany, Ochmann, J. and Laumer, S. (2020) 'AI Recruitment: Explaining job seekers' acceptance of automation in human resource management', in heine, M., Poustcchi, K., and Krasnova, H., *WI2020 Zentrale Tracks*. GITO Verlag, pp. 1633–1648. doi: 10.30844/wi_2020_q1-ochmann.

Part 6 - Readings

European Union's Coordinated Plan on Artificial Intelligence 2021 (https://digital-strategy.ec.europa.eu/en/policies/plan-ai) alongside point 1 (Explanatory Context) of the proposed Artificial Intelligence Act: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

EU proposal for an ePrivacy Regulation: https://digital-strategy.ec.europa.eu/en/policies/eprivacy-regulation

Veale, M. and Borgesius, F.Z., 2021. Adtech and Real-Time Bidding under European Data Protection Law. Available at: https://osf.io/preprints/socarxiv/wg8fq/

A Few Further Readings and Links

- European Data Protection Board Guidance https://edpb.europa.eu/our-work-tools/general-guidance/guidelines-recommendations-best-practices_en
- General Data Protection Regulation https://gdpr-info.eu/
- Human Rights Watch – Privacy News, e.g. https://www.hrw.org/news/2018/04/19/data-privacy-human-right and https://www.hrw.org/report/2021/06/16/my-life-not-your-porn/digital-sex-crimes-south-korea
- UNESCO https://en.unesco.org/artificial-intelligence
- United Nations – AI News, e.g. https://www.un.org/en/chronicle/article/towards-ethics-artificial-intelligence

A Few Further Readings and Links

- Almeida, D., Shmarko, K. and Lomas, E. (2021) ‘The ethics of facial recognition technologies, surveillance and accountability in an age of Artificial Intelligence: a comparative analysis of USA, EU and UK regulatory frameworks. Forthcoming, *AI Auditing, Assurance, and Certification. AI and Ethics*. Springer. ISSN: 2730-5961, Available at SSRN: https://ssrn.com/abstract=3878867 or http://dx.doi.org/10.2139/ssrn.3878867

- Kaminski, M., and Malgieri, G. (2020) ‘Algorithmic impact assessments under the GDPR: producing multi-layered explanations’, *International Data Privacy Law*, 2020;, ipaa020, https://doi.org/10.1093/idpl/ipaa020.

- Koshiyama, A. et al. (2021) ‘Towards algorithm auditing: a survey on managing legal, ethical and technological risks of AI, ML and associated algorithms’. Available at SSRN: https://ssrn.com/abstract=3778998 or http://dx.doi.org/10.2139/ssrn.3778998