Regulating explainable AI in the European Union
An overview of current legal framework(s)

XXXV Nordic Conference on Law & IT
Stockholm, 11 November 2020

Prof. Dr. Martin Ebers
University of Tartu (Estonia)
martin.ebers@ut.ee
www.ai-laws.org
2\textsuperscript{nd} RAILS-Conference
5\textsuperscript{th} April 2019, Berlin, Germany

“Governance of AI and Robotics: A European Approach?”
TECH & LAW CAMP

Samstag, 27. April 2019

Taylor Wessing RAILS LAWLIFT
We. Create.
Geriatronics.
About the lighthouse initiative.
Regulating explainable AI in the EU

- Introduction: The Black Box Problem
- EU Secondary Law
  - GDPR
  - Other Directives/Regulations
- EU Fundamental Rights and Constitutional Principles
- Conflicting Interests Militating against Explainability
- Future Regulation in the EU
Introducing: The Black Box Problem

Original Research Article

How the machine ‘thinks’: Understanding opacity in machine learning algorithms

Jenna Burrell

Nigerian 419 Scam:
My Dearest, Greetings to you
my Dear Beloved, I am Mrs Alice Walton, a citizen of United State. I bring to you a proposal worth $1,000,000,000.00 which I intend to

our (0.500810)
click (0.464474)
remov (0.417698)
guarante (0.384834)
visit (0.369730)
basenumb (0.345389)
dollar (0.323674)
price (0.268065)
will (0.264766)
most (0.261475)
pleas (0.259571)
Nigeria (- 0.001861)
Introduction: The Black Box Problem

Ethical Challenges covered across AI Principle Documents
Source: PwC based on 59 Ethical AI Principle documents.

- Fairness
- Interpretability and Explainability
- Transparency
- Accountability
- Data Privacy
- Reliability, Robustness and Security
- Human Control
- Safety
- Diversity and Inclusion
- Lawfulness & Compliance
- Multi Stakeholder Engagement
- Sustainability

Source: Stanford AI Index Report 2019, p. 149
Introduction: The Black Box Problem


  - The opacity of many AI systems “(...) may make it hard to verify compliance with, and may hamper the effective enforcement of, rules of existing EU law meant to protect fundamental rights. Enforcement authorities and affected persons might lack the means to verify how a given decision made with the involvement of AI was taken and, therefore, whether the relevant rules were respected. Individuals and legal entities may face difficulties with effective access to justice in situations where such decisions may negatively affect them.”

  - “(...) there is a need to examine whether current legislation is able to address the risks of AI and can be effectively enforced, whether adaptations of the legislation are needed, or whether new legislation is needed.”
Introduction: The Black Box Problem

• Different Types of Explanations
  • External and internal explanation
  • Model Transparency
    • How easy can we understand a model?
    • Entire Model: Simulatability
    • Individual Components: Decomposability
    • Learning Algorithm: Algorithmic transparency
  • Post-hoc Interpretability
    • How easy can we explain a decision?
    • Visualisation of learning outcome
    • Local Explanations by analyzing the parameters for a single decision
    • Explanations by Examples
Introduction: The Black Box Problem

Performance vs. Explainability

Source: Gunning, (2017) Explainable Artificial Intelligence (XAI), DARPA.
• **Prohibition of Fully Automated Decisions – Art. 22(1) GDPR**
  • The data subject shall have the right not to be subject to a decision *based solely* on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

• **Permitted (Fully) ADM – Art. 22(2) GDPR**
  • a) if necessary for *entering or performing a contract*
  • b) if *authorized* by Union or Member State law
  • c) if based on the data subject's *explicit consent*.

• **Rights of Data Subjects if ADM is permitted – Art. 22(3) GDPR**
  • Right to *obtain human intervention*
  • Right to *express their point of view*
  • Right to *contest the decision* that has significantly affected them.

• **Recital (71)**: safeguards should, inter alia, include “the right [...] to *obtain an explanation of the decision* reached after such assessment”.

**EU Secondary Law: The GDPR**
EU Secondary Law: The GDPR

• Ex ante Notification Duties of Data Controllers:

  ▪ **Art. 15(1)(h) GDPR**: data controllers are obligated to inform data subjects about “the **existence** of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, **meaningful information about the logic involved**, as well as the **significance** and the **envisaged consequences** of such processing for the data subject.

    ▪ Interpretation (1): It only refers to the general structure/architecture of the processing model. No need to provide explanations about individual decisions or concrete weights/features of the model

    ▪ Interpretation (2): Information can only be meaningful if it helps data subject to exercise her rights under Art. 22. Therefore: Specific explanations of the decision.
CJEU, joined cases C-141/12 and C-372/12 (YS. and M. and S.):

- Third country national who applied unsuccessfully for a residence permit has – under Art. 12 Data Protection Dir. 95/46 – no right to access the legal analysis (minutes) explaining the reason for that administrative decision.
- The legal analysis „is not in itself liable to be the subject of a check of its accuracy by that applicant and a rectification under Article 12(b) of Directive 95/46.”

Purpose of EU Data Protection:
- Fairness of Data Processing = Procedural Fairness
- EU Data Protection is NOT designed to assess the accuracy of decision-making process!
EU Secondary Law: Other Directives/Regulations

- Art. 6(1)(ea) Consumer Rights Directive 2011/83, as amended by Dir. 2019/2161
  - Information Duty of the Trader for distance and off-premises contracts that the price was personalised on the basis of automated decision-making

- Art. 5(2) P2B Regulation 2019/1150
  - Providers of online search engines shall set out the main parameters, which are most significant in determining ranking and the relative importance of those main parameters
  - by providing an easily and publicly available description
EU Fundamental Rights and Constitutional Principles

- **Fundamental Rights are Relevant if AI systems are used**
  - by **public administrations**, for example:
    - Public Security (smart borders, predictive policing)
    - Supervision of Financial Market
    - Tax Authorities
    - Social Welfare System
  - In the **judicial sector**, for example
    - For Evidence
    - To support a decision (ex.: COMPAS system)

- **Guarding Principles**
  - Rule of Law – Legal Certainty for individuals (Art. 2 TEU)
  - Due Process – Fair Trial – Effective Legal Protection (Art. 47 EU Charta of Fundamental Rights, Art. 6 ECHR)
Opaque algorithms which are **used for evidence** impair:

- adversarial proceedings
  - The parties cannot contradict the opponent’s allegations
- equality of arms principle
  - knowledge asymmetry between parties
- confrontation right
  - algorithm cannot be cross-examined

Opaque algorithms which are **used to support a decision** impair:

- presumption of innocence
- right to have a reasoned judicial decision
  - Algos do not justify how a particular decision has been reached
- right to appeal
  - Algos do not provide a motivation for their outputs
EU Fundamental Rights and Constitutional Principles

USA: State v Loomis 881 NW2d 749 (Wis 2016) 754 (US)

• Facts of the Case
  • Mr Loomis was imprisoned based (also) on the COMPAS risk-assessment
  • Loomis lodged a motion for post-conviction relief, requesting a new sentencing proceeding
  • Defence argued that the court’s reference to COMPAS violated his constitutional right to due process.

• Judgment of the Wisconsin Supreme Court
  • Appeal was rejected
  • Court acknowledged that risk scores failed to explain how COMPAS employed data to generate the results
  • However, according to the court, the risk assessments were “largely based” on available information such as criminal history and the answers to a list of questions provided.
Conflicting Interests Militating against Explainability

- **Trade/Business Secrets and IP Rights**
  - Recognized by Recital (63) GDPR as limits for transparency
  - Trade Secrets Dir. 2016/943
  - Computer Program Dir. 91/250
  - Database Dir. 96/9
  - European Patent Convention

- **Effectiveness of the System**
- **National Security**
- **Privacy Rights**
Conflicting Interests Militating against Explainability

• Evaluation
  • No general IP protection for algorithms!
  • Impact on Business depends on Type of Explanation

➤ Balancing Competing Interests
Conflicting Interests Militating against Explainability

• Tools for Balancing Competing Interests
  • Multi-tiered access regimes for sensitive information, which distinguish between
    • notification duties,
    • access rights to raw data
    • access right to aggregated data
  • Temporal restrictions for access rights
  • Use of information intermediaries,
  • Procedural safeguards, such as in-camera-proceedings, etc
Future Regulation in the EU

  - Overall Aim
    - EU-wide approach to avoid fragmentation in the internal market
    - Adjustment to existing EU Frameworks
    - New Legislation on AI, based on a risk-based approach
      - Mandatory legal requirements for “high risk” AI applications
      - Voluntary Labelling for “no-high risk” AI applications
  - High-risk AI systems, if
    - AI is employed in a high-risk sector, e.g. healthcare, transport, justice
    - AI used in such a manner that significant risks are likely to arise
Future Regulation in the EU

• European Commission, White Paper on AI, COM(2020), 65
  • Possible Requirements for High-risk AI systems regarding
    • Quality of training data
    • Keeping of Records and Data
    • adequate information and ensuring transparency regarding the use, limitations and capabilities of high-risk AI systems;
    • robustness and accuracy
    • Appropriate Human involvement
    • Specific requirements for certain particular AI applications, such as those used for purposes of remote biometric identification.
Future Regulation in the EU

Proposals of the German Data Ethics Commission
Future Regulation in the EU

• Civil Liability for AI
  • European Commission, Report on Safety/Liability, COM(2020), 64
    • Obligation of developers to disclose the design parameters and metadata of datasets in case of accidents?
    • Strict Liability of operators for risky AI applications?
    • For non-risky AI applications: Adaption of burden of proof concerning causation and fault?
  • European Parliament Report by Axel Voss, Oct. 2020
    • Operators of High-risk AI systems should be subject to strict liability
    • Low-risk AI system should be subject to fault-based liability
Future Regulation in the EU

- **EU Member States**
  - 14 EU Countries call for "soft-law solutions" in future AI regulation

- **Other Regulatory Options**
  - Ban of Certain Systems?
  - Premarket approval system
  - Monitoring and Oversight by Regulatory Bodies
  - Co-Regulation and Standardization
    - Certification of AI systems
    - Quality Seals
    - Presumption of Conformity if products comply with standards
Future Regulation in the EU

• A Plea for „Differential Explainability“
  • XAI is not a static, but a dynamic concept

• Requirements for Explainability depend on many factors
  • Domain where AI is used
  • Impact/Risk for Individual and Society
  • Target Audience: Who is the addressee of explanations?
  • Goals of XAI: Why is explanation necessary?
  • Institutional setting: How can the involved institutions (e.g. agencies or courts) mitigate the black box problem?