SUPPORT POOL
OF EXPERTS PROGRAMME

AI Auditing

Proposal for AI leaflets

by Dr. Gemma GALDON CLAVELL

As part of the SPE programme, the EDPB may commission contractors to provide reports and tools on specific topics.

The views expressed in the deliverables are those of their authors and they do not necessarily reflect the official position of the EDPB. The EDPB does not guarantee the accuracy of the information included in the deliverables. Neither the EDPB nor any person acting on the EDPB's behalf may be held responsible for any use that may be made of the information contained in the deliverables.

Some excerpts may be redacted or removed from the deliverables as their publication would undermine the protection of legitimate interests, including, inter alia, the privacy and integrity of an individual regarding the protection of personal data in accordance with Regulation (EU) 2018/1725 and/or the commercial interests of a natural or legal person.

Table of Contents

Document initially submitted in January 2023, updated in June 2024

Background

Since 2016, GDPR has laid out principles and procedures that have shaped how issues related to data protection and social impact are addressed in data-intensive technologies. The notions of transparency (articles 13 and 14 GDPR), "human intervention" (article 22.3 GDPR), information about the logic of the processing (article 14.2.g GDPR), accountability (article 5.2 GDPR), data protection by design and by default (article 25 GDPR) and auditability (including the notion of conformity assessment in the AI act) have shaped a shared, global understanding of what data protection means in practice.

To be accountable means, among others, a complete traceability of all the design decisions, taken by design, properly documented, analysed in advance, and backed with proof and evidence. But while the accountability principles have been laid out, it is still unclear how these principles can be implemented and checked in practice in ways that cover all relevant moments in the supply chain and facilitate enforcement by the supervisory authorities.

This is particularly relevant at a time when we see the accountability chain getting increasingly complex in AI, with companies often buying AI (foundational) models and services from third parties and retraining them with additional data or using them on their own decision-making processes.

We have developed AI leaflets as a key tool of effective AI transparency for AI users and implementors, but also as a mechanism to protect SMEs and provide a level-playing field for all industry actors. AI implementors and those using AI, both end-users and AI "clients", currently lack standardized tools to exercise free, Informed choice. In the absence of these tools, entities are forced to rely on marketing claims and unverified information which may create risks for their users and expose organizations to "inherited liability".

In this second report for the European Data Protection Board (EDPB), we develop a proposal for "AI leaflets", a concept exported from the medical domain to enforce a priori transparency for AI systems and products, and which draws on previous work developed for the Spanish DPA and the Spanish Ministry of Labor. AI leaflets complement existing tools like Model Cards, impact assessments, AI audits and algo-scores. Due to their technical nature, Ai leaflets are close to Model Cards. As the information in an AI leaflet is intended for a tech-savvy audience, Ai implementors should implement the algo-scores we proposed in our first report to facilitate and-user understanding and choice.

1. Basic definitions

Objectives of the algorithmic leaflet: to provide accessible information that promotes transparency, auditability and recourse to those buying, implementing or being impacted by AI systems. The leaflet facilitates compliance with requirements included in GDPR and AI Act.

Definition of algorithmic system: software that is developed with one or more techniques and Machine Learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; and statistical approaches, Bayesian estimation, search and optimization methods that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with (adapted from the definition of artificial intelligence contained in AI Act art. 3.1).

Examples of affected systems: those where one or more algorithms are at the center of a decision-making process that has implications for fundamental rights or individual/collective life-chances, including social media content recommenders, price/retribution models in consumer services, hiring

decisions, individual/group risk assessment in different settings (facial recognition as proof of life/identity, benefit allocation, recidivism, etc) and Large and Small Language and Image Models (Generative AI) used to interact with complex or unstructured data which produce new content users rely on to understand an issue or make decisions.

Inherited liability: When one entity buys AI products from another and uses them in their own decision-making processes or product design, it can be held legally responsible for any issues that lead to harmful, inefficient or discriminatory decisions or assessments. Leaflets provide key information to AI clients and users so they can make better decisions when choosing an AI system or provider.

2. Why algorithmic leaflets

In the last few years, at least 170 sets of ethical or human-rights based AI principles, frameworks, and guidelines have been developed to support responsible AI development and deployment in the public and private sectors.[1] Research has shown that a growing consensus has emerged around core principles, such as the need for accountability, privacy and security, transparency and explainability, fairness and non-discrimination, professional responsibility, human control, and the promotion of human values.[2] These principles and values have made it into discussions around how to regulate AI-related technologies, and both existing EU regulations such as the General Data Protection Regulation (GDPR) and the Digital Services Act (DSA) and new regulatory proposals being discussed right now, such as the AI ACT, echo this emergent consensus.

But while significant steps have been taken to align high-level approaches and principles, an important lesson from the GDPR, passed in 2016, is that enforcement can be a challenge. As AI principles gain acceptance within the public and private sectors, the focus is shifting to the development of appropriate strategies to operationalize them into responsible *practices*. Yet, as Nonnecke and Dawson highlight, "this process is not straightforward".[3]

One way to accelerate the adoption of enforcement practices is by drawing on the long history of how modern societies have dealt with the negative externalities of innovation, how complex scientific insight has been communicated to users and citizens in recent history, and the tools that have emerged to protect people and rights in highly innovative processes.

Looking at the history of the regulation of innovation, a relevant precedent and example for the effective regulation of AI systems and products is the medical sector. In the late 18th Century and early 19th century, many companies developing drugs and medicine would market their products under false, untested premises. In 1902, one advertisement for a medical product claimed, "No other preparation has had its therapeutic value more thoroughly defined or better established . . . [as] a remedy in the treatment of coughs, bronchitis . . . asthma, laryngitis, pneumonia, and whooping cough." The drug was heroin.[4]

---

[1] AI Ethics Guidelines Global Inventory," Algorithm Watch, https://inventory.algorithmwatch.org/
[2] Jessica Fjeld et al. (2020), Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI, Berkman Klein Center for Internet & Society, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482
[3] Nonnecke, B. and Dawson, B. (2021) Human Rights Implications of Algorithmic Impact Assessments: Priority Considerations to Guide Effective Development and Use. Carr Center Discussion Paper Series, https://www.hks.harvard.edu/sites/default/files/2023-11/2021_25_nonnecke_and_dawson_human_rights_implications.pdf
[4] Hamburg, M.A. (2010), Innovation, Regulation, and the FDA. N Engl J Med; 363:2228-2232. https://www.nejm.org/doi/full/10.1056/NEJMsa1007467

And while the 20th Century saw enormous and hugely beneficial advances in medicine, in its early decades many companies marketed their products with a variety of unproven claims. It was, as pharmacologist Louis Goodman called it', a "therapeutic jungle", not much different from a tech and AI industry that many have described as the "Wild West". It took several public health crises to pull medicine into the modern era by triggering new regulatory authorities and standards. This happened earlier in the US, where the Elixir Sulfanilamide case and its 107 victims prompted the passing of the Food, Drug, and Cosmetic Act in 1938. The law established that drugs intended to prevent or treat disease had to prove they were safe for use as labeled and receive a priori authorization by providing key data to the regulator. "For the first time, before pharmaceutical companies could market a drug, they had to show at least that the product was safe."[5] It was unclear at first what data had to be shared to prove compliance, but over time standardized assessments emerged and became standard practice across the pharmaceutical industry.

This early development of a regulatory framework for drugs meant that the US managed to protect its citizens from the health crisis that prompted the development of similar protections in Europe. The US regulator denied approval to thalidomide, a drug widely marketed in Europe as a sedative and antiemetic agent and recommended for use by women in their first trimester of pregnancy, because its manufacturer failed to show basic aspects of the product's pharmacologic and toxicologic characteristics. In the EU, many babies died and thousands were born with severe health problems. The thalidomide tragedy served as the catalyst for harmonized European pharmaceutical regulation, which is now centralised under the European Medicines Agency (EMA).

One of the key competencies of the EMA is to "provide guidance and templates […] with practical advice on how to draw up the product information for human medicines, which includes […] a package leaflet", defined as "The leaflet in every pack of medicine that contains information on the medicine for end-users, such as patients and animal owners."[6] This leaflet is the main piece of written information that citizens receive when using drugs that have been designed to help them but may harm them. Together with the medical prescription and the assistance of pharmacy staff, package leaflets are a way to protect and enforce rights, guide proper use and provide information that empowers citizens to understand the characteristics and uses of medical products, as well as ways to seek recourse should anything go wrong.[7]

---

[5] Ïbid.

[6] EMA website https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/product-information-requirements

[7] Directive 2001/83/EC of the European Parliament and of the Council established that such package leaflets must include information on: (a) the name of the medicinal product followed by its strength and pharmaceutical form, and, if appropriate, whether it is intended for babies, children or adults; where the product contains up to three active substances, the international non proprietary name (INN) shall be included, or, if one does not exist, the common name; (b) a statement of the active substances expressed qualitatively and quantitatively per dosage unit or according to the form of administration for a given volume or weight, using their common names; (c) the pharmaceutical form and the contents by weight, by volume or by number of doses of the product; (d) a list of those excipients known to have a recognized action or effect and included in the detailed guidance published pursuant to Article 65. However, if the product is injectable, or a topical or eye preparation, all excipients must be stated; (e) the method of administration and, if necessary, the route of administration. Space shall be provided for the prescribed dose to be indicated; (f) a special warning that the medicinal product must be stored out of the reach and sight of children; (g) a special warning, if this is necessary for the medicinal product; (h) the expiry date in clear terms (month/year); (i) special storage precautions, if any; (j) specific precautions relating to the disposal of unused medicinal products or waste derived from medicinal products, where appropriate, as well as reference to any appropriate collection system in place; (k) the name and address of the marketing authorisation holder and, where applicable, the name of the representative appointed by the holder to represent

3. From model cards to AI leaflets

The adaptation of the medical leaflet model to the AI and technical innovation space holds significant promise, but also challenges. The first main challenge is defining what needs to be shared in this exercise of "upfront" transparency. The complexities of doing transparency in practice have been acknowledged by the European Parliament, as evidenced by the release in 2019 of a report on "A governance framework for algorithmic accountability and transparency" [8] and the European Commission's creation of the European Centre for Algorithmic Transparency (ECAT) in 2023.[9]

In this proposal we move away from a notion of absolute transparency, which may imply sharing code or highly technical data that lay citizens may not be equipped to understand and use to protect their rights, and favor a notion of "meaningful transparency", drawing on Annany and Crawford, [10] Kaminski[11] and the excellent work of Safak and Parker for the Ada Lovelace Institute.[12] By meaningful transparency we mean information that "is realistically accessible to a member of the general public at the time of the request. It must be available in practice, not just in theory", as the ICO put it.[13] Here, we seek to make information accessible for the general public, but also regulators, civil society organizations and all relevant parties. This requires some level of "translation" of highly technical terms, but also the incorporation of non-technical information related to governance and impacts.

In order to engage in the required translation exercise, we also draw on efforts to foster the documentation of the decisions made during the development and testing of technology products, and specifically on the "Model Cards for Model Reporting" proposal developed by Mitchell *et al.* while working at Google, which have become a widespread tool.[14] In their paper, they define model cards as "short documents accompanying trained machine learning models that provide benchmarked evaluation in a variety of conditions, such as across different cultural, demographic, or phenotypic groups (e.g., race, geographic location, sex, Fitzpatrick skin type) and intersectional groups (e.g., age and race, or sex and Fitzpatrick skin type) that are relevant to the intended application domains. Model cards also disclose the context in which models are intended to be used, details of the performance evaluation procedures, and other relevant information."[15]

---

him; (l) the number of the authorization for placing the medicinal product on the market; (m) the manufacturer's batch number; (n) in the case of non-prescription medicinal products, instructions for use.

[8] Available at https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf

[9] See https://algorithmic-transparency.ec.europa.eu/index_en

[10] Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20(3), 973–989. https://doi.org/10.1177/1461444816676645

[11] Kaminski, Margot E., Understanding Transparency in Algorithmic Accountability (June 8, 2020). Forthcoming in Cambridge Handbook of the Law of Algorithms, ed. Woodrow Barfield, Cambridge University Press (2020)., U of Colorado Law Legal Studies Research Paper No. 20-34, Available at SSRN: https://ssrn.com/abstract=3622657

[12] Cansu Safak and Imogen Parker (2020) Meaningful transparency and (in)visible algorithms. Ada Lovelace Institute https://www.adalovelaceinstitute.org/blog/meaningful-transparency-and-invisible-algorithms/

[13] See ICO "Information in the public domain" https://ico.org.uk/for-organisations/guidance-index/freedom-of-information-and-environmental-information-regulations/information-in-the-public-domain/

[14] Amazon Web Services, for instance, uses a version of Model Cards called "servie cards".

[15] Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, Spitzer E, Raji ID, Gebru T. Model cards for model reporting. InProceedings of the conference on fairness, accountability, and transparency 2019 Jan 29 (pp. 220-229). https://arxiv.org/abs/1810.03993

One limitation of Model Cards which we seek to overcome with AI leaflets is their focus on the model alone. As the supply chains of AI become more complex, there is a need to develop mechanisms that capture both the existence of different developers and actors, the combination of different data sources and the likelihood of final AI implementors using AI models in ways that were not foreseen or with non-tested data.

The combination of the historical example and successes of the medical sector in protecting people and rights through meaningful, concrete practices, some industry efforts to promote greater transparency in engineering decisions, the many proposals that have emerged from civil society and public and private actors demanding action in the space of algorithmic explainability, accountability and enforcement, as well as the need from regulators to have shared standards to assess compliance, results in a proposal for an exercise of upfront meaningful transparency which we have called "algorithmic leaflet" and is described in detail in the next section

There are three notes worth highlighting before engaging in the description of the algorithmic leaflet fields. First, that the format and fields proposed are an attempt to overcome some of the issues that have made other policy tools difficult to implement in practice. Specifically, with algorithmic leaflets we suggest a format that is transparent by design in that leaflets are made available to end users, regulators and potential buyers to facilitate decision-making and understanding of how algorithmic systems work. Also, it is our experience that Data Protection Impact Assessments are often developed by legal teams who may not have access to or the skills required to assess technical processes. The leaflet we propose is highly technical in its conception to ensure that the relevant information is collected by the technical teams making the relevant decisions, and that the information is released publicly at the same time as the technology.

Second, that the algorithmic leaflet is not the only tool that can or should promote better transparency, accountability and trustworthiness around AI and technological systems and processes. The dynamic nature of many algorithmic systems means that any attempt to capture their functioning and impacts may be short-lived or incomplete, and so leaflets need to be complemented by dynamic exercises like audits, and clear instructions on how often and when to update them. AI leaflets are not a perfect tool either. But they are a good enough tool that translates and standardizes very real concerns around the need to better understand how AI systems work, to empower citizens and civil society to engage with technical systems and to provide the AI industry with clear instructions as to what constitutes compliance.

Third, we want to highlight that in order to promote the effective incorporation of AI leaflets, it is recommended that a process of expert consultation and industry piloting is designed and implemented before its terms are finalized. Standards become meaningful when they are either imposed through laws and regulations or the result of collaborative processes that allow them to consolidate. Due to the rapidly changing nature of the technical field, and the implementation challenges observed in other technology-related regulation, it is desirable that the practice and implementation tools that will need to emerge to make legal protections effective and meaningful are embraced by as many actors as possible.

4. AI leaflet template

This section starts by providing an overview of the leaflet categories. Specific definitions for each item are provided below.

Leaflet categories:

General information
○ System name/code and version (5.2 GDPR) ○ Leaflet version and version history (5.2 GDPR) ○ System owner and suppliers data ○ Suppliers' role ○ Risk level (AI Act) ○ Governance roles (Chapter IV GDPR) ○ Distribution date (5.2 GDPR) ○ Existing documentation
Information on process
○ Description of intended purposes, uses, context and role/service provided (Article 5.1.b, 5.2 and 24.1 GDPR) ○ Stakeholder involvement ○ Organizational context ○ Human role/s (Article 22 GDPR)
Information on training/validation data
○ Data sources/collection methodology (Articles 5 and 9 GDPR) ○ Data types and characteristics (Article 5.1.a, b GDPR) ○ Privacy by Design (Article 25 GDPR) ○ Datasheets for Datasets (Article 5.1.a, b GDPR)
Information on the model
○ Method/s used and justification ○ Simplified output/s ○ Decision variables ○ Objective function/s (Article 5.1.d GDPR)
Information on bias and impacts (in lab/operational settings)
○ Metrics (Articles 5.1.a and 5.1.b GDPR) ○ Protected categories (Articles 13.1.e, 14.1.e and 35.9 GDPR) ○ Impact rates per category and profile before and after each technical intervention (Article 5.1.d GDPR) ○ Auditability and audit score (Articles 5, 22, 24 and 25 GDPR)
Information on redress, if relevant:
○ Explainability profiling (Recital 71 GDPR) ○ Redress or review (Articles 13.2.f, 14.2.g and 15 GDPR) ○ Redress metrics

Definitions

System name and version: if any

Leaflet version: specify if it is the first instance. Leaflets should be revisited with a any major system change, or earlier if unsupervised machine learning is used.

System owner and supplier/s data: including contact details and name of the team in charge of product development, and any external organisation or person that has been contracted to develop the whole or parts of or the algorithmic tool.

Suppliers' role: description of the role the external supplier had in the development of the algorithmic tool.If multiple organisations have been contracted or there are multiple companies involved in the delivery of the tool, these relationships should be described clearly and concisely.

Risk level: as defined in AI Act or other relevant legislation. If a system has different risk levels in different regulations, this should be specified.

Governance roles: identification of controller/s, processor/s, DPO/s, auditor/s

Distribution date: the date the system started to operate

Existing documentation: for instance data reuse permissions/authorizations, data sharing agreements, ethics/IRB approval, DPA approval, algorithmic audit, proportionality assessment, impact assessment, transparency report, academic paper/s, GitHub/public repositories, etc. Information should be provided on whether these documents exist, where they can be found (if they are public) and who is/was responsible for developing them.

Description of the purpose and role/service provided by the algorithm, including,

- Organizational context (how the algorithmic tool is integrated into the decision-making process and what influence the algorithmic tool has on it)
- Whether it is a new role/service or the automation of an existing role/service
- Purpose of the algorithmic tool
- Description of its use
- Excluded uses (potential uses that the tool was not designed for to help avoid misconceptions about the scope and purpose of the tool)
- Benefits

Stakeholder involvement: description of any stakeholder consultation processes performed, including UX studies

Human role/s: description of how system outputs are handled. If humans are involved, description of their role and procedure to approve/reject algorithmic decisions, statistics on impact of human involvement

Data sources/collection methodology: including,

- Legal basis for access
- List of sources and link to GDPR compliance policies
- Time frame and geographical coverage of all data used, including APIs
- If the datasets are public, link to their location/repository and sharing policy
- Information on preprocessing
- Information on prohibitions stated in Article 9, GDPR

Data types and characteristics: for each data source, describe data type (number, string, image, etc.), whether data is personal and/or sensitive, and what information is included in the data (age, gender, location, etc.)

Privacy by Design: description of measures taken to minimize, anonymize or otherwise protect personal data

Datasets: name, content, format and use of all datasets involved

Method/s used and justification: linear regression, logistic regression, decision tree. SVM algorithm, Naive Bayes algorithm, KNN algorithm, K-means, random forest algorithm, etc

Simplified output/s: score, tag, categorization, recommendation, ranking, etc.

Decision variables: description of the types of variables or features used to train, test and run the model - for example 'age' or 'address'. In certain cases, it might not be feasible for a team to disclose all the variables in a dataset. In this case, teams should disclose - at a minimum: whether the data contains personal and special category information; variables of interest, such as protected characteristics and potential proxies; and variables with high predictive power or that have a significant bearing on the model.

Objective function: define the mathematical expression that represents the quantity that the algorithm seeks to optimize.

Metrics: accuracy metrics (such as precision, recall, F1 scores), metrics related to computational efficiency obtained in lab and real world settings and evaluation metrics. Include error, false Positives and false negative rates.

Protected categories identified (such as women, the old, children, those in specific locations, etc, both collected and extracted), source of protection and justification

Impact rates per category: impact can refer to selection/scoring rates or any measure of how a specific category, profile or protected group is treated by the system. If impact rates differ from group representativity (women are selected for a specific job/role less than their demographic weight), any deviation needs to be justified.

Auditability: can the values be externally verified? If so, how (using an API for instance)? Has the system been audited by an internal or independent body? If so, link to audit results.

Explainability profiling: description of all tests performed to test the explainability of the system, and the results obtained.

Redress or review: identification of the mechanisms in place for redress, appeal or review of the decision/s available to affected groups, their representatives of the general public.

Redress metrics: aggregate on redress petitions received and outcomes (decision reviewed/changed or upheld).

References

To prepare this document, a comprehensive literature review has been carried out to ensure that this proposal builds on existing models while addressing the shortcomings identified in some of them. The list below provides links to the references and texts reviewed.

The references have been organized in APA style using ChatGPT and corrected manually:

Bandy J. (2021). Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits. Retrieved from https://www.researchgate.net/publication/349125003_Problematic_Machine_Behavior_A_Systematic_Literature_Review_of_Algorithm_Audits

Blueprint for an AI Bill of Rights. (n.d.). OSTP - The White House. Retrieved from https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Gilbert T. K., Snoswell A.J., Dennis M., McAllister R & Wu C. (2022). A Sociotechnical Approach to the Design and Evaluation of Autonomous Vehicle Policy. arXiv preprint arXiv:2205.07395. Retrieved from https://arxiv.org/pdf/2205.07395.pdf

co-cddo. (n.d.). algorithmic-transparency-standard/template_table.md at main · co-cddo/algorithmic-transparency-standard · GitHub. Retrieved from https://github.com/co-cddo/algorithmic-transparency-standard/blob/main/template_table.md

Cihon P, Future of Humanity Institute. (2019). Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development. University of Oxford. Retrieved from https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf

Goicoechea, I. (2020). Omitted Variable Bias in Machine Learning models for Marketing and how to avoid it. Retrieved from https://medium.com/bedrockdbd/omitted-variable-bias-in-machine-learning-models-for-marketing-and-how-to-avoid-it-674137fb2c26

Verma S & Rubin J. (2018). Fairness Definitions Explained. University of Massachusetts Amherst. Retrieved from https://fairware.cs.umass.edu/papers/Verma.pdf

H.R. 8152. (2021). Energy Commerce. House.gov. Retrieved from http://web.archive.org/web/20221120045608/https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/BILLS-117hr8152ih.pdf

ISO/IEC TR 24027:2021(en). (2021). Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making. International Organization for Standardization. Retrieved from https://www.iso.org/standard/83841.html

Lam, M.S., Gordon M.L., Metaxa D., Hancock J.T., Landay J.A. & Bernstein M.S. (2022). End-User Audits: A Design Space for Evaluating and Improving Algorithmic Decision-Making. Stanford University.
Retrieved from https://hci.stanford.edu/publications/2022/Lam_EndUserAudits_CSCW22.pdf

Lewinson, E. (2019). Explaining Feature Importance by example of a Random Forest. Towards Data Science. Retrieved from https://towardsdatascience.com/explaining-feature-importance-by-example-of-a-random-forest-d9166011959e

Barett A.M., Hendrycks D., Newman J. & Nonnecke B. (2022). Actionable Guidance for High-Consequence AI Risk Management. arXiv preprint 2206.08966. Retrieved from https://arxiv.org/abs/2206.08966

Radiya-Dixit E, Minderoo Foundation Tech and Society Solutions Lab. (2022). A Sociotechnical Audit: Assessing Police use of Facial Recognition. Retrieved from https://www.mctd.ac.uk/wp-content/uploads/2022/10/MCTD-FacialRecognition-Report-WEB-1.pdf

Danaë Metaxa, Joon Sung Park, Ronald E. Robertson, Karrie Karahalios, Christo Wilson, Jeff Hancock and Christian Sandvig (202) Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. Retrieved from https://hci.stanford.edu/publications/2021/FnT_AuditingAlgorithms.pdf

Leslie D. (2019) Project ExplAIn. (n.d.). The Alan Turing Institute. Retrieved from https://web.archive.org/web/20230204123556/https://www.turing.ac.uk/news/project-explain

Harini, S. & Guttag, J. V. (2019). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. arXiv preprint 1901.10002. Retrieved from https://arxiv.org/abs/1901.10002

Think Tank. (n.d.). Auditing the quality of datasets used in algorithmic decision-making systems. European Parliament. Retrieved from https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729541

Minkkinen M., Laine J. & Mäntymäki M (2022). Continuous Auditing of Artificial Intelligence: A Conceptualization and Assessment of Tools and Frameworks. Retrieved from https://www.researchgate.net/publication/364156896_Continuous_Auditing_of_Artificial_Intelligence_a_Conceptualization_and_Assessment_of_Tools_and_Frameworks


edpb
European Data Protection Board