Algorithm Audit

Public standard profiling algorithms

Qualitative and quantitative safeguards for responsible use of profiling algorithms

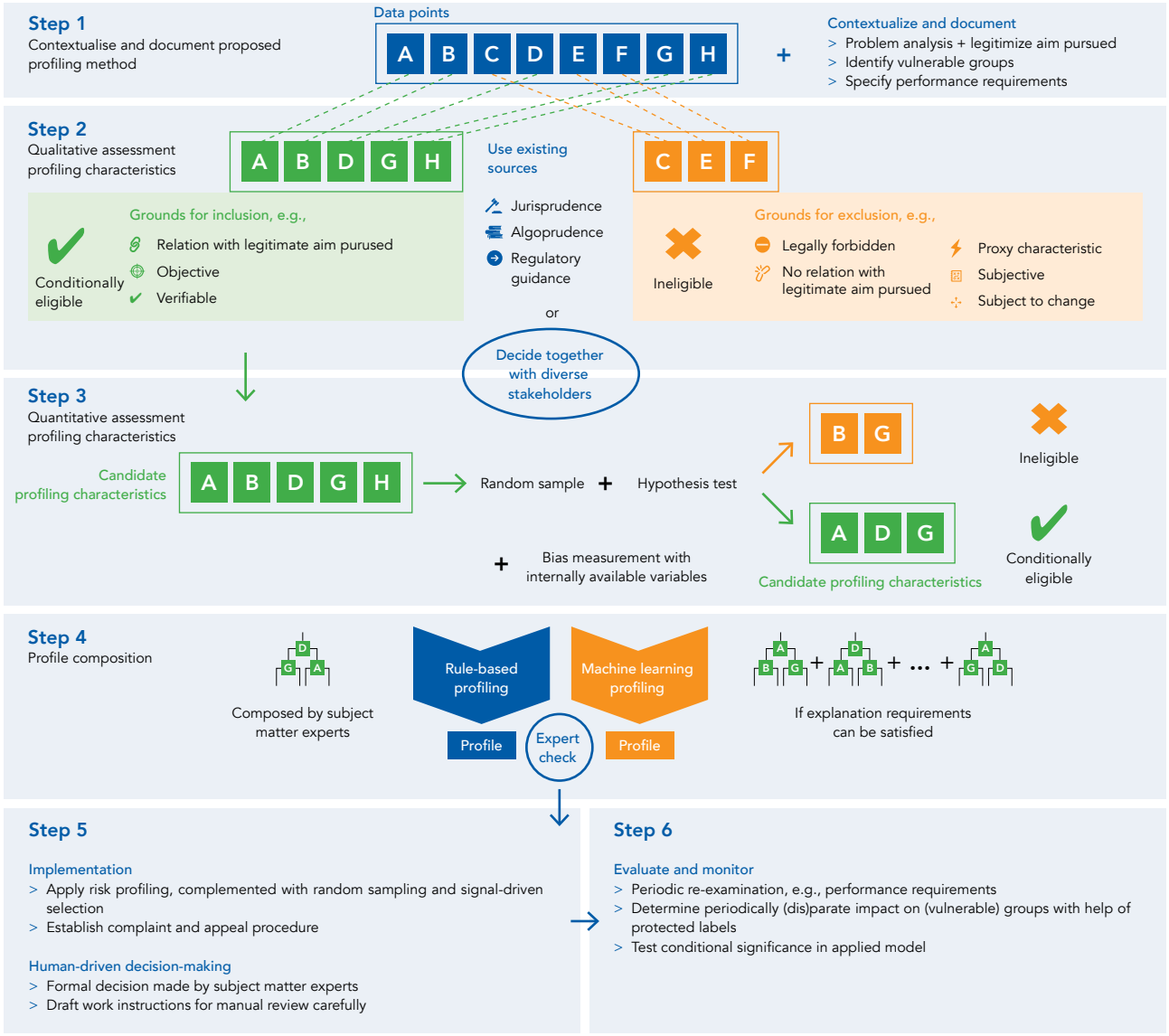
October 2024

Summary

This document outlines a step-by-step guide for the responsible use of profiling algorithms in the public domain. This standard applies to both machine learning and rule-based algorithms. By examining algorithms against this standard, (indirect) discrimination and other undesirable effects of profiling can be combated. The method follows a step-by-step process consisting of a qualitative and quantitative assessment and additional organizational control measures. The assessments help to fill in openly formulated requirements for profiling methods in anti-discrimination law, the European AI Act and other laws and regulations. The methods have been developed from practical experience with validating profiling algorithms. The standard focuses primarily on application in the public domain, but can also be utilized in the private sector.

Step-by-step guide

The infographic below briefly outlines the step-by-step guide. A full explanation is further developed in Handout – Public standard profiling algorithms.



Qualitative and quantitative safeguards for responsible use of profiling algorithms

A full explanation is further developed in Handout – Public standard profiling algorithms.

Explanation step-by-step guide

Step-by-step guide

Below step-by-step plan elaborates on how profiling algorithms can be used responsibly. However, it does not provide any guarantees for this, as it depends on the way in which the steps are carried out and the corresponding choices that are made. The requirements from the Algorithm Framework (Algoritmekader of the Dutch Ministry of the Interior) are the guiding principles for all the different phases in the life cycle of the profiling algorithm under review.[1] This standard is specifically aimed at profiling algorithms. It is therefore an addition to Fundamental Rights Impact Assessments (FRIAs), where this guide gives a better understanding of assessing specifically profiling methods. This standard does not address requirements for information security. More information regarding this can be found in Baseline Information Protection Government (BIO). The General Data Protection Regulation (GDPR) serves as the primary framework for the lawful processing of personal data.

Step 1 – Contextualize and document proposed profiling method

This step can potentially be integrated into other methods, such as conducting a Fundamental Rights Impact Assessment (FRIA), applying the Algorithm Frameworks or the Algorithm Research Framework of the Netherlands National Audit Service (ADR).

1.1 Describe the legal basis on which enforcement and supervision through risk profiling is based.

1.2 The following aspects must be carefully motivated:
- > The problem the algorithm is intended to solve;
- > The consideration whether, and if so which type of algorithm is the best to solve the identified problem efficiently.

1.3 Identify vulnerable groups in the population.[2] Investigate what the adverse effects could be for these groups.

1.4 Examine and document which quantitative bias metric is most relevant for the given context.[3]

1.5 Document performance requirements.

1.6 Identify the available variables in the database that represent a characteristic of natural persons or organizations.

---

[1] The requirements and measures from the Algorithm Framework align with the provisions of the AI Act when the profiling algorithm falls under the definition of an AI-system. If the profiling method does not fall under the high-risk category of the AI Act, the Algorithm Framework can still be used as a management framework for handling 'impactful algorithms'.

[2] Protected grounds under non-discrimination law: religion, belief, political opinion, race, sex, nationality, heterosexual or homosexual orientation or civil status (Dutch context). And also grounds that are not formally protected by law, but on the basis of which discrimination may still be ethically undesirable, such as obesity, level of education and professional appearance.

[3] The relevant quantitative metric depends per context.

Step 2 – Qualitative assessment of profiling characteristics

2.1 Review whether existing normative judgments are available for the correct use of profiling characteristics in a comparable context. Consider: case law, algoprudence[4] and regulatory guidance. If this information is available, skip Step 2.2 and perform the subsequent steps based on the available judgments.

2.2 Establish a diverse group of stakeholders, consisting of among others an algorithm developer, a subject matter expert, citizens subjected to the algorithm or their representatives, and legal, statistical and ethical experts.[5]

2.3 Together, review the characteristics identified in Step 1.6 and determine whether each characteristic meets the following or other possible grounds for exclusion:

> Prohibited by law: prohibited differentiation based on non-discrimination law[2], or unlawful processing of personal data in the context of the algorithm according to the GDPR;

> No relation with legitimate aim pursued: characteristic has no clear and substantive relationship with the aim pursued by the algorithm;

> Proxy characteristic: characteristic has a strong relationship with a vulnerable group[6];

> Subjective: characteristic cannot be measured objectively and is based on a subjective value judgment;

> Subject to change: characteristic is unreliable because it is based on a snapshot of a characteristic that changes over time.

Review the remaining criteria and check whether they meet the following ground for inclusion and motivate why:

> Relation with aim pursued: characteristic has a clear and substantive relationship with the aim pursued by the algorithm;

> Objective: characteristic is independent of subjective perception or value judgment;

> Verifiable: the correctness of a characteristic can be checked.[7]

The characteristics that do not meet the grounds for exclusion but do meet the grounds for inclusion proceed to Step 3 and are referred to as *candidate profiling characteristics*.

Step 3 – Quantitative assessment of profiling characteristics

3.1 Take a random sample from the target population.

3.2 Formulate a hypothesis about the relationship between profiling characteristic and the aim pursued.

3.3 Apply statistical hypothesis testing to the random sample and examine whether there exists a statistically significant relationship.[8]

3.4 If there is no statistically significant relationship, remove the characteristic from the set of candidate profiling characteristics.[9]

---

4 Algoprudence: transparent collective judgements of responsible regarding use of algorithms. https://algorithmaudit.eu/knowledge-platform/knowledge-base/white_paper_algoprudence/

5 Guidelines for this process: https://algorithmaudit.eu/algoprudence/how-we-work/#guidelines

6 For instance: differentiation based on Dutch language is strongly related to migration background. Differentiation based on technical professions is strongly related to gender.

7 Savings in a foreign bank account is an example of a characteristic that is objective, but not always verifiable. An indicated amount is objective, but may not always be verifiable by a governmental institution.

8 An example can be found in section 3.1 of the report *Preventing prejudice*, Algorithm Audit (2024)..

9 In the handout guidance is provided what statistical test to use in specific cases and how to deal with multiple hypothesis testing.

3.5 If possible, perform a bias test using internally available data on vulnerable groups.[10]

3.6 Weigh the quantitative insights from the bias test using the qualitative method from steps 2.1-2.2.

Step 4 – Profile composition

4.1 Create a risk profile based on remaining candidate profiling characteristics. This profile can be composed by subject matter experts or by a variable selection algorithm. Document and motivate the choice how the profile is composed. Public sector organizations can only apply variable selection algorithms (machine learning) if explainability requirements can be met.[11]

4.2 Validate a composed candidate risk profile by (external or internal, but independent) competent experts who were not involved in the design process of the risk profile.

Step 5 – Implementation

5.1 Determine a division in the population that samples persons or organizations based on the risk profile, based on random sampling and by signal-driven sampling (for example reported complaints or other signals from within the organization). A suggested ratio is 2:1:1.

5.2 Establish a procedure to uphold the rights of affected parties, such as the possibility of complaint and appeal procedures. Ensure that complaints are picked up by the organization.

5.3 Let domain experts make the final decision to actually investigate people or organizations selected by the risk profile. Carefully establish work instructions for domain experts. Avoid repetition of the profiling characteristics in the step of algorithmic profiling and the step of manual inspection by domain experts. Take vulnerable groups into account. Document why a person or organization is ultimately selected for a control procedure, in such a manner that explainability requirements are met.

5.4 Manage the model using version control.

Step 6 – Evaluation and monitoring

6.1 Periodically repeat this guide, taking into account any changed circumstance.

6.2 Determine periodically the composition of groups selected for the control procedure. Labels for protected grounds can be used, available either internally in the organization or by request at the national office of statistics. Alternatively, if protected group labels are not available, a clustering analysis can be performed to assess what groups deviate from the average performance of the profiling method.[12] Test the results against the previously established accepted amount of bias per group.

6.3 Determine the predictive value of criteria given the risk profile using conditional significance testing.[13]

6.4 If during a periodic test the defined requirements are not met, stop the profiling method and archive it.

---

[10] Note that data on gender, language, age, socio-economic status do not always fall under Art. 9 of the GDPR and can therefore be processed for this purpose.

[11] Explainability requirements for ML-driven risk profiling can be found in algoprudence *Risk profiling for social welfare re-examination* (ALGO:AA:2023:02:A).

[12] More information on bias testing without access to protected labels: https://algorithmaudit.eu/technical-tools/bdt/

[13] Does the difference in predicted outcome deviate significantly from zero when a profiling characteristic is removed from the profile?

About Algorithm Audit

Algorithm Audit is a European knowledge platform for AI bias testing and normative AI standards. The goals of the NGO are three-fold:

Normative advice commissions

Forming diverse, independent normative advice commissions that advise on ethical issues emerging in real world use cases, resulting over time in algoprudence

Technical tools

Implementing and testing technical tools for bias detection and mitigation, e.g, bias detection tool, synthetic data generation

Knowledge platform

Bringing together experts and knowledge to foster the collective learning process on the responsible use of algorithms, see for instance our AI Policy Observatory and position papers

Structural partners of Algorithm Audit

SIDN Fund

The SIDN Fund stands for a strong internet for all. The Fund invests in bold projects with added societal value that contribute to a strong internet, strong internet users, or that focus on the internet's significance for public values and society.

European AI&Society Fund

The European AI&Society Fund supports organisations from entire Europe that shape human and society centered AI policy. The Fund is a collaboration of 14 European and American philantropic organisations.

Dutch Ministy of the Interior and Kingdom Relations

The Dutch Ministry of the Interior is committed to a solid democratic constitutional state, supported by decisive public management. The ministry promotes modern and tech-savvy digital public administrations and govermental organization that citizens can trust.

Building *AI auditing* capacity from a *not-for-profit* perspective

www.algorithmaudit.eu

www.github.com/NGO-Algorithm-Audit

info@algorithmaudit.eu

Algorithm Audit

Stichting Algorithm Audit is registered as a non-profit organisation at the Dutch Chambre of Commerce under license number 83979212