Algorithm Audit

Empirical methods for supervising algorithmic profiling systems

Assessment protocol for examining indirect discrimination

June 2025

Summary

When regulating algorithmic profiling systems, not only legal but also statistical information plays a key role. Using a Dutch public sector risk profiling algorithm as an example, we demonstrate that current frameworks, guidelines and soft law fall short in providing sufficient guidance for the interpretation of open norms within European non-discrimination law. We show that established methods from empirical science can help clarifying these norms. Building on the case-based example, we propose an assessment protocol designed to assist supervisory authorities in formulating targeted questions to examine indirect discrimination through algorithmic profiling systems used in the public and private sector. This approach builds upon existing legal frameworks, enabling supervisory authorities to effectively monitor algorithm-driven decision-making processes, even with limited resources.

Assessment protocol for examining indirect discrimination

Organisational responsibilities

1	Which algorithms and AI systems are used within the organization?

Note that algorithms and AI systems are not always recognized as such

Analysis of problem

2	Does the process in which the profiling method is used pursues a legitimate aim?
3	What problem is solved with the use of the profiling method?
4	Have vulnerable groups in the population been identified? Has the adverse impact on these groups been examined? If not, why not?

Data quality

5	Was a random sample of the population drawn during the development phase of the profiling method? If not, was the used data during development analyzed to assess its representativeness of the population?
6	To what extent are labels assigned by employees reliable?
7	Are profiling features excluded from the profiling method in advance, because they are subjective, subject to change or known as proxy features for a protected characteristic?
8	Has the proxy nature of profiling characteristics been established by analysis of population statistics? If not, why not?
9	Are selected profiling characteristics linked to aim pursued? Are the profiling characteristics objective and verifiable?

Profiling algorithm

10	Was a random sample selected to compare against the results of the profiling algorithm? If not, why not?
11	Have assumptions in the profiling method been tested for suitability based on a statistical hypothesis test? If not, why not?
12	Is the necessity of the profiling algorithm evaluated by examining alternatives? If not, why not?
13	Has equal treatment of vulnerable groups been investigated by analysing population statistics? If not, have other analyses been carried out to identify possible unequal treatment?
14	Has the effectiveness of the profiling algorithm been determined? Are corrections, as a results of unequal treatment been investigated and are its effects on the expected effectiveness of the algorithm quantified?
15	How was the proportionality trade-off between equal treatment and effectiveness assessed?

Use

16	Which portion of the conducted investigations was selected using the profiling method? What portion of the selected group was selected randomly?
17	Is the composition of the selected group for an investigation being monitored? Has correction of the selection been considered? If not, why not?

1. Introduction

Within the democratic rule of law, supervisory authorities play an important role in providing legal protection and legal certainty for citizens, consumers and organisations. To fulfill this role with regard to the responsible use of algorithms and artificial intelligence (AI), new skills are required. Quantitative methods play an increasingly important role in this. For example, in research into a risk profiling algorithm in the College Grant Check (controle uitwonendenbeurs, abbreviated as CUB) process of the Netherlands Executive Agency for Education (DUO), indirect discrimination was established on the basis of a data study in which random samples, statistical hypothesis tests and the migration background of more than 300,000 students in the period 2012-2023 were retrospectively analysed.[1] Building on this quantitative study, controlling state powers – both the judiciary and the Dutch Parliament – have been able to formulate a concrete normative judgement regarding the control process.[2] [3]

However, it is the task of supervisory authorities to proactively prevent such fundamental rights violations. With the ever-increasing digitization and the limited extent to which organizations are able to manage the risks of AI systems, questions arise.[4] How is it possible that after the Dutch childcare benefits scandal *"very little has changed"* to prevent discriminatory algorithms used by the Dutch public sector organisations? [5] Will the AI Act strengthen the supervision of algorithms by developing concrete standards for profiling algorithms? And how can supervisory authorities carry out their tasks more effectively with limited resources?

This article introduces an assessment protocol on the basis of which supervisory authorities can provide targeted supervision of indirect discrimination through algorithms and AI. In doing so, it builds on the proposal formulated in the Risk Profiling Assessment Framework (Toetsingskader risicoprofilering) of the Netherlands Institute for Human Rights (NIHR) to use empirical methods to investigate indirect discrimination.[6] On the one hand, this is a topical and timely issue – the use of algorithms and AI systems will only increase in the near future. On the other hand, the DUO case shows that methods from empirical science, which have been established for decades, are only slowly finding their way into policy and supervisory practice.[7]

---

1 Misusage college grant PwC (2024), Preventing predjuice Algorithm Audit (2024) and Addendum Preventing prejudice, Algorithm Audit (2024).

2 District court of Overijssel 29 October 2024, ECLI:NL:RBOVE:2024:5627.

3 Dutch Parliamentary papers 2023/24 D21614, Dutch Parliamentary papers 2023/24 24724 nr. 229, Dutch Parlimantery papers 2023/24 24724 nr. 231.

4 The report Focus op AI bij de Rijksoverheid Dutch Court of Auditors (2024) shows that for 35 percent of AI systems currently in use by Dutch public sector organisations, the performance is unknown. This outcome, among other things, led to 82 parliamentary questions from the permanent committee on digital affairs, Dutch Parliamentary Papers Z16114.

5 Interview Aleid Wolfsen – chairman of the Dutch Data Protection Authority, de Volkskrant (2024).

6 Toetsingskader risicoprofilering – Normen tegen discriminatie op grond van ras en nationaliteit, the Netherlands Institute for Human Rights (2025), in particular p.25-30.

7 Since 2003, the Dutch national legislation enables government organisations to request specific population statistics for statistical research purposes. Drawing random samples and performing statistical hypothesis tests to measure effects, for example of medicines, has been a tried and tested method since the 1950s.

This white paper has been written together with researchers of the Netherlands Bureau for Economic Policy Analysis (Centraal Planbureau) based on a joint article in Dutch, see Jurriaan Parie, Ylja Remmits, Brinn Hekkelman and Mark Kattenberg (2025) *Praktische handvatten voor empirisch toezicht op profileringsalgoritmes*, Tijdschrijft voor Toezicht.

DUO's risk profiling algorithm illustrates how empirical methods can help to resolve normative value tensions, among others between effectiveness and unequal treatment. In hindsight, it turned out that the profiling characteristics from the risk profile (education level, age and distance to parent(s)) were strongly related to students' migration background. Based on aggregated data provided by Statistics Netherlands (STATISTICS NETHERLANDS), it was possible to determine exactly at population level that in 2014, 63.3 percent of the 16,023 vocational training (MBO 1-2) students and 13.2 percent of the 104,814 university (WO) students had a non-European migration background. Since a higher risk score was assigned to MBO 1-2 students in the risk profile, a higher risk score was therefore indirectly assigned to students with a migration background. On the basis of these quantitative insights, it was possible to make an informed assessment of whether the distinction made was proportionate.

In March 2024, the Minister of Education, Culture and Science (OCW) came to a decision: on behalf of the cabinet, Mr. Dijkgraaf apologized for indirect discrimination in the CUB process.[8] In November 2024, a compensation scheme for more than 10,000 students worth more than €61 million was announced.[9] At the same time, there remains parliamentary support for the use of profiling by the Dutch government. In the same debate in which the apologies were offered, Minister Dijkgraaf stated that: *"[...] we are aware of the value of risk-based supervision. […] It is important that risk-based supervision has the right safeguards."*[10] Quantitative methods such as those discussed in this article will play an important role in this.

The DUO case is not an isolated one. Ill-considered risk profiling is exemplary of the broader use of algorithms by Dutch public sector organizations. *"With just about every tile we lift, we discover discriminatory algorithms,"* says Aleid Wolfsen – chairman of the Dutch Data Protection Authority (AP).[11] Discrimination through profiling is also a risk in the private sector. In February 2025, journalists of Argos reported that Rabobank used discriminatory features in transaction monitoring systems and the NIHR ruled that Meta indirectly discriminates based on gender when showing advertisements for vacancies.[12] [13] In this case as well, data studies have played an important role in demonstrating unequal treatment.

Now that profiling algorithms have penetrated the digital fabric of society, it is important that supervisory authorities offer citizens, consumers and organisations legal protection against discrimination. To this end, this article analyses the framework of non-discrimination law (section 2). Subsequently, based on the DUO case, empirical methods are discussed that are helpful in testing profiling algorithms against legal frameworks (section 3). We will then introduce an assessment protocol based on which supervisory authorities can formulate targeted inquiries for algorithm producers and deployers about indirect discrimination through profiling algorithms (section 4). The article ends with a conclusion (section 5).

8 Kamerstukken II 2023/24 24724 nr. 220.

9 Kamerstukken II 2023/24 24724 nr. 243.

10 Kamerstukken II 2023/24 D07566.

11 Supra note 5.

12 Rabobank discrimineerde bij klantcontroles | NPO Radio 1 (2025).

13 Verdict: Meta Platforms Ireland Ltd. maakt verboden onderscheid op grond van geslacht bij het tonen van advertenties voor vacatures aan gebruikers van Facebook in Nederland, College voor de Rechten van de Mens (2025); Investigation that gave rise to the complaint: New evidence of Facebook's sexist algorithm, NGO Global Witness (2023).

2. The open legal norm: who is responsible for its interpretation?

Distinction is ever present. Both in classical antiquity and in today's society, a distinction is made for access to education, employment and housing, among other things. Today, however, some forms of discrimination are prohibited. Constitutions, European treaties and national legislation prohibit direct discrimination based on protected grounds, such as ethnicity and nationality.[14] Prohibited discrimination can also occur indirectly through so-called *proxy characteristics*. In that case, distinction is not made directly on protected grounds, but on seemingly neutral selection criteria that lead to a disproportionate disadvantage for these groups. Examples of proxies for the protected grounds of ethnicity and nationality are postal code, level of income, license plate, family member abroad and low literacy.[15] Since correlations occur by definition in statistical modelling, profiling characteristics always have a proxy character to a greater or lesser extent. Making a distinction based on proxy characteristics is not necessarily prohibited. However, it must be possible to objectively justify its use. European non-discrimination law provides a framework for this. This section analyses how the numerical world of algorithms and AI relates to the letter of the law, including developed soft law frameworks.

2.1 Legislation: too general for responsible profiling algorithms

The General Data Protection Regulation (GDPR), the Dutch Public Administration Law (Awb) and Dutch Equal Treatment Law (Awgb), among others, regulate the use of profiling algorithms. The most concrete standards for equal treatment come from non-discrimination law. The Awgb, treaties of the European Union and the Council of Europe state that there prohibited indirect discrimination occurs when an apparently neutral provision, criterion or practice particularly affects persons in a protected group compared to other persons. There is no discrimination if the distinction made can be objectively justified. This justification test consists of three parts:[16]

> Suitable: are the profiling characteristics sufficiently relevant and objective to contribute in a (more) effective way to the achievement of the legitimate aim pursued?
> Necessary: are there other, less invasive, ways to achieve the aim pursued?
> Proportionate: does the legitimate aim pursued carry enough weight to justify the distinctions made?

The above norms describe what the general legal frameworks are, how these open norms should be implemented for profiling depends on the context. In the DUO case, it is not immediately clear how the suitability, necessity and proportionality of the three profiling characteristics can be tested.

The AI Act (AIA) does not appear to offer more concrete guidance. Product safety legislation ensures the responsible development and use of AI systems by public and private organisations, protecting the safety, health and fundamental rights of European Union (EU) citizens. According to the AIA, specific applications of AI systems that have been designated as 'high risk' must comply with harmonised standards – also known as CE

[14] The prohibition of discrimination follows from Treaties of the European Union (EU) and the European Convention on Human Rights (ECHR). In the Netherlands, the prohibition of discrimination is laid down in Article 1 of the Constitution and elaborated in the Awgb. Article 1 Awgb mentions the following protected grounds: "religion, belief, political opinion, race, sex, nationality, heterosexual or homosexual orientation or marital status". Protected grounds listed in EU treaties vary by context. For example, age is a protected ground in the Employment Equality Directive (Art. 1 2000/78/EC), but not in the Racial Equality Directive (Art. 2 2000/43/EC).

[15] Supra note 6.

[16] The objective justification test only applies if a legitimate aim is pursued. In the case of DUO, this is met, namely: compliance with the College Grant Act 2000 (Wsf2000).

marking.[17] The European Commission has therefore requested standardisation organisation CEN-CENELEC to develop 10 standards for high-risk AI systems that provide the presumption of conformity with the AIA, including a risk management and quality management system.[18] However, there are concerns among experts as to whether, under the considerable time pressure, useful standards for specific applications of AI systems, such as profiling, will emerge from this standardisation request.[19] The harmonised standards will primarily be procedural and will not provide prescriptive guidance for explainability, human intervention and non-discrimination. In addition, many Dutch profiling algorithms, such as the DUO algorithm do not qualify as an AI system, which means that the AIA's requirements will not apply to this type of algorithm.[20] As a result, it is necessary for national legislators and regulators to continue to explore national initiatives to concretize open legal norms for the responsible use of profiling algorithms.

2.2 Soft law frameworks: too soft for responsible profiling algorithms

The need for clearer guidance on the responsible use of profiling algorithms has not gone unnoticed by supervisors and policymakers. To provide more clarity on how open norms should be interpreted various institutions have developed soft law frameworks. The practical useful of these frameworks is demonstrated through the DUO case.

The Dutch Ministry of the Interior and Kingdom Relations has brought together various guidelines, instruments and frameworks that have been developed in recent years in an *Algorithm Framework*.[21] In the Algorithm Framework, reference is made to the NIHR's Risk Profiling Assessment Framework for an *"explanation of how to carry out the justification test"*.[22] The Algorithm Framework and Profiling Assessment Framework had not yet been published in the summer of 2023, when journalists reported a suspected overrepresentation of students with a migration background during the college grant control process. At the time, the researchers consulted other frameworks that offered little guidance to test the specific profiling characteristics against the applicable standards from non-discrimination law.[23] [24] When unequal treatment is suspected, how to assess the suitability, necessity and proportionality of the profiling features?

If there is a suspicion of unequal treatment, the next step is to investigate whether an apparently neutral profiling method disadvantages a person or group sharing the same characteristics. In the DUO case, this means checking whether students with a migration background have been disproportionately affected by proxy characteristics in the profiling algorithm. The Profiling Assessment Framework specifies that *"'disproportionate' means that applying a risk profile increases the chance for an investigation due to a characteristic falling under 'race' or nationality above the average chance of being selected for investigation in the case of random sample checks."*

---

17 Among others Article 1(2) and Article 40 of the AI Act.

18 Draft standardisation request amending implementing decision C(2023)3215 on a standardisation request in support of Union policy on artificial intelligence, Europese Commissie (2023).

19 The authors affiliated with Algorithm Audit are active in various CEN-CENELEC working groups through the Dutch standardization body NEN.

20 Guidelines for AI Regulation Implementation – Definition of an AI System, Algorithm Audit (2025).

21 Algorithm Framework v2.1, Ministry of the Interior and Kingdom Relations (2025).

22 Supra note 6.

23 Students with a migration background remarkably often accused of fraud, minister wants to thoroughly investigate system, NOS (2023).

24 Impact Assessment Human Rights and Algorithms (IAMA), Ministry of the Interior and Kingdom Relations (2021) and the Research framework Algorithms of the Central Government Audit Service (2023).

The reference to average probabilities and random sampling presupposes a quantitative methodology to examine this proportionality trade-off.

The next step is to carry out the justification test. This starts with assessing the appropriateness of profiling characteristics. The Profiling Assessment Framework explains that *"The essence of this test is to determine whether the use of risk profiling actually contributes to the identification of more violations of norms than if only random checks were used."*. However, the framework does not prescribe how the *"effectiveness of each of the individual profiling characteristics"* must be established. The Profiling Assessment Framework does state that this is "*in fact an empirical test*". In the DUO case the effectiveness of investigations that were randomly selected from 2014 and 2017 was 3.6% (n=387) and 3.8% (n=293) respectively. The effectiveness of the CUB process – of which the risk profile was a part – produced an effectiveness of 38.9% and 35.3%.[25] Due to the interconnectedness of the application of the risk profiling algorithm and manual selection in the control process, only the effectiveness of the entire control process could be determined in the DUO case. How the appropriateness of individual profiling features has been empirically tested we discuss in Empirical methods.

When it has been demonstrated that the risk profile increases the effectiveness of investigations, the next step in the justification test is to examine its necessity. It must be motivated whether there are no less invasive alternatives for achieving the legitimate aim pursued. The Profiling Assessment Framework states that *"comparing alternatives is a matter of measuring and comparing effectiveness but also requires a trade-off: an alternative profile that has less or no (indirect) profiling characteristics also counts as a reasonable alternative if it costs marginally more and or is marginally less effective than the initial risk profile"*. Compared to the random sample, the risk profile is effective, but DUO did not sufficiently investigate which alternatives were available during the development of the algorithm.[26] This does not sufficiently motivate the need for the use of the risk profile.

The Profiling Assessment Framework states that if the profiling method is considered appropriate and necessary, the proportionality of the risk profile should be weighed: *"are the objectives of the risk profiling proportionate to the negative effects that this form of risk profiling causes"*? To this end, all *"positive and negative effects of the risk profile should be carefully mapped out"*. There is no standardized solution for resolving the trade-off between effectiveness and infringement of equality rights. Based on the DUO case, in Empirical methods we describe data analysis techniques that inform this proportionality assessment.

3. Empirical methods

Digital information systems enable organizations to apply empirical methods at scale during the development and use of algorithms.[27] DUO's CUB process could be carefully examined because data on random samples and the migration background of students could be analysed retrospectively for the period 2012-2022. The empirical methods used in this study are discussed in more detail in this section. In addition, it is explained how the methods can support the objective justification test.

3.1 Data quality and the random sample

If empirical methods are used to inform the objective justification test, the available data must be reliable. At least the following questions must be answered positively: Are the data representative for the corresponding population? Are the data points

[25] Supra note 1.

[26] Supra note 1.

[27] Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, Cass R Sunstein, Discrimination in the Age of Algorithms, Journal of Legal Analysis, Volume 10, 2018, Pages 113–174, https://doi.org/10.1093/jla/laz001.

objective and verifiable? Can manually assigned labels be trusted? Trust in data quality stands or falls with documentation standards. This includes documentation of the meaning, reliability and topicality of the data. Organisations affiliated with the Dutch Ministry of Justice and Security (JenV), for example, have set up the JenV Data & Algorithms Scheme, which includes policy and protocols how data should be processed.[28]

During research into the DUO case, the available data on students living on their own were reliable. The values of the profiling characteristics (type of education, age and distance to parents(s)) could all be determined unambiguously and objectively.[29] However, the data on the outcomes of the CUB process were not representative, because young MBO students were checked disproportionately often and this group is not representative of the entire population of students living away from home. Due to this so-called *magnifying glass effect*, the data on the results of the checking process is therefore not representative. The housing situation of young MBO students is usually different from the housing situation of university students.[30] As a result, there is insufficient information available about the effectiveness of the CUB process for higher professional education and university students.

In addition, the quality of the labels – in the DUO case the results of the inspection process ('lawfully' or 'unlawfully' received college grant) – must be trustworthy. It is a possibility that due to human bias or ambiguous work instructions, inspectors will more often judge certain students' grant receival as 'unlawful'. In that case, the collected labels are biased and possibly incorrect.[31] It is also important that successful objection procedures are corrected in the data. In the DUO case, the reliability of the labels and any bias in them was not investigated as such, because unequal treatment was the purpose of the study on the basis of a data study of all the different steps in the CUB process.[32]

Drawing a random sample combined with a carefully designed data collection process is therefore a best practice.[33] In a random sample, data subjects are selected for verification without a profiling method being used to select persons or organisations for an investigation. Preferably, a profiling method is designed solely on the basis of data drawn using a random sample. When this is not possible, the profiling method should be evaluated based on the results of a random sample drawn parallel to the algorithmic-driven selection process. In the DUO case, the CUB risk profile that had already been applied was not based on a random sample. The risk profile used was evaluated based on 387 and 293 randomly selected students for an investigation in 2014 and 2017. [34] Based on these random samples, assumptions in the risk profile were tested.

3.2 Testing assumptions

If the data is trusted, empirical methods can be used to support the objective justification test. Based on the random sample, the appropriateness of profiling characteristics from a risk profile, the necessity of the entire model, and the proportionality can be

[28] JenV Data Standard Reference Data, Judicial Information Service (2023).

[29] Education data (MBO 1-2, MBO 3-4, HBO or WO) of students are collected by DUO from educational institutions, age is tracked via the Basic Registration of Persons (BRP) and distance to parent(s) is determined internally on the basis of the distance between the postal code of the parents' address and the student.

[30] National Student Housing Monitor 2014, Knowledge Centre for Student Housing.

[31] A so-called noise study can provide insight into the extent to which human assessments differ from each other, see Kahneman, D., O. Sibony and C.R. Sunstein, 2021, Noise. See also 'Bias experiment influence labels on decision', Parliamentary Papers II 2023/24 2024D17779.

[32] Supra noot 1.

[33] Report AI & Algorithm Risks Netherlands, Edition 3 (2024).

[34] Supra noot 1.

investigated. The methods are applied to the DUO case.

On the basis of random samples drawn by DUO, it can be statistically investigated whether there is a link between profiling characteristics from the risk profile and unduly use of the college grant.[35] Assumptions from the profile can be tested by means of a hypothesis test: is it indeed the case that younger students are more likely to make unlawful use of the college grant than older students? What about older students who live far away from their parents? After applying the statistical hypothesis test (Z-test) to frequency counts observed in the random sample, only the characteristic 'distance to parent(s)' turned out to have predictive value.[36] The other characteristics are therefore considered not to be appropriate from this simple statistical test, regardless of possible qualitative justifications.

More advanced statistical methods can also be used to investigate the suitability of profiling features. This is necessary when different profiling characteristics are interrelated. This may be the case when there is no link between the feature and the outcome for individual profiling characteristics, while there is such a link within the risk profile used. The risk profile would then still use inappropriate characteristics.[37] This can be overcome by investigating the individual relationship between a characteristic and the outcome by testing conditional statistical significance.[38]

The random sample is also relevant to assess the necessity of a profiling method. The effectiveness of profiling characteristics considered appropriate should be compared with the effectiveness of the random sample. As mentioned earlier, the effectiveness of the random samples drawn by DUO from 2014 and 2017 was 3.6% and 3.8% respectively. The effectiveness of a risk profile must be compared to these figures. These results inform the proportionality assessment. This procedure must be repeated for each suitable characteristic added to the profiling method.

3.3 Population statistics

One of the most complex aspects of mapping the effects of the risk profile is determining the proxy nature of the profiling characteristics. A sound methodology to determine the proxy character is a data study based on population statistics. If the degree of indirect discrimination has been determined, this must be weighed against the necessity and appropriateness to distinguish on the basis of this characteristic during the proportionality assessment. For the DUO case, we explain how the proxy nature of the profiling characteristics has been determined.

First of all, it must be determined in respect of which protected characteristic the proxy character is determined. European non-discrimination law states that 'race' and ethnicity are protected grounds at all times.[39] This is not always the case for age.[40] Many public sector organisations, but also other institutions such as banks, have data at their disposal (such as nationality) based on which the proxy nature of profiling characteristics can be determined. However, these special categories of personal data

[35] Sample size guidelines can be found in 'Size random sample', Algorithm Audit (2024).

[36] Supra note 1. Note that 'distance from parents' only has predictive value in the analysis of the 2014 sample and not in the 2017 sample. This is partly due to differences in the underlying population. In 2014, the population consisted of students in secondary vocational education (MBO), higher professional education (HBO) and university education (WO). In 2017, following the abolition of the college grant, the sample consisted only of MBO students. The predictive value of the characteristic 'distance to parent(s)' therefore depends on the specific purpose for which the profiling method is applied and must be examined and substantiated for each situation.

[37] The opposite can also be true, so that the risk profile is deprived of relevant variables and is less effective.

[38] In this way, two versions of the risk profile can be made, a model with and without the relevant characteristic, and it can then be tested whether there is a difference in both predictions.

[39] 'Race' is a legal term for personal characteristics such as skin colour, ethnicity and national or ethnic origin. Supra notes 6 and 14.

[40] Supra note 14.

may not always be processed according to the GDPR.[41] The GDPR itself offers some exceptions to be allowed to process this data for research into possible bias and the AI Act also offers an exception for this. [42] [43] If these data are not available, or cannot be processed for this purpose, Dutch public sector organisations can submit a request to the Dutch national office of statistics Statistics Netherlands.[44] Based on population statistics, the proxy nature of profiling characteristics can be determined this way.[45]

During the DUO study, aggregation statistics provided by Statistics Netherlands were used to determine the proxy nature for the profiling characteristics (type of education, age and distance to parent(s)) for the protected ground 'migration background'. The results are shown in Figure 1. The red line shows the proxy nature per profiling attribute. It follows that type of education has a strong proxy character: 13.2% of university students have a non-European migration background, compared to 63.3% of students in vocational education (MBO 1-2). In terms of age, older students are more likely to have a non-European migration background. The further away students live from their parent(s), the more often they are of Dutch origin. Note that for each profiling characteristic has a proxy nature to a greater or lesser extent. A detailed analysis of these aggregation statistics can be found in the report *Addendum Preventing prejudice*.[46]

41 Article 9 GDPR.

42 Van Bekkum, M. (2025). Using sensitive data to de-bias AI systems: Article 10 (5) of the EU AI act. Computer Law & Security Review, 56, 106115.

43 Article 10 AI Act.

44 Articles 41-42 of the Dutch Law on the National Office of Statistics. See Maatwerk en microdata, Statistics Netherlands (2025).

45 Data can be processed securely by means of protocols drawn up by Statistics Netherlands. Data is analyzed via a secure environment and only aggregation statistics are shared rather than privacy-sensitive data of individual persons. Results for groups smaller than 10 people are not published.

46 Supra note 1.

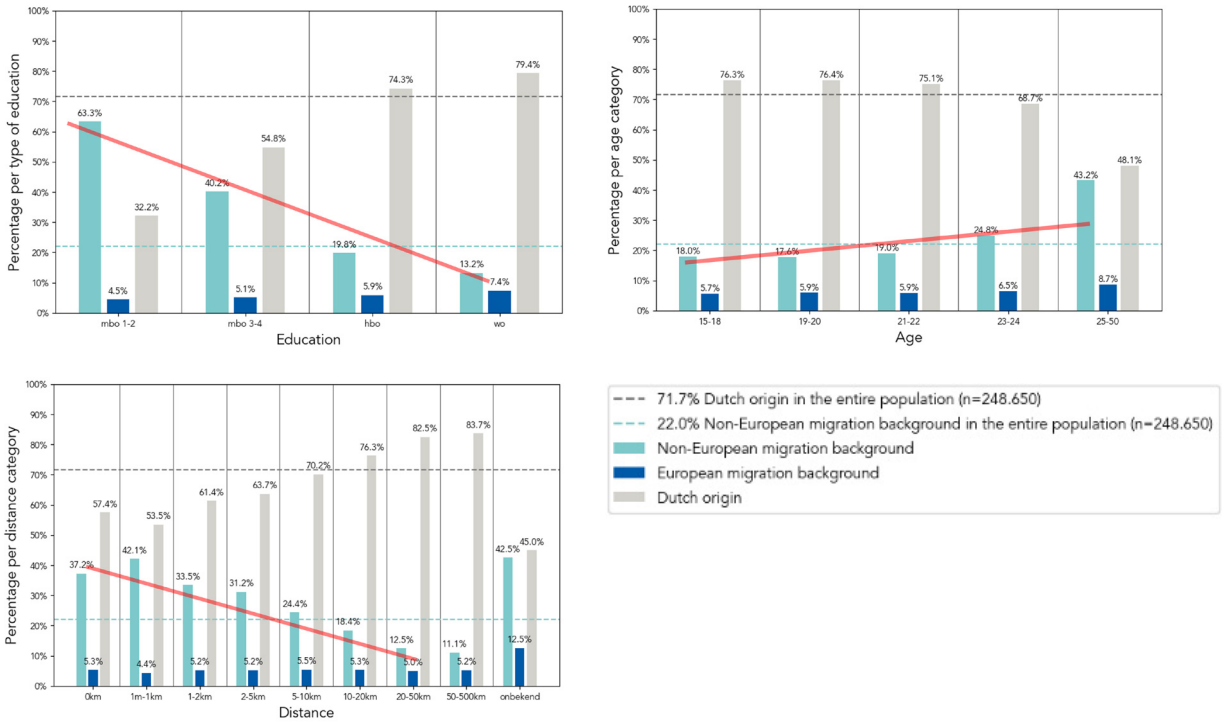

Figure 1 - Distribution of students with a (non-)European migration background and students with Dutch origin pre type of education, age and distance category in the college grant population 2014 (n=248,650).

Source: Addendum Preventing prejudice, Algorithm Audit (2024).

Now that the proxy nature of the profiling characteristics is known, the proportionality of the risk profile can be weighed. Based on the statistical hypothesis testing, we know that the profiling characteristics type of education and age are inappropriate to use, as there is no statistical support for a relationship between the profiling characteristic and the aim pursued. These characteristics do not need to be considered in this step. How the profiling characteristic distance to parent(s) should be weighed in relation to its proxy nature is a value-driven consideration that must be settled transparently and in consultation with stakeholders.[47]

If data on the protected characteristic is available at the individual level, the effect of proxy characteristics on the outcome of the profiling method (who is or is not selected for an investigation) can be monitored. In this step, (un)equal treatment through proxy characteristics can be identified most directly. This is relevant because a risk profile based on multiple criteria can ensure that the proxy nature of an individual characteristic in combination with other proxy characteristics has a different effect than if the profile consists of only one characteristic. Based on the results, a link can also be established between the expected effectiveness of the profiling method and the extent to which the selection deviates from the representative sample. This link can be established, for example, by means of adjusted selection and helps to further inform the proportionality assessment. This technique is elaborated on in more detail in 3.5 Adjusting based on characteristics.

How supervisors can gain insight into the choices made with regard to the proxy nature of profiling methods is described in Assessment protocol for supervisory authorities.

3.4 Demographic groups not available

Even if population statistics cannot be analysed, for example because internal data or Statistics Netherlands data are not available, there are still ways to investigate indirect discrimination. These methods are typically referred to as unsupervised learning or anomaly detection. Proven statistical methods can be used for this purpose, such as clustering.

In the context of scientific research, clustering was applied to the DUO case to investigate the degree of bias that could have been detected if Statistics Netherlands aggregation statistics on the migration background of students had not been available. Based on profiling characteristics (type of education, age, distance to parent(s)), students were grouped (in clusters) who were more often than average classified as 'high risk' by the profiling algorithm. The most disadvantaged cluster consists of MBO students who on average live relatively close to their parent(s).[48] Without access to students' migration background, this outcome could have served as a first signal for domain experts to further investigate a suspicion of unequal treatment in the CUB process.

[47] Algoprudence: Jurisprudence for algorithms, A. Meuwese, J. Parie, A. Voogt, Nederlands Juristenblad 10 (2024).

[48] Auditing a Dutch Public Sector Risk Profiling Algorithm Using an Unsupervised Bias Detection Tool, F. Holstege, M. Jorgensen, K. Padh, J. Parie, J. Persson, K. Prorokovic, L. Snoek, Preprint arXiv (2025).

3.5 Adjusting based on characteristics

By carefully monitoring the results of a profiling algorithm, unequal treatment can be mitigated by making adjustments.[49] Adjustment means that the producer of deployer of the algorithm adjusts the selection such that the proportion of people with a certain characteristic reaches a desired amount.[50] Adjusting offers a way to control unexpected and undesirable discrimination due to proxy characteristics. By taking the random sample as a reference point, unequal treatment can be prevented. An advantage of adjusting is that the loss in expected effectiveness of the risk profile is minimal.[51] However, this requires access to protected characteristics, such as migration background.

Adjustment is part of monitoring a profiling algorithm during use. After the allocation of risk scores, an extra step is introduced in which the selection is adjusted based on a specific characteristic. In the DUO case, this would mean that the students with a migration background with the lowest risk scores in the 'high risk' category would be replaced by the same number of students without a migration background with the highest risk scores who were not already in this category. As a result, the share of students with a migration background in the 'high risk' category decreases, while the expected effectiveness based on risk scores remains as high as possible. In this way, adjustments are made to obtain a selection that is more effective than a random sample (because a profiling algorithm is used), but that is very similar in composition to the sample (because it is adjusted accordingly). The challenge lies mainly in deciding when and to what extent adjustments should be made. This is a normative choice that relates to the justification test. In this way, adjustments contribute to weighing proportionality, because the relationship between effectiveness and unequal treatment is quantified.

4. Assessment protocol for supervisory authorities

Supervisors – both internal and external – have an important function to safeguard that profiling algorithms are used responsibly. Especially when this technology is used for risk-based selections. Building on the applicable legal frameworks for equal treatment, including those described in the Profiling Assessment Framework of the Netherlands Institute for Human Righs, the assessment protocol below supports supervisors in asking specific questions about the responsible use of profiling methods. The assessment protocol is empirical in nature and focuses on preventing undesirable indirect discrimination. The answers collected inform the objective justification test from non-discrimination law on the basis of which prohibited indirect discrimination can be established. The questions relate to different phases of the algorithm lifecycle and help to prevent citizens, consumers and organizations from being discriminated. We encourage supervisory authorities to start asking questions from the assessment protocol to public and private organisations.

[49] See Kleinberg, J, J. Ludwig, S. Mullainathan and A. Rambachan, 2018, Algorithmic Fairness, AEA Papers and Proceedings 108, pp 22 – 27 or Hekkelman, B., M.A.C. Kattenberg and B.J. Scheer, 2023, Eerlijke Algoritmes, CPB Document for an application to Dutch data and the term 'bijgestuurde selectiemethode' (adjusted selection method).

[50] It is possible to adjust selections on several characteristics, for an elaboration of this see: Hekkelman, B., M.A.C. Kattenberg and B.J. Scheer, 2024, The Costs of Affirmative Action: Evidence from a Medical School Lottery, CPB Discussion Paper.

[51] See Kleinberg, J, J. Ludwig, S. Mullainathan and A. Rambachan, 2018, Algorithmic Fairness, *AEA Papers and Proceedings* 108, pp 22 – 27.

Assessment protocol

Organisational responsibilities

1	Which algorithms and AI systems are used within the organization?

Note that algorithms and AI systems are not always recognized as such

Analysis of problem

2	Does the process in which the profiling method is used pursues a legitimate aim?[52]
3	What problem is solved with the use of the profiling method?
4	Have vulnerable groups in the population been identified? Has the adverse impact on these groups been examined? If not, why not?

Data quality

5	Was a random sample of the population drawn during the development phase of the profiling method? If not, was the used data during development analyzed to assess its representativeness of the population?
6	To what extent are labels assigned by employees reliable?
7	Are profiling features excluded from the profiling method in advance, because they are subjective, subject to change or known as proxy features for a protected characteristic? [53] [54]
8	Has the proxy nature of profiling characteristics been established by analysis of population statistics? If not, why not?
9	Are selected profiling characteristics linked to aim pursued? Are the profiling characteristics objective and verifiable?

Profiling algorithm

10	Was a random sample used to compare against the results of the profiling algorithm? If not, why not?
11	Have assumptions in the profiling method been tested for suitability based on a statistical hypothesis testing? If not, why not?
12	Is the necessity of the profiling algorithm evaluated by examining alternatives? If not, why not?
13	Has equal treatment of vulnerable groups been investigated by analysing population statistics? If not, have other analyses been carried out to identify possible unequal treatment?
14	Has the effectiveness of the profiling algorithm been determined? Are corrections, as a results of unequal treatment been investigated and are its effects on the expected effectiveness of the algorithm quantified?
15	How was the proportionality trade-off between equal treatment and effectiveness assessed?

Use

16	Which portion of the conducted investigations was selected using the profiling method? What portion of the selected group was selected randomly?
17	Is the composition of the selected group for an investigation being monitored? Has correction of the selection been considered? If not, why not?

[52] Examples: protecting public safety, preventing crime, enforcing immigration policy, combating fraud, etc.

[53] Some examples of proxy characteristics for protected attributes race or nationality are zip code, level of income, license plate, relative living abroad, low literacy. Supra note 6.

[54] Public standard profiling algorithms, Algorithm Audit (2024).

5. Conclusion

Forming a normative judgement about indirect discrimination is a complex task. Our analysis shows that non-discrimination law contains open norms that offer users too little guidance to use profiling methods responsibly in practice. Developed soft law frameworks are also of limited use for this purpose. This article describes empirical methods that can assist in determining whether indirect discrimination can be justified based on appropriateness, necessity and proportionality. The use of random sampling, hypothesis testing, population statistics and adjusting based on characteristics are central to this. Modern digital infrastructure makes it possible to analyze these data retrospectively at the population level. Based on practical experience, we have composed an assessment protocol that helps supervisory authorities to request relevant information about discrimination through profiling algorithms. This way, supervisors can improve their knowledge position without having to form a normative judgement directly. A question can easily be asked. Now that researchers have worked out examples, it is up to supervisors to scale up the practical use of these methods.

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:

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

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

Project work

Support for specific questions from public and private sector organisations regarding responsible use of AI

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