Autoriteit
Consument & Markt

Position Paper

# Oversight of algorithms

Muzenstraat 41
2511 WB Den Haag
www.acm.nl
070 722 20 00

Kansen & keuzes voor bedrijven en consumenten

# Contents

# 1 Introduction

### *Algorithms are everywhere*

Algorithms and their applications are centuries-old.[1] In the abstract, an algorithm is nothing more than a set of steps or instructions that are executed in order to achieve a certain objective. For example, a recipe for a particular dish could be considered an algorithm: on the basis of the recipe (the 'function' or the 'model'), certain actions are performed involving the ingredients (the input), which ultimately lead to a particular dish (the execution). However, algorithms are mostly known for their application in mathematics and computer science.

Algorithms play a critical role in automation. For example, think of relatively simple tasks, such as converting physical mouse movements into cursor movements on a computer screen. The use of algorithms in various fields has grown tremendously as a result of the continuing growth in computational power and the amount of available data. With the aid of algorithms (self-learning or otherwise), it is now possible to process a constant stream of data (which is largely unstructured data), and, for example, to make predictions about the weather or climate, or to spot patterns in images that are not or hardly visible to the human eye.

When people talk about algorithms these days, they often refer to software applications (which algorithms are a part of) that, in an automated manner, make predictions, take decisions, or give advice. In that particular sense, algorithms usually operate within a specific context, where the data that is used, the individuals involved, and the implementation in an organization are all relevant elements for the execution of the algorithms. That is why it would be better to refer to them as 'algorithmic applications' instead of 'algorithms' as such.

### *Oversight of algorithmic applications*

By now, people have come to realize that algorithmic applications more and more often play a role (decisive or otherwise) in activities that directly impact people, businesses, organizations, as well as society as a whole. Businesses increasingly use algorithmic applications in their operations, for example, for an optimal design of their production processes, for delivery routes, personalized offers, adjusting supply dynamically, or for dynamic pricing. This offers society many benefits. However, there are also many societal concerns about algorithmic applications that businesses use in order to offer consumers their products and services,[2] because, for example, online businesses have increasingly become better at nudging consumers in their choices and purchases. They use all sorts of techniques in order to influence online consumer behavior. At what point does persuasion turn into deception? The Netherlands Authority for Consumers and Markets (ACM) recently clarified this by setting boundaries to online persuasion in its Guidelines on the protection of the online consumer.[3]

Much of the political and societal attention is currently focused on the effects of algorithmic applications on individuals, businesses, and society as a whole. In that context, people often talk about artificial intelligence (AI). In 2019, the Dutch cabinet presented its Strategic Action Plan for Artificial Intelligence (in Dutch: Strategisch Actieplan voor Artificiële Intelligentie)[4], in which it explains its plans with regard to AI policy. One of the roadmaps in this plan describes how public interests must continue to be protected amid AI developments. In separate letters to the Dutch House of Representatives, the cabinet in 2020 responded to the parliamentary motion filed by Dutch MP Middendorp[5] and to three studies into algorithms.[6] In these

[1] One of the most well-known algorithms, the Euclidean Algorithm, dates back to Ancient Greece, and the name giver of the concept of algorithms, Al-Khwarizmi, who contributed greatly to algorithmics, lived around 800 CE in Persia.
[2] The concerns about the use of algorithmic applications by businesses are obviously broader than the concerns that directly concern ACM's duties, such as concerns about unequal treatment and discrimination.
[3] https://www.acm.nl/en/publications/guidelines-protection-online-consumer
[4] https://www.rijksoverheid.nl/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiele-intelligentie
[5] https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2020Z07093&did=2020D15106
[6] https://www.rijksoverheid.nl/documenten/kamerstukken/2020/11/20/tk-kabinetsreactie-op-drietal-onderzoeken-naar-algoritmen

letters, the cabinet talks about the risks and opportunities of AI developments in businesses, but also talks about oversight of algorithms.

Concerns about the effects of algorithmic applications on individuals and businesses are not unique to the Netherlands. Various international initiatives have been set up to develop standards for algorithmic applications. As part of its strategy with regard to data and AI[7], the European Commission, for example, presented a white paper about AI, containing a framework for excellent and trustworthy AI.[8] This strategic EU framework based on fundamental values must ensure that people are able to trust AI, and it must encourage businesses to develop AI solutions.

Merely expanding the existing standards to make them fit for the digital economy is not enough to take away the societal concerns. In an economy that is becoming more and more digitalized, regulators such as ACM must also be able to enforce the standards that they regulate if businesses use algorithms for determining their market conduct. Only then will individuals and businesses be able to have confidence that digital markets continue to work well for them.

This position paper is a starting point from which ACM wishes to develop its oversight of algorithmic applications further. It offers general guidance for investigations into violations in which algorithmic applications play a role. The paper first describes why algorithmic applications are relevant to ACM (chapter 2) and what exactly are algorithms (chapter 3). Next, it is explained how ACM is able to investigate algorithmic applications in practice (chapter 4).

[7] See https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/shaping-europe-digital-future_nl.
[8] White paper on artificial intelligence – A European approach to excellence and trust, see https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_nl.pdf.

# 2 In what situations are algorithmic applications relevant to ACM?

ACM contributes to realizing a healthy economy by ensuring that markets work well for people and businesses, now and in the future. When markets function well, businesses compete fairly with one another, and people and businesses are not harmed by unfair practices. Algorithmic applications of market participants may result in markets not functioning properly. To ACM, algorithmic applications are relevant if they play a role in activities that touch on the areas that ACM oversees. So it is about the actual application thereof in activities that have an effect on consumers and market participants. Algorithmic applications that, for example, determine prices, influence supply and demand on the energy market, or personalize the offerings to consumers are relevant to ACM. For example, think of algorithmic applications that lead to price discrimination, cartel behavior among market participants, or to a certain design of the online choice architecture for purchasing products where the consumer, against their own economic interests, takes a decision about a transaction that they would not have taken otherwise. Algorithmic applications that, for example, influence what messages are presented in what order on a company's internal site are not likely to be relevant to ACM.

The most frequently used functions for which market participants use algorithmic applications in the digital economy are:[9]

- **Search functionality**: presenting and ordering information based on certain input.
- **Aggregation**: the collection, categorization, and reordering of information from different sources. For example, think of the collection of price information of the products of competitors.
- **Observation (surveillance):** the observation of behavior and patterns in order to identify deviations. Think of network detection or fraud detection in transaction data.
- **Prediction**: predicting future behavior or scenarios.
- **Filters**: the filtering (mostly in the background) of information and data. For example, think of spam filters or filters to exclude copyrighted material (including allegedly copyrighted).
- **Recommendation systems**: recommending certain information or products mostly on the basis of data (including behavioral data) about the user, the product and/or other parameters.
- **Scoring and ranking**: the scoring or ordering of information, products, businesses and/or consumers. Think of online review scores and credit scores of consumers.
- **Information production**: the production of information. Think of automated news reporting, or automated reporting about sports events, stock markets, and share prices.
- **Communication**: automated communication with consumers and/or businesses. Think of the communication between consumers and chatbots or virtual assistants that communicate with third parties on behalf of consumers.
- **Allocation**: the automated execution of transactions, and the distribution and allocation of supply and demand. Think of the automated selling of online advertisement space (real-time bidding) or linking a customer with an available taxi.

[9] The below typology is largely taken from: Latzer, M. & Festic, N. (2019). A guideline for understanding and measuring algorithmic governance in everyday life. Internet Policy Review, 8(2).

# 3 What are algorithms and what variants exist?

In the abstract, an algorithm is nothing more than a set of steps or instructions that are executed in order to achieve a certain objective. For the meaning of the term ‘algorithm’, we follow the definition of the Netherlands Scientific Council for Government Policy (WRR): “an automated series of steps that converts input data into output data”.[10] Algorithms can be categorized in different ways.[11] This chapter limits itself to a categorization based on several common methods used by algorithms.

The way algorithms work can be distinguished into different variants. For example, an algorithm may consist of a simple decision tree of which the basic rules were determined beforehand. Those decision rules are subsequently executed in a mostly automated manner, but it is also possible that an individual executes these rules. There are also variants that, on the basis of training data, identify patterns and relationships, which are subsequently used to generate or adjust a model (which in itself is also an algorithm). The model can subsequently be used for generating a specific output (on the basis of input data), for example a prediction. This method is also referred to as ‘machine learning’. Within the category of ‘machine learning’, even more methods can be distinguished. We will discuss the most common variants below.

## 3.1 Statistical algorithms

Statistical algorithms are algorithms where the rules (or the instructions) of the algorithm are programmed, on the basis of knowledge, in advance by humans. These are predominantly instructions in the form of ‘if X then Y’ on the basis of which decisions are taken or predictions are made. For example, a software developer uses certain knowledge about credit risks for determining the rules of an algorithm that can be used to assess whether someone is eligible for a loan, and if so, at what interest rate. In this example, it can be an algorithm that, on the basis of several data points (such as income level, debts, credit score, age, loan period, collateral value, current interest rates) what the maximum loan can be and at what interest rate. Such algorithms can be characterized as ‘static’ because the algorithm’s rules, without any human interference, remain unchanged. There is no automated feedback mechanism in place on the basis of which the algorithm can ‘learn’ and adjust itself.

## 3.2 Self-learning algorithms and machine learning

The term ‘machine learning’ is mostly used for algorithms that, using training data, extract patterns and relationships on the basis of which a model is generated or adjusted. The model can subsequently be used for generating a specific output (on the basis of input data), for example a prediction. It is a data-driven learning process. With self-learning algorithms, the model continuously adjusts itself during the execution of the task on the basis of built-in feedback mechanisms. There are various methods of machine learning. Some of the most used methods are explained below.

### *Supervised learning*

With supervised learning, the model is trained using example data of which the input and the expected output are known. The expected output is also referred to as ‘ground truth’ or labeled data. During the training phase, the algorithm learns what features of the input have an effect on the output, and adjusts the model accordingly. It thus learns from historical example data. Next, the model can be applied to new data. This method is therefore often used for predicting future situations on the basis of historical data. One example of supervised learning is automated spam detection where an algorithm is trained using emails that are labeled by individuals as spam or non-spam.

[10] WRR Report no. 95, ‘Big data in een vrije en veilige samenleving’, 2016, p. 21. See also the investigative report that was carried out at the request of the Dutch Ministry of the Interior and Kingdom Relations (BZK): Hooghiemstra & Partners, ‘Toezicht op gebruik van algoritmen door de overheid’, 2019.

[11] For an extensive categorization of algorithms (function/objective, method, input data, interpretability, and nature of the developer), see: Bundeskartellamt, Autorité de la concurrence, ‘Algorithms and Competition’, November 2019, p.4-11.

### *Unsupervised learning*

Unsupervised learning does not involve the use of example data where the input has already been classified, labeled or categorized. This means no direction takes place using examples. The algorithm will thus have to systematize the input data by itself, for example, by clustering the data on the basis of shared or similar features. This method is used, for example, for making recommendations of certain products to consumers.

### *Semi-supervised learning*

This variant is a combination of supervised and unsupervised learning. Both labeled data (input data with the thereto-related output data) and non-labeled data are used for training the algorithm. This is a variant that is used regularly in practice because people often have large data sets that, for the most part, have not been labeled. Having data labeled by humans is a time-consuming process, and can be quite costly. With this model, the model is trained by using the labeled data first, for example, a model for speech recognition will first be trained on the basis of spoken words the meaning of which is already known (labeled data). The model is subsequently trained using non-labeled speech data.

### *Reinforcement learning*

In reinforcement learning, the algorithm is trained using 'trial and error'. Actions are rewarded or punished depending on whether steps in the right direction are made. This is an iterative learning process where the algorithm learns by maximizing the reward and minimizing the punishment. The main difference with supervised and unsupervised learning is that, in this method, the algorithm is not trained using training data (labeled or otherwise). This was used, for example, for training the algorithm of the computer program 'AlphaGo,' which made it possible to beat the best human players of the board game Go.[12] One major advantage of this method is that it does not rely on historical data. After all, high-quality historical data that has been carefully labeled is often hard to come by.

### *Neural Networks and Deep Learning*

A separate category of algorithms, where developments are following each other in rapid succession, are the algorithms that involve 'deep learning'. In this variant of machine learning, people use the structure of neural networks. This method is inspired by the way the human brain works. A neural network consists of a network of neurons, just like the brain. In a neural network, there is an input layer of neurons, one or more 'hidden' layers of neurons, and an output layer of neurons. The neurons in one layer can transmit a signal (output data) to the next layer (input data), which, in turn, can transmit another signal to the next layer, and so forth. The statistical output of one layer forms the input for the next layer. Whether the output of the previous layer is actually used in the next layer depends in part on the threshold values that are used.

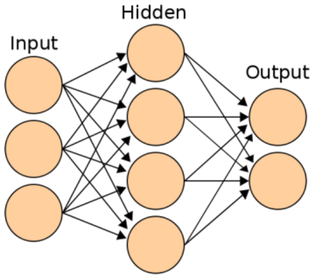

**Figure 1 – Schematic diagram of a neural network**

The term 'deep learning' is used for neural networks that contain multiple hidden layers. This method is particularly useful for pattern recognition in unstructured data. For example, this method is used in applications such as speech recognition and image recognition. One key characteristic of this method is that the algorithm's functioning and behavior are less transparent because of its complexity.

[12] https://deepmind.com/research/case-studies/alphago-the-story-so-far.

# 4 How can ACM investigate algorithmic applications?

## 4.1 Investigative powers

ACM is able to launch investigations into the use of algorithmic applications by market participants in order to gather information or evidence regarding a possible violation. The scope of such an investigation will depend on the elements that need to be proven and on the information that such an investigation may yield.

ACM is able to use the following powers for launching investigations into algorithmic applications:[13]

- Entering premises (dawn raids) (Section 5:15 Dutch General Administrative Law Act (Awb))
- Demanding information (Section 5:16 Awb)
- Demanding access to business data and documentation (Section 5:17 Awb)

If ACM demands access to digital data under Section 5:17 of the Awb, it will comply with the rules laid down in the 2014 ACM Procedure for the inspection of digital data.[14] This procedure also applies to situations where ACM investigates algorithmic applications, and, in that context, demands digital data.

## 4.2 Experience with digital investigations

ACM has considerable experience with using these investigative powers for the collection of digital evidence. In that context, ACM also has experience with the analysis of algorithmic applications or the assessment of the use thereof. In the past, ACM carried out such analyses, for example, as part of investigations into possible violations of the Dutch Competition Act, sector-specific regulation, misleading practices vis-à-vis consumers, and violations of the Dutch Telecommunications Act. In those investigations, the below investigative methods were used.

### *Administrative assessment*

Operational processes are examined, on the basis of documentation, interviews, and communications, which includes an examination of the functioning of algorithmic applications, how these are created, and who was responsible for what.

### *Simulation and code analysis*

An analysis is made of the ways in which systems within a company and/or between multiple companies communicate with each other, and what data is exchanged in those processes, on the basis of simulations or source code analyses. In simulations, ACM looks at the behavior of algorithmic applications by running them in an isolated and safe (virtual) environment. In source code analyses, ACM analyzes, using the source code, the functioning of algorithmic applications that cannot be simulated.

## 4.3 Investigations into the role, behavior, and functioning of algorithmic applications

In investigations into algorithmic applications, three key research questions can be distinguished:

1. What is the role of an algorithm in the activity under investigation, and what procedures have been followed in that context (procedural transparency)?
2. What is the behavior of the algorithm (explainability)?
3. What is the functioning of the algorithm (technical transparency)?

[13] Under Section 5:20 of the Awb, companies are required to cooperate with ACM investigations.

[14] https://www.acm.nl/en/publications/publication/12772/2014-ACM-Procedure-for-the-inspection-of-digital-data

In investigations into activities of businesses where algorithmic applications play a role, ACM does not need to analyze the functioning or the behavior thereof in each and every case. Other information, such as correspondence (internal and external), documentation, and statements, may produce sufficient evidence for establishing a violation. Such evidence will primarily be collected when assessing the role of an algorithmic application. However, regular investigative methods, such as conducting interviews, demanding information and businesses data, and on-site investigations may also be suitable for the collection of evidence when analyzing the behavior and functioning of an algorithmic application.

### 4.3.1 Investigations into the role of an algorithm

An investigation into algorithmic applications often starts with the question of what the role of an algorithm is in the activities that are under investigation, and what procedures have been followed in that context. With this type of investigation into the **procedural transparency** of an algorithm, ACM will get a better overall picture of the purposes for which the algorithm is used, what the underlying assumptions, objectives and interests are, what data is used in that context, what choices have been made, what risks have been identified, and how these are mitigated, and, finally, who were involved in the process, and what their specific roles and responsibilities were.[15] When answering this question, it will also become clear quickly whether it is about a relatively simple algorithmic application (for example a simple recommendation system) or whether it is about a complex application in which potentially hundreds of different algorithms (perhaps some from different market participants) and/or data streams play a part. Below are several questions that may be relevant when assessing the organizational context in which algorithmic applications are used.

#### *The algorithm's objective, assumptions, and changes*

- What objectives does the company wish to achieve with the algorithm?
- What assumptions, interests, and preferences form the basis of the design and/or the implementation of the algorithm in the organization? What choices were made in that context?
- Why was this solution chosen specifically? Were there also alternatives, and, if so, what were these, and why was not chosen for any of those alternatives?
- What data was used for the training of the algorithm? Where did this data come from? The same applies to the information that serves as input data.
- How was the algorithm tested? What were the results? Did this lead to any changes? Why so or why not?
- What changes to the algorithm have been implemented over time? Why were changes implemented? Has this been recorded?

#### *Role in operational processes and activities*

- For what operational processes is the algorithm used?
- What activities or decisions are supported by the use of the algorithm? What role does the algorithm play therein?

#### *Risk identification and risk mitigation*

- What method(s) is/are used for identifying risks?
- What risk(s) has/have been identified?
- What measures have been taken in order to mitigate these risks?

[15] The term 'procedural transparency' is taken from the draft purchasing conditions for fair and transparent algorithms that the Municipality of Amsterdam, together with Pels Rijcken and KPMG, has drawn up. See: https://www.amsterdam.nl/wonen-leefomgeving/innovatie/digitale-stad/grip-op-algoritmes/ (visited most recently on 10 April 2020).

### *Roles and responsibilities of actors*[16]

- Decision-makers: who is involved in and/or responsible for determining the objective and the assumptions (including margins of error) of the design and/or the implementation of the algorithm in the organization?
- Developers: Who are involved in the development of the algorithm, the implementation, and maintenance?
- Users: Who in the organization uses the algorithm in operational processes and activities, and what is their role? What is the division of roles between the algorithmic application and users in decisions that have an effect on consumers and market participants? To what extent do human interventions occur in these decisions?
- Inspectors: Who are involved in the standardization process with regard to the development and use of algorithmic applications, as well as compliance therewith within[17] and outside of the organization? What decisions have they taken, and why?
- External parties: Are external parties involved in the development and/or implementation of an algorithm? What is their role and what arrangements have been made?

## 4.3.2 Investigations into the functioning and behavior of an algorithm

### *Transparency*

The level of transparency of an algorithmic application is relevant to the determination of the functioning and behavior of an algorithm. This largely depends on the algorithm’s complexity. The more complexity increases, the more opaque an algorithm’s reasoning process usually becomes.[18] In the literature, there is a lot of discussion about what exactly is meant by ‘transparency’ of algorithms, but, roughly speaking, two forms of transparency can be distinguished at this point: **technical transparency** (functioning) and **explainability** (behavior).[19]

Technical transparency is about how the algorithm works. In that context, relevant information includes information about the functioning of an algorithm, the source code, the documentation (technical or otherwise), the used training data, and the applied variables, parameters, and threshold values.

Explainability is about how the algorithm behaves. In that context, relevant information includes information with which the algorithm’s output can be explained. A further distinction that can be made here is whether the explanation relates to the algorithm or the applied model in general, regardless of the input (*model-centric*), or relates to a specific input-output relationship (*subject-centric*).[20] One such example is for example information about what changes to the input data lead to a different output or decision.

### *Investigation into functioning and behavior*

In simple algorithms, behavior can mostly be explained by the functioning of the algorithm. It can, in certain situations, be useful to take a look ‘under the hood’ and analyze the source code itself. However, there are limitations to this research method. First of all, effective code reviews are dependent on the existence of good documentation. If that does not exist, code reviews quickly turn into time-consuming and costly affairs. This also applies to complex codes. In addition, code reviews of complex algorithms are less likely to be of added value, because they provide little insight into the algorithm’s behavior. Many developers use source code that is developed and shared by others (for example through platforms such as Github). In those

[16] The characterizations of ‘decision-makers, developers, and users’ have been taken from: M. Wieringa, ‘What to account for when accounting for algorithms: A systematic literature review on algorithhmic accountability’, FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, January 2020, p1-18.

[17] Within organizations, think of company lawyers, compliance officers, the privacy officer or ethics offcer.

[18] See also the guidelines on the application of algorithms by governments, published by the Dutch Ministry of Justice and Security.

[19] See, among other documents, the two reports that were commissioned by the European Parliamentary Research Service Panel for the Future of Science and Technology: R. Koene A. Clifton C. Hatada Y. Webb H. Patel M. Machado C. LaViolette J. Richardson and D. Reisman, ‘A governance framework for algorithmic accountability and transparency’, 2019; and C. Castelluccia, D. Le Métayer, ‘Understanding algorithmic decisionmaking: Opportunities and challenges’, 2019.

[20] L. Edwards, M. Veale, ‘Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For’, 16 Duke Law & Technology Review 18 (2017).

cases, the original source can be found (using a search query based on specific source code) containing information about the functioning. Besides finding out the functioning of an algorithm, code reviews can offer insight into the intentions of the developer(s). For example, it can give insight into the changes (transformations) to the input data before they are run through the algorithm.

There are also methods for getting a better overview of the behavior and underlying rationale and the relationships (causal or otherwise) between the input and output without code reviews. Various research methods are available for that, ranging from traditional statistical and econometric methods such as regression analysis to more complex methods where the algorithm's behavior is calculated using algorithms.[21]

### *Input-output analysis: historical data or self-defined data*

The functioning of an algorithm can be analyzed using, among other methods, input-output analyses. In this kind of investigation, ACM can use existing historical input and output data or using self-defined input. One benefit of self-defined input is that, in such cases, ACM is able to analyze in a more controlled manner what effect any differences in input data can have on the output. However, with self-learning algorithms, ACM should take into account that the input data itself can change the model and the eventual behavior. This can have consequences for the replicability of a specific outcome. It is therefore important to take into account that the model itself can change, and that there is an overview of this, for example, by recording the changes.

### *Input-output analysis: live environment or controlled environment*

If you use self-defined input, you could decide to test that input in a live environment, which means in the context in which the algorithm functions in practice. By demanding the thereto-related output (or access thereto), you can assess the real-life behavior of the algorithm. Another option is testing the self-defined input using a copy of the algorithm, which runs in a controlled environment (sandboxing). In that situation, ACM has more control over the algorithm's functioning. This can be relevant with self-learning algorithms in particular. One drawback of sandboxing is that the exhibited behavior may deviate, to a greater or lesser extent, from the algorithm's real-life behavior. That may have consequences for the evidentiary value of results of analyses in controlled environments. With analyses in controlled environments, it is therefore important to find out whether, and if so, what deviations may occur compared with a live environment.

### *Explainable AI*

As indicated earlier, there are also methods where algorithms are used for explaining (fully or by approximating) the behavior of algorithms. One well-known example of such a method is LIME.[22] Such solutions can be used for getter a better overview of the behavior of complex algorithms, such as algorithms that use neural networks. Other examples of technical methods providing more insight into the behavior more are Anchors[23], Sunlight[24] and TREPAN.[25]

When choosing such solutions, it is important to think about what you want to analyze, and to check whether the chosen method is fit for that purpose. For example, some methods are better suited for explaining the behavior of an algorithm with specific input, while other methods are more suited for obtaining a general idea of the behavior regardless of input. It should be noted that there is a debate about

[21] For an overview of different technical methods, see: A. Adadi, M. Berrada, 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)', IEEE Access 6 (2018).

[22] M. Tulio Ribeiro, S. Singh, C. Guestrin, '"Why Should I Trust You?": Explaining the Predictions of Any Classifier', http://arxiv,org/abs/1602.04938.

[23] M. Tulio Ribeiro, S. Singh, C. Guestrin, 'Anchors: High-Precision Model-Agnostic Explanations', AAAI 2018: 1527-15355. https://homes.cs.washington.edu/~marcotcr/aaai18.pdf.

[24] M Lecuyer e.a., 'Sunlight: Fine-grained Targeting Detection at Scale with Statistical Confidence', CCS '15: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, October 2015 p. 554–566. http://www.cs.columbia.edu/~djhsu/papers/sunlight.pdf.

[25] These examples are derived from C. Castelluccia, D. Le Métayer, 'Understanding algorithmic decisionmaking: Opportunities and challenges', 2019, p. 48-50.

the reliability of such methods ('explainable AI') to explain the behavior of algorithms accurately.[26] After all, they are merely a statistical interpretation or approximation of what the underlying model does.

### *Relevant information for research into the functioning and behavior*

When examining the functioning and/or behavior of algorithms, the following information can be relevant:

- The relevant documentation (technical or otherwise) about the functioning and underlying assumptions of the algorithm, including internal user manuals;
- The source code, including older and alternative versions (that are available through the used version-control system such as Git);
- The used input data and/or training data;
- The used variables, parameters, and threshold values;
- Log files and other diagnostic information such as test and debugging reports, and documentation about implemented changes; and
- Scrum boards, communication within collaborative environments that developers use (such as Slack, Mattermost and Github) and other communication channels (for example email and chat apps).

## 4.4 Challenges in investigations into algorithms

Below, several challenges are discussed that ACM could be dealing with in concrete investigations into algorithmic applications. The businesses that use algorithmic applications, too, might have to deal with these challenges when answering to their internal and external stakeholders.

### 4.4.1 Transience

Self-learning algorithms and/or the data that play a role in the training and functioning thereof may be transient. Self-learning algorithms adapt themselves, and can thus behave differently from one moment to the next. However, the context in which they operate, too, is transient, both in terms of input for the functioning, as well as the behavior that is linked to the outcome of the algorithm. The volume of data can be so enormous that storage (long-term or otherwise) may be difficult or very costly. In addition, it may also concern, for example, personal data that are aggregated but that are not (or cannot) be kept for privacy reasons. If ACM investigates a structural violation that was committed or was started in the past, this transient nature of algorithms (self-learning or otherwise) and/or the thereto-related data can be a challenge.

### 4.4.2 External parties / chain problem

Algorithmic applications do not operate in a vacuum, and, within an organization, they are part of an IT environment with links to systems (or subsystems) and datasets that could belong to or fall under the responsibility of other parts of the organization or external parties. Particularly when applications that use complex algorithms are concerned, companies often use third-party algorithmic applications. In those situations, such applications can run locally (as regular software) or be accessed remotely (cloud applications).

The use of algorithmic applications of external parties does not mean that ACM is unable to investigate them. Third parties, too, are required to cooperate with any request of ACM for information or access to business data and documentation, taking into consideration that any such request must be proportional. If multiple third parties are involved in the algorithmic applications, it may complicate ACM's investigation. However, the basic principle remains that the company that takes out services and products from external parties for the purpose of its operational activities continues to be responsible for its own operational activities.

[26] C. Rudin, 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead', Nature Machine Intelligence 1, 206–215 (2019).

### 4.4.3 Cross-border aspects

In some of its investigations, ACM is often faced with data that is stored outside the Netherlands with an external party or where the company under investigation uses services that are offered by an external party abroad. ACM takes the position that, in investigations, data can be copied if one of the following three situations (or a combination thereof) applies. The company under investigation:

1. owns the data;
2. manages the data; and/or
3. uses the data.

Real-life experience shows that this works well for data, depending on the situation. This assumption can also be applied to investigations into algorithmic applications. If the algorithmic application can be copied in its entirety as files or as a system (including a virtual system), it can be compared to copying a collection of data.

### 4.4.4 Privacy and necessary data for investigations

It is widely known that there are algorithmic applications that involve the processing of huge volumes of personal data. For investigations into the functioning and behavior of such algorithmic applications, it may be necessary for ACM to have access to such personal data (or a part thereof). This may mean that ACM gets access to large amounts of personal data, which comes with inherent risks. In that case, ACM as data controller will have to meet the requirements laid down in the GDPR. That is why ACM must ascertain in advance whether any personal data will be processed in its investigations into algorithmic applications, and if so, what measures ACM will take in that concrete case in order to do so in a lawful manner with sufficient safeguards in place.

### 4.4.5 ACM’s IT investigation infrastructure

Algorithmic applications can be limited to a simple infrastructure, and use one or more of an undertaking’s computer systems. In that case, it is likely that such an environment can be copied and taken by ACM, and that ACM can activate and investigate it in a protected and isolated manner. The use of third-party cloud services by undertakings, including the use of third-party algorithmic applications, may result in ACM only being able to investigate an algorithmic application if ACM also uses the same third-party cloud services. In that case, ACM would have to deviate from the regular procedure in which digital investigations are conducted in isolated environments.