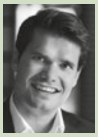
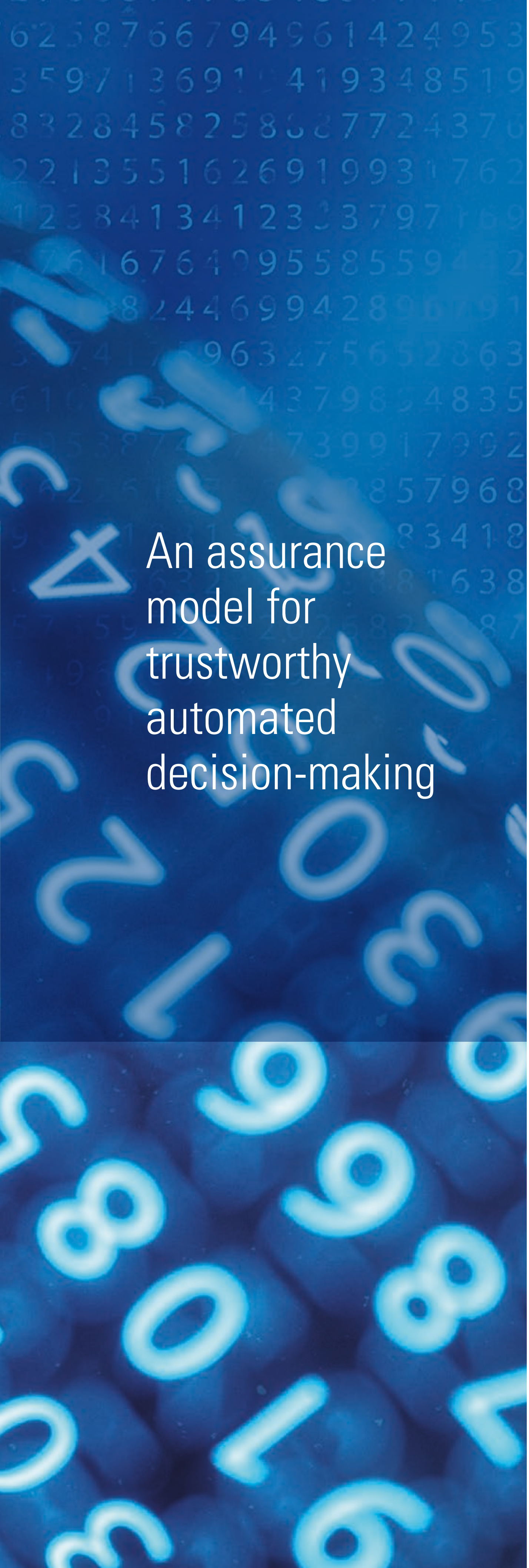
In algorithms we trust

Can we trust the analysis and decision-making processes that takes place under the hood of systems that guide us? Early signs indicate a growing societal agitation about algorithms although the common norms and values we attach to them are far from crystal clear at this point in time. We recognize strong similarities with the events that resulted in the rise of the financial audit profession well over a hundred years ago. Back then there was significant distrust of the general public in annual reports. We propose, in analogy with financial audit, to develop an assurance model for the governance of algorithms as the foundation of societal trust.

Prof. Dr. Sander Klous
is a partner at KPMG and a professor at the University of Amsterdam.
klous.sander@kpmg.nl



Frank van Praat MA MSc RE
is a manager at KPMG.
vanpraat.frank@kpmg.nl



An assurance model for trustworthy automated decision-making

INTRODUCTION

What is the common denominator of investment decisions, elevator buttons, medical diagnoses, news feeds and self-scanning cash registers? The answer: these are all examples of where decisions are increasingly fuelled by algorithms. In the past we used to be afraid of a scenario where Big Brother was *watching* us. The reality turns out to be different: Big Brother is *guiding* us in almost everything we do. That leads us to a new challenge: how do we ensure that this automated guidance of our lives is done properly? In other words, if we enter an age of *governance by algorithms*, we need to think about the *governance of algorithms* as well.

Can we trust the analysis and decision-making processes that take place under the hood of the systems that guide us? Early signs indicate a growing societal agitation about algorithms although the common norms and values we attach to them are far from crystal clear at this point in time. We recognize strong similarities with the events that resulted in the rise of the financial audit profession well over a hundred years ago. Back then there was significant distrust of the general public in annual reports. We propose, in analogy with financial audit, to develop an assurance model for the governance of algorithms as the foundation of societal trust.

Although there are significant differences, we can surely learn from the three lines of defense model that was developed for controlling financial risks over the last decades. In this article we discuss how a similar model can be developed for algorithm assurance. Our main conclusion is that the most difficult element is the creation of connecting tissue between the functional domains in the first line of defense and the risk domains in the second.

> If we enter an age of governance **by** algorithms, we need to think about the governance **of** algorithms as well

GOVERNANCE BY ALGORITHMS

We have become addicted to algorithms in almost everything we do. These algorithms typically have a positive and a negative side. The positive side is that they help us make better decisions or make our lives more convenient. One of the main negative sides is the risk that algorithms may guide us in an inappropriate way. A growing group of concerned citizens voices concerns about this adverse impact on our lives. This makes sense, as algorithms quickly gain significance in a society where the amount of data grows exponentially and many algorithms are far from transparent about how decisions are made "under the hood".

Algorithms for instance dictate our credit scores (in China, credit scores are even based on social behavior these days); In some cases, even jail sentences are partially based on algorithmic assessments; and medical professionals diagnose patients based on data, with software that contains algorithms to come up with personal advice.

The old motto of George Orwell states that Big Brother is watching you. In fact, reality has developed beyond that. As a consequence of the increasing impact of digital technology on our human actions, it's safe to say that nowadays Big Brother is guiding you. One can only hope that this guidance is based on the correct models and is in accordance with our values and needs. Technology such as algorithms mediates ([Verb15]): we can't just see this technology as "neutral stuff" as it shapes the way we interact with each other. As a consequence it has effect on our personal acts (micro) and on the functioning of society as a whole (macro).

We expect platforms to behave in accordance with societal values that are continuously changing

What is needed to create an algorithm that we can trust? Let's look at a simple example: a navigation system. As a user, you expect that such a system will lead you from A to B in the best possible way. That requires at least three things: 1) the quality of the (card) data must be valid; 2) the route must be calculated in an effective and reliable manner under varying circumstances; and 3) the results must serve the best interests of the user. For example, the algorithm should not have a preference for routes along particular commercial outlets or gas stations (unless asked for).

This sounds simple but in reality is a lot more complicated. One of the reasons is that historically, the pace of technological developments is faster than regulators can cope with. Also the awareness and implications on societal norms and values seems to consistently lag behind.

A beautiful example is found in the case of airline reservations platform SABRE. In the 1970s travel agents could complete a near-instantaneous booking for most airlines via dedicated SABRE terminals. It was a breakthrough concept compared to manual reservations and was very successful. It was also heavily criticized as SABRE could favor American Airlines. Looking back, it was remarkable that American Airlines didn't even deny their manipulative efforts. The president of American, Robert L. Crandall, boldly declared in a senate hearing that biasing SABRE's search results to the advantage of his own company was in fact his primary aim. In his words: It was the raison d'etre of having such a platform.

Decades later, this argument sounds silly, mostly because societal norms and values have changed. Should Mark Zuckerberg have been so bold in his recent senate hearings about Facebook practices it would have been hopelessly naive and would probably have a disastrous effect for his company. We now expect platforms to behave in accordance with societal values. However, these values are far from crystal clear and are continuously adjusted to cope with fast technological advances. One important aspect in the discussions is the "black box effect" of algorithms and the lack of assurance around the operating intricacies. Policymakers and politicians have only started to discuss how algorithms govern our lives and to develop an accompanying vision on what that means for the way algorithms should be regulated. The European Parliament formulated a resolution in the spring of 2017 calling on the European Commission to take the lead. The emergence of algorithms, according to the resolution, has beautiful and less beautiful aspects. Learning machines have "immense economic and innovative benefits for society" but at the same time they offer new challenges.

One of the key elements in the resolution is the explainability of algorithms. This explainability is already manifest in the European GDPR legislation that gives citizens a "right to get an explanation" in case of algorithmic decision-making. An example: it entitles them to understand why they were rejected for a bank loan when the decision was based on an algorithm.

GOVERNANCE OF ALGORITHMS

When society is governed *by* algorithms, the question pops up how to organize the governance *of* algorithms. This subject is addressed in scientific initiatives such as "Verantwoorde Waardecreatie" (Responsible Value Creation) with Big Data (VWData) in the Netherlands. The goal of this initiative is to develop instruments and an architecture for fair, reliable and trustworthy use of Big Data in order to ensure value creation for business, society and science.

The audit profession can play a key role in contributing to this goal by providing a new type of assurance around the use of algorithms. However, this is a relatively new domain. The profession is currently in need of a model as a foundation for this assurance. The goal would be to give an integrated opinion on a set of characteristics of how organizations develop and deploy algorithms. This opinion would then enable organizations to demonstrate publicly that they properly govern algorithm results and helps them to build and maintain public trust.

In a number of perspectives, this new challenge has analogies with the model of the audit of financial statements and the accompanying control frameworks that has been around for decades. Let's explore the similarities and see how we can learn from these.

Assurance on behalf of society

The first observation is that both financial audits and algorithm assurance are carried out on behalf of a broad societal need. In the case of financial audits, the objective is to make sure that users of financial information can trust this information and use it for their decision-making. In practice, there is a wide array of decisions. One example is an investor who uses the audited information to make an informed judgment. Another is a job seeker who uses a financial report to gain some background information before a job interview. It is evident that the first user has higher demands on accuracy than the second. There's a variety of public expectations, based on what's at stake and its dependency on the information.

To deal with this, auditors apply the concept of so-called materiality (or relevance). The audit profession has learned how to draw the line between what is big enough to matter or small enough to be immaterial. It depends on factors such as the size of the organization's revenues, its position in society, and the type of the business it is in. Ultimately, it's a matter of professional judgment which misstatements (or omissions) could affect the decision-making of the users. In addition, materiality is influenced by legislative and regulatory requirements and public expectations. Materiality – defined in the planning phase of the audit – then defines the level and type of testing to be done.

There is a striking similarity between the requirements for financial information and the deployment of algorithms. Algorithms also serve a wide variety of user needs. Some of them may be very critical and have profound impact – such as an algorithm that advises jail sentences based on data – while others are less impactful – such as the Netflix recommendation algorithm that guides you seamlessly into your next binge watch. These differences in potential impact should be a primary driver in determining the materiality of algorithm audits.

Part of assessing the impact of an algorithm is the number of users that may be affected by it. The algorithms determining the newsfeed of Facebook have impact on how hundreds of million users view their world. Even though fake news might not be impactful if it misinforms a single individual, the scale of it is defining for the materiality in giving assurance on the algorithm.

The discussion about materiality might actually provide interesting guidance on another heated debate around algorithms: the norms and values that society expects algorithms to meet. This discussion includes questions like: when is an algorithm good enough? Or, who is responsible if an algorithm fails? Legislation, rules and regulation typically lag behind these issues, caused by technological advances. The burden will probably be on the various courts to set boundaries by means of case law.

Three lines of defense

The three lines of defense model has been around for many years as a design philosophy of how organizations can be in control of (financial) risks. The basic assumption is that senior management needs to rely on the effectiveness of an organization's risk management that is carried out by functions in different lines in the organization. The model defines the relationship between these functions and describes the division of responsibilities.

The first line of defense owns and manages risks. This is often a business unit responsible for realizing operational and strategic goals. The responsibilities of this line include providing sufficiently reliable information, which means they need to have monitoring and controls in place that makes sure the provided information is indeed sufficiently reliable.

The second line of defense consists of functions that oversee or specialize in risk management and compliance. Professionals in this second line facilitate the first line in areas such as planning & control, financial risk, process control, information processing, etc.

The third line of defense provides assurance by monitoring how the first and second line operate in accordance with the system of controls. The third line also reveals inconsistencies or imperfections in this system and reports to senior management.

This model has become the standard for the majority of large organizations. External auditors base their opinion on how this model works – in some interpretations of the model they are part of the third line. They carry out tests to see the workings of controls and observe how the model works. Their opinion is based on monitoring the processes and controls.

The second line plays a pivotal role in the model. By functioning properly, it provides the first line with advice for the business to perform better (and leaves the primary responsibility for monitoring and control in the business). In addition, it gives the third line the possibility to be able to rely on their work, so that this third line can obtain assurance without building an excessive parallel "control tower".

Algorithm assurance is about enabling new opportunities so that a trustworthy outcome is intrinsic to the process

What happens if we apply this model to the development and deployment of algorithms? The first line consists of data-analysts and programmers who work on developing, improving and deploying algorithmic applications, their responsibility is to build high quality models and software coding and the use of reliable data. The second line consists of professionals who are responsible for risk management on topics such as security and privacy. The third line consists of auditors whose challenge is to give an opinion on the algorithms based on the control framework that is implemented by the first and second line.

This may sound logical and simple. However, the hard part is to build a framework that ties the first and the second line together in a logical way, as the primary focus of these lines is very different.

The first line is focused on building the best algorithm for a specific need. Professionals are responsible for quality and have a variety of instruments to guarantee this. They organize themselves around functional topics and are responsible for elements such as quality control, architecture, data management and testing in their projects.

The second line aims to be in control of risks, partly based on compliance. In the case of algorithms there are a number of domains such as security, privacy and ethics. Professionals in this line monitor how the first line takes its responsibility to control these risks. Currently, a number of building blocks of algorithmic governance are already in place such as security audits, information security standards and ISAE 3402 statements.

The challenge is to build a framework of controls that seamlessly connects these to the daily work of the first line. This would allow governance of algorithms to be built into the regular processes instead of being added as an extra layer of bureaucracy. Such an additional layer of bureaucracy would be far more destructive for the fast-paced world of algorithm assurance than for the yearly cycle of financial audit.

Nevertheless, in the (financial) risk domain there is already a lot of experience with optimizing the efficiency and effectiveness of the second line. One of the challenges in the financial sector was to "merge" the legal requirements from various different compliance programs into one framework, building a robust but efficient model: "One control for different purposes". This is very similar to one of the challenges in algorithm assurance, where controls on various layers of the enterprise architecture, like infrastructural controls and data management controls have to be merged into a single framework as well to be able to provide efficient and effective assurance on an algorithm.

There is a lot at stake to get this right. Algorithm assurance is not about avoiding the risks that come with building algorithms or about creating rigid structures that stifle innovation and flexibility. It's about enabling new opportunities in such a way that a trustworthy outcome is an intrinsic part of the process. In fact, it would probably be better not to speak of three lines of defense but rather of three lines of responsibilities.

Changing the audit profession

This means that the audit profession has to take their responsibility and lead the way in defining principles, guidelines and frameworks that contribute to the need for greater oversight on algorithms. Auditors need to work closely together with developers and data scientists to set the standards for oversight; the complexity of the topic requires a combined effort in an open ecosystem instead of individual auditors trying to reinvent the wheel.

A certain level of understanding is needed for auditors to play this leading role. Some decades ago, there was intense debate on the question if external auditors could carry out an audit around the computer. Nowadays every auditor has a basic understanding of how systems work and how these contribute to a control framework. The same is true for algorithms. We cannot and shouldn't expect from auditors that they can grasp the logic of what's under the hood – especially when it comes to advanced examples involving e.g. self-learning algorithms. But we can expect them to grasp how the development and deployment of algorithms can be monitored and controlled. In other words: you don't need to understand the electronic circuits in your household appliances to understand the need for a circuit breaker.

Successful standards for oversight will guide organizations to gain greater control over their algorithms, and give auditors the opportunity to extend their impact beyond financial statements. They will not be on the seat of the analyst or the programmer and are not going to assess or judge what takes place under the hood. Instead, auditors will represent society in assessing if an organization has taken its responsibility when it uses an algorithm that impacts our daily lives.

Reference

[Verb15] P. P. Verbeek, *Beyond interaction: a short introduction to mediation theory*, Interactions 22, 3, April 2015, pp. 26-31.

About the authors

Prof. Dr. S. Klous is a partner at KPMG in the Netherlands where he leads the data and analytics practice and is professor in Big Data Ecosystems at the University of Amsterdam. He holds a PhD in High Energy Physics and worked on a number of projects for CERN, the world's largest physics institute in Geneva. He contributed to the research of the ATLAS experiment that resulted in the discovery of the Higgs Boson (Nobel prize 2013). His best-selling book *We are Big Data* about the future of our information society was runner-up in the management book of the year election of 2015 and is published in Dutch and English. His new book *Trust in a Smart Society* appeared in 2017 and was leading the management book charts in the Netherlands after its introduction. It has been nominated for book of the year by the order of organization advisors and will appear in English in October 2018.

F. van Praat MA MSc RE is a manager at the IT Assurance & Advisory group of KPMG and currently leads the development of KPMG's Artificial Intelligence (AI) Assurance services. His expertise lies in IT Risk Management and control, Data & Analytics, and the human side of AI. He has worked in both national and international environments where he has developed and implemented systems of Internal Control over IT and worked on tooling to enable data-driven audits. Besides his work for KPMG, Frank is also involved in the Post-Master IT-Audit, Compliance and Advisory at the Vrije Universiteit in Amsterdam.