Keeping Control on Deep Learning Image Recognition Algorithms

Tjitske Jager and Eric Westhoek

1 Introduction

Can computers become smarter and faster than humans? This question is hard to answer. Yet, the learning capacity of systems provides rich insights into things that we as humans simply cannot see. This involves patterns and connections that have hitherto taken place outside our field of vision. The applications to provide insight into this not only make use of criteria or business rules devised by humans, but also independently search for emerging patterns and deviating observations. Not surprisingly, AI has been recognized by several governments as a key technology for the future. There is broad consensus among practitioners, scholars, and governments AI offers many and new opportunities. Algorithms for instance often support and improve the business operations and service delivery processes of organizations. In addition, algorithms also offer opportunities to make decision-making processes transparent and more controllable.

Using AI algorithms also introduces novel threats to organizations. The complexity of these algorithms (too many variables or components) and the fact that AI oftentimes entails the use of neural networks means that the processes of how the algorithm attained its results become a black box. In addition, AI algorithms and the data that has been used to train the algorithm can contain biases. Further, it is not known or predictable in advance what the algorithm learns, which can lead to undesired effects, especially with algorithms that learn themselves. Another threat relates to algorithms sourced from third-party vendors, where data and algorithms

T. Jager
3Angles Audit, Risk and Compliance, Harkema, The Netherlands

E. Westhoek (✉)
Achmea, Den Haag, The Netherlands
e-mail: westhoek@ese.eur.nl

E. Berghout et al. (eds.), *Advanced Digital Auditing*, Progress in IS,
https://doi.org/10.1007/978-3-031-11089-4_6

are often owned by the third-party vendors. Organizations need a framework to control for these risks while reaping the benefits of AI.

Like most organization insurers have also started to employ AI for their own operational processes. An important process for an insurer is to assess damage to an insured object in case of an insurance claim. As an insurer, how do you quickly identify this damage and help the customer get back on track? In this chapter we present a case study of an insurer ABC that uses image recognition via machine learning to damage to glass horticulture greenhouses. The main benefit to ABC of using image recognition is that it decreased the time to assess the damage, thereby potentially saving more crops that are grown in the greenhouse and thus reducing the claim amount. Using this case study we will present and explain a framework to control and monitor ML algorithms.

Hereafter we will first introduce some aspects of machine learning and image recognition. Then, we will discuss other related frameworks aiming to provide organizations with more control over their algorithms in Sect. 3. In Sect. 4 we present the case study that has been used to establish the framework, and that will aid in explaining how the framework is used. We briefly discuss the case study to demarcate any interesting observations in Sect. 5. The framework that is based upon this analysis is presented in Sect. 6. Because the framework seeks to aid IT-auditors in their work when auditing ML algorithms, we discuss the role of the auditor in Sect. 7. The chapter is concluded in Sect. 8.

2 Machine Learning and Image Recognition

Machine learning allows computers to learn using algorithms. Machine Learning (ML) is about creating algorithms that can learn from data. The novel developments in the field ML have sparked a revolution in which people no longer program (if this, then that) rules within programs, but in which machines themselves derive rules from data. A machine learning algorithm is able to independently extract patterns from data, build models, and make predictions about various things without pre-programmed rules.

Learning in the context of ML differs from programming rules, as a rule-based system does. In a rule-based system, strict rules must be followed by the IS that are programmed into software in advance by humans. The problem with rule-based systems is that the program needs to be instructed step by step what it is supposed to do, while considering its impediments and ensuring that it only does what it is supposed to do. This is a time-consuming and error prone activity as all possible scenarios/situations that might or might not occur in the future must be taken into consideration. In theory, ML has the potential to relate to the intelligence level of a human being, because it is possible to let the system think like a human being, so that the system itself proposes a solution for the established situation. Mimicking this intelligence can be achieved by training the system. ML algorithms can be used to

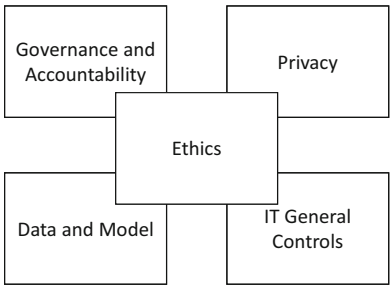

Fig. 1 Five perspectives on algorithm controls. (Source: Netherlands Court of Audit, 2021)

recognize things on an image. In this context, recognize means that the algorithm can classify whether something is on the image or not.

A simple example of such a training exercise is for instance providing an ML algorithm several pictures of Chihuahuas and muffins which can be presented to a computer (input), telling which picture is what (output). If the computer gets enough pictures, it learns to make connections between the different pictures and the computer is able to tell if there is a Chihuahua or muffin in the picture. So, there has been no person who has told the algorithm what the rules are for recognizing a Chihuahua or a muffin. However, humans are required to tell once what the correct output should be, so that the algorithm can make the connections itself between the input and output. This technique has developed enormously in recent years.

3 Related Frameworks

Despite the great social attention for ML algorithms, hitherto there are little concrete instruments to test or analyze algorithms, which is why the testing framework presented in this chapter has been developed. The assessment framework has been established based on existing guidelines and frameworks presented in other works. One of the prime foundational sources used to create our framework, is the framework presented by the Netherlands Court of Audit, (2021). This framework encompasses five perspectives (depicted in Fig. 1), where ethics is not considered separate but integrated in the other four perspectives. This is visually shown in the figure below and will be briefly explained hereafter:

3.1 Steering and Accountability

The "management and accountability" perspective concern the recording of various aspects related to governance: the assignment of roles and responsibilities, gathering of expertise, lifecycle management of the algorithm, risk assessments when using algorithms, and agreements with external parties about, for example, liability.

COBIT (Control Objectives for Information and related Technology) was used to design the assessment of these elements.

3.2 Data and Model

In this perspective, the aspects that deal with the quality of the data and the development, use and maintenance of the model underlying the algorithm are discussed. Whereby possible prejudices (based on the ethical perspective) in the data, data minimization and/or the output of the model are also recognized and tested. The assessment framework is based on scientific literature and machine learning practice. The focus of this perspective lies with the development of the model. Within the perspective, attention is also paid to the operation, use, and maintenance in practice of an algorithm. The researchers note that the testing framework has been made applicable for the entire spectrum of algorithms: from simple decision models to machine learning models. This can lead to a part of the assessment framework not being applicable to a specific algorithm.

3.3 Privacy

This perspective addresses the requirements that the GDPR poses and relevant considerations regarding the processing of personal data, in particular personal data. Legal requirements for an algorithm in the context of the General Data Protection Regulation (GDPR) must be met. Therefore, the GDPR is an important source for the assessment framework.

3.4 ITGC

Traditional IT arrangements should also be in place when using algorithms. Examples of such arrangements are the management of access rights, continuity of the algorithm, and change management. This concerns the embedding in the application and the underlying components that are relevant for the functioning of the algorithm, such as the database and the operating system.

3.5 Ethics

The starting point for the elements of the ethics perspective is the ethical framework proposed by the European Union that describes several ethical principles. Ethics are

not considered a separate element in the testing of algorithms but should be interwoven in the four other perspectives that make up the testing framework. These aspects from this perspective address:

- Respect for human autonomy.
- Preventing damage.
- Fairness (a fair algorithm).
- Explainability and transparency.

Different perspectives come together in the assessment framework. Although various guidelines/testing frameworks were available for these aspects, there was nowhere available an integrated testing framework specifically aimed at an algorithm. The testing framework is a general framework in which the various elements that are important in the control of an algorithm are addressed. The testing framework serves as a practical instrument for the auditor and is a means of control afterwards. Of course, this framework can also be of great value and input at the front end for the quality requirements surrounding the creation and use of algorithms, at the front end of the process. The assessment framework addresses the following aspects:

- Management & accountability
- Model & Data
- Privacy
- ITGC
- Ethics

The assessment framework is generic in nature, which has advantages and disadvantages. The framework provides a good solid foundation to be aware of the risks associated with an algorithm. Prior to the application of this testing framework, general questions were formulated in order to obtain a general picture and the context of the algorithm. The context in practice must guide the interpretation of the assessment framework in practice. Organizations must be aware of all risks that may arise and determine for themselves which aspects apply in this context. This can also mean that other risks can be identified from the specific situation. It is therefore not as simple as finishing the frame and that there is then a controlled algorithm.

The assessment framework first defines which risks are related to the various perspectives. Tied to these risks several safeguards and measures are proposed to control these risks. One element of "People" or "Culture" is not pointed out as a separate aspect in the assessment framework. The culture aspect is less prominently discussed in the assessment framework. However, literature suggests that this is an important aspect not to be overlooked. Ultimately the people within an organization will implement and work with the algorithm and that is why it is important to involve them early in the development so that no resistance to the use of the algorithm might emerge. The framework partially addresses this need by suggesting that multidisciplinary teams should be set up to involve a diversity of people from the organization. As indicated, the testing framework functions as a retrospective check on the algorithm and is not so much focused on the development phase. However,

the assessment framework can serve as input there. It is precisely in this phase that it is important to address these risks.

Outsourcing is not specifically mentioned separately as an important aspect but is briefly mentioned under the perspective of management and accountability and does not appear explicitly in the other perspectives. However, the outsourced processes should be assessed as they might lead to an increased risk. The fact that the part of the process has been outsourced does not mean that you are not responsible as an organization, on the contrary. It is therefore important to recognize this aspect, to estimate the risks and to include them in the research. We note that the nature of the critical questions will not change if the process is internally organized or outsourced.

The Netherlands Court of Audit treats privacy as a separate perspective. The question is whether privacy is an aspect that must be considered when controlling an algorithm. In the context of this research, this aspect is less relevant. The privacy aspect is covered by the data that is used as input for the algorithm, but also access to this data, etc. This is where the risks surrounding privacy come back. If only reliable operation of an algorithm is considered, the privacy aspect is irrelevant. However, when considering the data as important input for the algorithm the privacy aspect is equally relevant.

4 Case Study

In this chapter it will be discussed how models/algorithms are applied in practice. In this chapter we discuss the case study Project Greenhouse. We will first explain the case and then continue to explain the control aspects of the algorithm used using this case. Based on the case in ABC, we will discuss the relevant aspects regarding IT controls in order to realize a complete testing framework for the assessment of robotics algorithms. The ABC has built up considerable knowledge in the various sub-areas of AI. Several AI initiatives have been put into practice. A good example of this is the greenhouse project that focuses on recognizing damage based on aerial photos using a machine learning image recognition component.

4.1 Motivation for the Project

The idea to use robots to inspect damages for the insurance coverage was sparked during the aftermath of a major hailstorm that caused severe damage to greenhouses in two provinces. A helicopter was employed to make an estimation of the damage to the greenhouses. The helicopter flight yielded several aerial photos that provided a basis for the assessment of the damage and provided ample information on how to repair it. The speed of the assessment is important in this context because if the greenhouses remain damaged for too long the ABCs grown in it will be destroyed. A swift assessment of the damage enables countermeasures that prevents further damage. The IT department of ABC was directly involved in the process and

mapped the photos made from the helicopter to coordinates on a map. This enabled other staff that assessed the damage to directly link the helicopter photos to the reported damage. Not only does this process accelerate the assessment of the damage, but by doing so also allow the firm to inform their clients faster about the extent of the damage.

4.2 Image Recognition Greenhouse Damage

The greenhouse project has started at ABC. The aim of the project is to use image recognition to determine the damage to insured greenhouses within 24 h, so that experts have all the information about the insured in the affected area the day after a storm or hailstorm. Within 24 h, ABC wants to know the extent of the damage, and which insured objects are present in the area. For example, a loss adjuster can estimate based on the information whether the ABCs in the greenhouse can still be saved and where repair work must first take place.

After a disaster, an estimate is made of the damage to greenhouses by means of image recognition. This makes it possible to prioritize which greenhouses should be visited first by the damage-experts. This is displayed in a dashboard for the claims adjusters. The dashboard provides practical benefits for ABC who can prevent claims by responding in a timely manner and for customers who can continue to use part of their cash. If greenhouses are damaged, the crops being grown can be lost if, for example, the temperature drops due to broken and damaged windows. As a result of the above case and its evaluation, the company asked itself the following question: Can this be done smarter, easier and could machine learning do something in this?

With this question in mind, a project/innovation team set to work using machine learning and an image recognition algorithm to analyze these aerial photos from an aircraft or drone. The aim is to determine the damage to insured greenhouses via image recognition within 24 h after a major storm or hailstorm. This makes it possible to quickly analyze which crops can still be saved with rapid recovery. In the long run, the amount of damage can possibly be determined based on aerial photos. What is the greatest need and where ABC can still be of added value to limit further damage.

If action is taken promptly, temporary solutions can be used to limit the damage. In order to display the results in a usable dashboard, it is necessary to link the estimated greenhouse damage to the geographical data of the insured greenhouses. It is necessary to geo-code the data of the greenhouses insured within ABC. The coordinates of the greenhouse have been added to the policy for this purpose. A dedicated dashboard for damage-experts was developed that provides all the necessary information to prioritize which greenhouses should be visited first and to act immediately. The data required to make the prioritization process possible consists of a combination of internal data about the greenhouse and the results of a machine learning process that applies image recognition. An estimate can then be made of the damage to a greenhouse.

4.3 Process

In order to get a picture of the situation after a disaster, an external party is used that supplies aerial photos of the affected area within 1 day. The photos are automatically retrieved from the database of the third party BirdsEye, with a dedicated third-party server. The photos are then treated in the database. A roster is then created that contains tiles (squares) using the photos in combinated with GPS coordinates, effectively linking the coordinates to the pictures. The photos are assessed by the algorithm, whereby each tile is assessed in order to be able to determine whether there is damage to the respective pane or not. To assess the damage the tiles are processed by an IS that encompasses different machine learning algorithms.

The first of these algorithms determines the damage and a second algorithm determines whether it is a checkerboard or corner damage (type of damage). In the case of checkerboard damage, the damage is spread over the greenhouse. When this type of damage occurs, windows are broken on several points and little can be done to save the crops of the insured. However, if there is only limited corner damage to the greenhouse, actions will be initiated to limit the damage to the crops of the insured, and to help the insured get back into operation. These outcomes are then combined with the known data of the insured. Using this combined data, a rule-based system then determines the damage compensation that the insured attains based on whether the greenhouse is classified as a "total loss" or whether it can still be saved, also taking into consideration the type of crop harvested. Are they expensive orchids, for example, or is it lettuce, in other words, is it a plant that costs a few euros or a plant that costs a few cents. The results are presented in a power BI dashboard and the damage is prioritized based on these parameters. This ends up in the dashboard that is made accessible to damage-experts.

4.4 IT Department

The IT department is organized at a central level within ABC. ABC has set up the Internet of Things (IoT) platform in collaboration with a large third-party software provider. Within this platform, a private environment in the cloud was realized where the project could be brought into operation.

The IT department focused on building infrastructure gathers gigantic amounts of photos in a few hours, linking them to the firm's data, classifying them, and then providing this information to the loss adjuster using the dashboard. Once the damage-experts had finished their job, the resulting assessment should then be provided to the management after a (major) calamity.

From day one of the project the IT department was closely involved in the project, as it was new within ABC to develop a project in Blue which is a third-party platform. An external consultant from the large third-party software provider was involved in the project to help the organization with the development of the project.

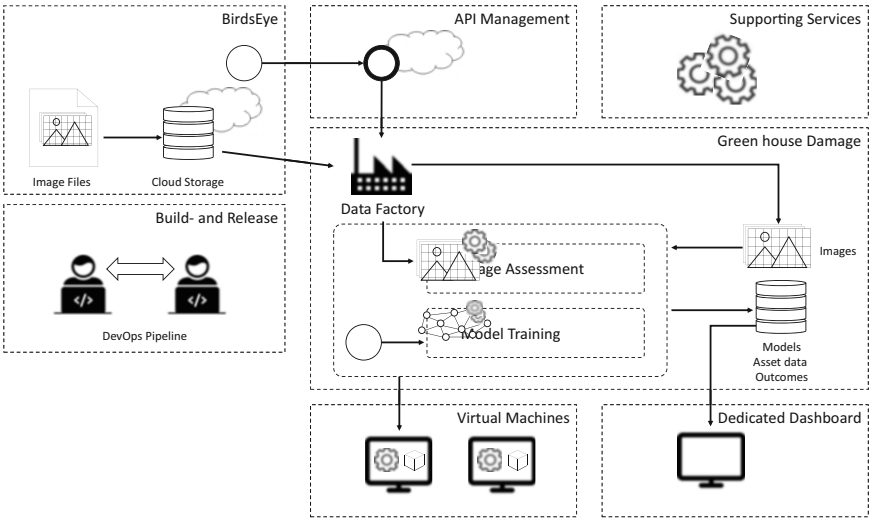

Fig. 2 Graphical depiction of the IT architecture

A development, test, acceptance, and production environment were created for the greenhouse project. Within these environments, all components were deployed. Via IDM it has been arranged who has access to these environments and who has which rights. The IT architecture developed for the project is portrayed in Fig. 2.

The environment that includes both the infrastructure and the code of the application was developed and deployed with Blue DevOps. The data factory takes care of the data transport of the data from the supplier to the storage environment that Databricks uses. The flowchart in Fig. 3 below provides insight into how the AERIAL application processes the data and provides it to the dashboard.

An external party is used to supply photos of the area affected by the calamity within 1 day. The conditions of the photos and other agreements are laid down in a Data Delivery Agreement (GLO). The supplier and recipient of the data have agreed that the photos will be delivered in accordance with a set of quality requirements. The quality of the results from the AERIAL application depends on the timely and correct delivery of greenhouse and ABC data. In the event of an emergency, it is essential that the data in the AERIAL application is up to date.

The quality of the photos is checked before they are offered as an entrance check. Some control aspects are whether the photos are not corrupted and conform to the correct projection as agreed in the GLO. If "errors" appear here, these are logged in the database whereafter the application discards them. The photos are delivered in one set, this is also recorded in the GLO. Upon receiving the set of photos, a sample is taken from that set and if there are no errors, the set of photos is approved. If the photos are removed because they have not been approved, this set of photos (which contained the error) will not be accepted. In this situation, the GLO is serves as a guideline that decide which photo does not meet the requirements and will not be

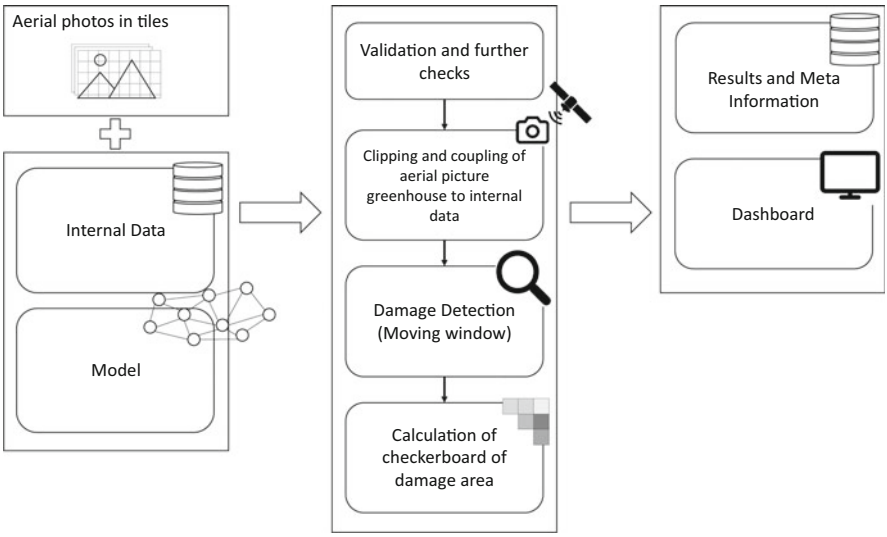

Fig. 3 Flowchart of process image recognition

accepted. The result of this check is provided as feedback to the external party. The aerial photos that are being used are placed in a database on the storage environment and sent to the Databricks environment. Data stemming from internal sources, like as customer data, data about the crop, the insured amount, the coordinates, etc. are included during this process using the Datafactory. The most recent, accurate and most up-to-date model stored in Databricks is used to classify the photos.

Access to the models is arranged via Identity Access Management (IAM) that contains IDM roles. When adjustments to the models are needed, the correct IDM role is required to perform that action. Based on meta data associated with the photos and the internal data of the greenhouses whose coordinates are known, the greenhouses are identified in the photos. Then these photos are classified with an algorithm. The results of the classification process are made available to Power BI via an IDM link. Experts have an IDM role that allows them to consult the database. At ABC there are two administrators who can also change the database, but only in terms of how data is displayed. The management roles to recreate or adjust the models have been assigned to the Data Science department. Any output of the process is thereafter made available to the claims adjuster.

4.5 Project Output

The product resulting from the process is a prioritization dashboard. The loss adjuster sees for each insured that has been affected, what percentage of his greenhouse is damaged, is it corner or checkerboard damage, or is there anything

that can be saved, can measures still be taken to save the crops together with the insured? The address is displayed as a location on the map. The estimated percentage of damage can be seen per cash/policy number. It also states the insured amount, the name of the greenhouse owner, which crops grow in it, etc. Not all data is automatically disclosed. The policy data is now manually updated every few months by someone from ABC, after which it is transferred by the data scientist to the database in Blue. This concerns advice to the loss adjuster based on a prioritization dashboard on which the loss adjuster still makes his own decision. There is no direct decision towards the insured without a human act, assessment, having taken place.

4.5.1 Training and Testing the Model

A machine learning model has been developed that is able to recognize damage on the greenhouses. This is based on classification. This first model was developed with the aim of being able to process a lot of data and train the model as simply and quickly as possible. The photos were tagged using Google Capture. The data scientist has built an application for this. A random photo of a greenhouse is taken and then zoomed in on a part, after which it is labelled by the assessor. This can click on these pieces (see opposite) based on the question is there damage "yes" or "no." A dataset was obtained from the external party to train the model. The prediction of the model was compared with the assessment of the loss adjuster. This results in a total overview, as shown in Fig. 4 below. The damage is plotted on the photo via points.

By training the model it learns to identify the greenhouses. For the training damaged and undamaged photos are provided as input each of them reviewed and tagged by a data scientist. The model learns from these examples. The tagging process is currently still performed by the data scientist. The intent for the future is that this is carried out by the loss adjuster, after which these labelled photos are presented to the model to further train it by employing supervised learning.

Actions have not yet taken in case of deviations from the expectations of the model, at least not automated. When a deviation occurs, a data scientist needs to take an action. The backlog for the further training of the model is developed to automate

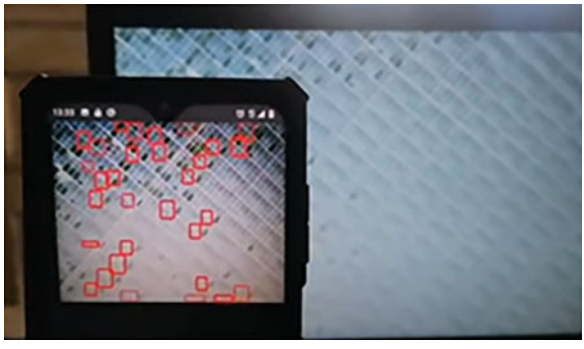

Fig. 4 Plotted damage points on a screen

this process. The "new" models are further trained on the initial model. The model can be trained with many variables and parameters. Each of these variables optimized by looking at a lot of photos that already have a label on them. Depending on the context, a model trained for a specific situation performs better than another. Therefore, the model to classify the images must be selected based on the context as it affects the accuracy of the predictions. Which model is chosen depends on the weather, for example? If there is a lot of cloud cover, the model is chosen that performs well when there is a lot of cloud. If there is also reflection from the sun, then another model is chosen that performed better under these conditions. It is important that the loss adjuster has flexibility in the choice of model. The system now chooses the model itself and projects the model on the data.

Hundred percent accurate classifications are the ideal but will never be achieved. This has to do with the circumstances, which can be different every time. A percentage of 90–95 is more plausible; this number is increased using the feedback loop in the process that allows for further refinements of the model. However, as explained this feedback loop is not yet in place, at least not automated. Currently, the loss adjuster informs the data scientist if there are doubts as to whether something went wrong, after which the data scientist adjusts this in the model, so that the model is improved.

Furthermore, currently there is still no structural recurring process to ensure that the model continues to do what it is supposed to do, that a test run is carried out once every 3 months during which it is checked whether everything still works technically. The following parts can be distinguished here:

- Assignment of ABC-crisis team to start up the IT-system
- Process the photos
- Interpretation of the photos
- Linking the photos to GEO and customer data
- Provide advice to experts

What has not been tested is whether the IS correctly links to other parts of the organization, such as reinsurance, and the back office to receive feedback from the experts. This has not yet been set up in the process. The documentation of the user stories describes the requirements of the end users and what tests need to be performed using what scenarios. All materials related to the tests for the components are included in the use cases. The management team takes care of the automatic regression test.

4.5.2 Finetuning the Model

In a neural network, labels are added that form the recognition of the damage. Based on the training set, the algorithm learns to recognize tiles as "damage" or "no damage." With a limited data set, machine learning models are less accurate, because too much value is assigned to noise. This problem is resolved by offering more than a thousand photos of greenhouses to finetune the model with this larger dataset.

The difference between the old and the follow-up model lies in the technology, namely classification or detection. In the new model, a classification technique is applied to divide a photo into many planes. This technique is potentially much less accurate than the detection technique and can never reach the level that a Yolo V3 or similar new detection models can achieve. Data scientists involved in the project have built an app to show the power of this technique. The latest model is placed in a mobile device, which can then be used to "screen" a photo of a greenhouse for damage. The entire photo is interpreted in one go and the damage, if any, is detected. The center, length, and width of the damage are also identified. Therefore, the output of the neural network is detecting these two aspects.

The follow-up model that will be used is based on the detection technique. This model has already been trained once; however, it still needs to be trained with labelling. As such it has not yet been implemented and remains a task for the loss adjuster. That means that a loss adjuster needs to keep developing the model. For this task, a new front end has been developed together with the damage-experts, to enable the loss adjuster to carry out the task himself. Taken together this also enables the damage expert to train and implement his assistant (model) himself, within the Cornerstone environment that allows for data analytics. In the future, the same flexibility will also enable to remove a model and transfer it to a drone to bring it along to a location.

4.6 Organizational Aspects of the Project

4.6.1 Involvement of the Business Unit

ABCs damage-experts themselves came up with the project proposal themselves. Therefore, there was strong support for the project from the business. During the development of the IS the support of the damage-experts was invaluable as their knowledge as experts was required to label the data and to receive their input on how it would be presented. As a person you very quickly can discern damaged greenhouses from not damaged greenhouses. When training the model, we soon found out that for machines this is far less easy. For instance, glass is transparent and confuses the model, greenhouses are not all the same. Moreover, there are different types of glass. Some greenhouses are partly covered with cloth, others have chalked windows. Not all greenhouses have crops in the greenhouse, others do, and here too a difference can be discerned, one growing orchids, the other tomatoes. Combining these factors makes it really complex to train a model that can account for all these different parameters.

4.6.2 Involvement of Other Departments

Besides the IT departments involvement for obvious reasons, other aspects within the organization also required attention. Within ABC quality and manageability of

data and algorithms was a new topic when the project commenced. At that time, there was far less know how to manage these aspects then there is now. Through this project ABC has gained significant experience.

During the project the innovation team has initiated the Project Impact Assessment (PIA) process, which results in a PIA. Compliance, legal, and security were also involved in this process to provide input from their perspectives. The departments jointly went through the process of creating the PIA. The project/innovation team, of which the business was also part, provided a description of the initiative based on a set of questions. With the help of this set of questions, each specialism then answered the set of questions from the perspective of their own discipline. The answers to the questions laid bare the possible (negative) consequences of the use of personal data for the persons and organization(s) involved where then mapped in a structured manner. In addition, the risks were identified as much as possible. In a joint session between the departments, a coordinated plan was set out to answer the outstanding questions. In addition, during the joint session answers to the questions were discussed. Finally, actions are defined. The process is summarized in Fig. 5 below.

From the assessment of the data, the data stewards were also involved, and this resulted in a BIA that deals with the aspects of information classification, availability, integrity, and confidentiality. All those involved have made separate, individual plans/given advice in their area of expertise. There is not one place of central recording, but this is distributed in the organization in the departments where specialism is invested.

4.6.3 Compliance

Compliance participated in the development of the DPIA and the risks were mapped out. Control measures are then formulated based on this. Specifically, from the perspective of Compliance, the privacy aspect has also been assessed here by means of a Data Protection Impact Assessment (DPIA). DPIA is a risk inventory prior to the processing of personal data. Whether a DPIA should be performed can be determined using the PIA test. To assess whether a DPIA test should be performed, nine criteria have been drawn up by the European privacy supervisors to assess whether the intended processing of personal data poses a high privacy risk for the persons involved. As a rule of thumb, it is prescribed that a DPIA must be performed

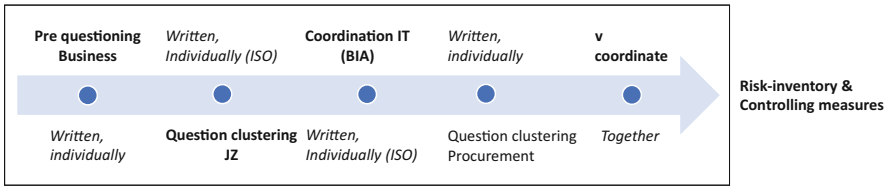

Fig. 5 Process to involve other departments

if the processing meets two or more of the nine criteria. In addition, the project needs to satisfy some criteria drawn up by the Autoriteit Persoonsgegevens (AP). Based on another assessment against these criteria, the conclusion from compliance was that: "Performing a DPIA is not necessary for the project." This is based on the fact that there is no large-scale and/or systematic processing of location data from or can be traced back to natural persons if we take photos incidentally (for example after a calamity or damage report) in execution of the insurance contract. However, it is stated as a point of attention that this should be considered in the contracts with the parties with whom we work together in this regard.

4.6.4 Security

Prior to the implementation, "Threat Modelling" was applied by the Security department, to control the security threats as much as possible. The process to develop sufficient security controls involves identifying potential threats and developing tests or procedures to detect and respond to those threats. It is important to understand how threats can affect systems. A threat model was developed for this purpose, which is based on STRIDE (Kohnfelder & Garg, 1999) threat modelling. STRIDE is a threat model created by Microsoft engineers intended to guide the discovery of threats in a system. The STRIDE model is meant to assess several types of threats to the security of an application. Table 1 shows the different types of threats that can be used to mount a cyber security attack:

Table 1 STRIDE: the different types of threats

Threat	Definition	Example
Spoofing	Impersonating something or someone else	Pretending to be any of Bill Gates, Paypal.com, or ntdll.dll
Tampering	Modifying data or code	Modifying a DLL on disk or DVD, or a packet as it traverses the network
Repudiation	Claiming to have not performed an action	"I didn't send that email," "I didn't modify that file," "I certainly didn't visit that web site, dear!"
Information Disclosure	Exposing information to someone not authorized to see it	Allowing someone to read the Windows source code; publishing a list of customers to a web site
Denial of Service	Deny or degrade service to users	Crashing Windows or a web site, sending a packet and absorbing seconds of CPU time, or routing packets into a black hole
Elevation of Privilege	Gain capabilities without proper authorization	Allowing a remote internet user to run commands is the classic example, but going from a limited user to admin is also EoP

4.7 Benefits of the Project

The "new" process offers many advantages. Without projects like these, damage-experts are less likely to have a clear picture of the damage after a disaster. After a few weeks, claims are still being received from a greenhouse that might have been "saved." The time gain is since there is faster insight into the damage, which means that prioritization can be done more quickly. This provides practical benefits for ABC, which can prevent consequential damage by reacting in a timely manner, and for customers who can continue to use part of their greenhouse. This insight also means that policyholders can be proactively approached to ensure that parts of their greenhouse remain in operation. The model is being further developed and expanded, for example:

- Automatic retrieval of policy data
- Tool to train model for experts
- Algorithm to count number of diamonds or in other words the amount of damage
- Analyze drone photos
- Unlocking photos to customers
- Automatically create a claim and inform the insured

The data scientist emphasizes the essence of the feedback loop, when this is part of the process, the model will get better and better. As a result, the expert is ultimately in charge of training models and giving feedback. Whereby everything around that is automated, so when a data scientist is superfluous.

5 Analysis of Case Study

First of all, it is good to mention that the challenge in this project was to keep the project small in order to make it manageable. The innovation manager indicated that you quickly become enthusiastic about the project and the technology that you quickly think bigger in terms of possibilities. The future wishes have been placed on the backlog. The model, the backlog, and the experience gained form the basis for further development in the coming year.

It has been a good choice to keep the project small, clear, and manageable. A project/innovation team has been started as a basis, in which the necessary disciplines have been involved, namely the data scientist who built the model and the IT department for setting up the environment within Blue. By involving the expertise in the project in this way, attention is also paid to the specifics from each specialism. An example of this is the configuration within Blue that had not been done before and where the expertise from the third-party software provider is used at the initiative of IT. By involving IT in the project in a timely manner, which in the beginning mainly focused on the layout, it shows that a good foundation has been established. As a result, no problems with regard to the technical infrastructure arose in the further

course of the project. Gradually, all relevant departments have been involved in the development and have provided their input. This concerns the compliance, legal, and security departments, but also the data experts, mainly focused on the privacy aspect.

Documentation of activities, assessments, and evaluations are recorded within the department. Within ABC this has been arranged per department, so that the recording is not fixed in one place but is spread throughout the organization. You could also opt for a multidisciplinary approach in which the input is recorded in a central project file. That's a choice. Most importantly, it is implemented and well documented and the relevant disciplines are involved.

The employees who are involved in the project in practice are also directly involved in the project. Separate training program has not been discussed here. The size of the people involved makes it possible to realize direct training on the job. It is a project that came about together and of which everyone saw the added value. This culture and motivation certainly aided in making the project successful.

It could have been better in some respects. These aspects mainly focus on the model itself. In the process, the feedback loop is not adjusted, where the damage-experts provide feedback to the algorithm on the basis of the output, but this is a condition for making the model better or for training. Continuous improvement of the model maximizes the benefits. This is also recognized by the organization and is a wish that is high on the backlog in terms of prioritization. Given the aim of the project to give direction in terms of prioritization where the loss adjuster should go, good results have been achieved here, namely:

- Better prioritization with a focus on saving ABCs.
- Customers back in business faster.
- Faster information from reinsurer.

It mainly serves the customer's interest, which indirectly also entails a financial interest. This is closely related. When is the model good enough? That depends on accuracy and practice. Hundred percent accuracy will not be achieved in practice, because it is different every time. An accuracy percentage of 90–95% should be feasible is also indicated by the data scientist. This is also related to the feedback loop, which can then be adjusted within the process, so that the model becomes more accurate. The desire to further develop the model and the changes and extensions to the backlog are a good basis for arriving at an improved model. Taking together we can identify several management aspects from the case study. These aspects are outlined in Table 2 and will be further discussed in the following section.

6 A Framework for ML Algorithms

Algorithms are getting smarter and are getting ever closer to rivalling human intelligence. The possibilities that machine learning has to offer are developing rapidly. Machine Learning is about creating algorithms that can learn from data. Machine learning allows computers to learn using algorithms. Algorithms are

Table 2 Aspects related to AI-control. Specific use of the aspects is situation- and context-dependent. The maturity level of the organization with regard to the use of these aspect plays an important role in the implementation of the controls

Controlling aspect	Aspect	Orientation
Controlling aspects aimed at:	Control	System oriented
	Processing (incl. feedback loop)	System oriented
	Contents	System oriented
	Outsourcing	System oriented
Prerequisite aspects	ITGCs	System oriented
	Governance	Data oriented
Other controlling aspects	Culture	Data oriented

increasingly influenced our decision-making and are replacing humans evermore for several tasks. An algorithm in the context of computers can be described as a set of instructions that serve to carry out a task. This concerns systems, with "simple" calculation rules based on data, to make decisions or give advice, but also to constitute to more complex learning and/or predictive systems. For rule-based algorithms it is possible to determine how they have produced a certain outcome. However, the complexity of ML algorithms has proven to be far more difficult to unravel.

Therefore, these novel developments in the field of ML also bring about additional risks and have prompted the desire within organizations to get a firmer grip on this technology. ML has a profound impact on the decision-making process within an organization and understanding that impact is key when exerting control. If the decision-making process takes place in a transparent way, firms can also take responsibility for it. Understanding how to create transparency in the decision-making process of an ML algorithm requires insight in what ML is and how algorithms are used. This insight can be harnessed to gain insight into what management aspect is relevant when controlling ML algorithms.

6.1 Fostering Trust in ML Algorithms

The research that study human–robot interaction trust in an algorithm is defined as: "the willingness of users to provide confidential information, accept the recommendations, and follow the suggestions of a robot" (Siau & Wang, 2018, p. 49). Although this definition was originally used in context of robotization, Siau and Wang suggest that the same definition could be applied to ML algorithms. The demand for trustworthy algorithms is only increasing as their influence on society can already be heard felt. In the article: "What /IF—What if auditors play a role in taming algorithms," the Dutch association for accountants (NBA) has outlined three societal trends they observe with regard to the influence of algorithms (NBA, 2020):

- Firstly, our decisions are increasingly driven based on data and the algorithms that use this data.
- Secondly, we use the technology slavishly and trust it blindly without questioning the inner workings of the algorithm.
- Thirdly, if something goes wrong, a culprit is sought as soon as possible without further investigating the underlying problem in the algorithms.

Algorithms that aid in decision-making are in fact not a novel phenomenon; however recently they have become more commonplace and are increasingly being used in a broader sense due to the emergence of Big Data applications. It is relatively easy for these algorithms to determine whether the calculation rules are "good," or whether they meet the standards set for them. These calculation rules have gradually become more complex over time because there are more (input) variables, and the underlying neural network is more complicated. This makes it not only more difficult to check the algorithm, but also to explain how the algorithm works. As a result, decision-making rules have become much more complex due to AI, with learning systems also doing their own reasoning to arrive at a decision. Some of the reasoning that the system then follows to arrive at a decision cannot (or is not easy to) make transparent. The decisions of an AI-based system are difficult if not impossible to analyze. Therefore, frameworks should not focus on testing the technology, but more about testing whether the development of that technology meets the standards to be set. We will now discuss the control areas that will serve as the basis for these standards.

6.2 Control Areas of the Algorithm

Quality of and trust in an algorithm must start at the source by setting clear, unambiguous requirements for the functioning of the algorithm and making careful choices when designing, developing, and implementing it. The creation of the algorithm precedes its use. However, this aspect will be disregarded for further elaboration on the control of the ML algorithm. The management tasks that are involved in exerting control over the algorithm are mainly focused on the aspects of control, process, and content. The preconditions that can be recognized particularly in the field of IT and governance are also important here.

6.2.1 Control

The first aspect to manage in ML algorithms is who is responsible for the algorithm and its functioning. Another aspect of ownership is the responsibility for the data from different sources that serves as input for the algorithm. This data and the associated resources are often managed by different departments within an organization. This also means that they have a different owner that is responsible for the

data provided and the associated quality aspects thereof. This raises the question who is responsible for entering this data as input into the algorithm. The responsibility for and ownership of the algorithm should be recorded. This will be further discussed in the governance section.

6.2.2 Process

An important factor in more complex forms of algorithms like machine learning algorithms is that the creation of such algorithms is fundamentally different from traditional algorithms. Traditionally, the development of a system is a static and well-organized process, and an auditor can make a statement with a certain degree of certainty about the functioning of the system using a conventional audit approach. However, developing systems with predictive algorithms (based on AI) involves a semi-autonomous and iterative process. Under human supervision or even without, an algorithm then processes a large amount of data, which autonomously creates a predictive algorithm. Statistical methods and mathematical techniques are then used to determine that the predictive algorithm does what it is intended to do.

If deviations arise, the same statistical methods and mathematical techniques are used to optimize the algorithm to the desired result. The end-goal of the ML algorithm is ultimately to predict an outcome. Therefore, a relevant question is how well the algorithm performs this task. Signals from other sources like a complaint process for the algorithm should also be gathered. A tool to recognize these signals and undertake action if necessary is recommended in such instance. Concluding, a form of output monitoring that assesses the output generated by the algorithm is relevant here. To further improve the functionality and performance it is recommended to create a feedback loop for the algorithm, so that the algorithm can continuously be evaluated and improved.

6.2.3 Contents

The dataset to train the ML algorithm is crucial to attain the desired results, as confirmed by several studies (Liebchen & Shepperd, 2008). If the data for the machine learning algorithm is inconsistent or inaccurate, the results will also be inaccurate and inconsistent. The principle of garbage in, garbage out is very much applicable in this context. A dataset must always be structured and well-balanced. By (structured) we mean that data should be annotated consistently with labels that describe the data. The more labels you add to the data, the more options there are to train models for specific solutions in the future. In addition, a qualitative dataset must be well-balanced, meaning that for each case (class) the algorithm has to identify there should be an equal number of training examples. An unbalanced set will contain a “bias” to an item and thus make inconsistent predictions. Once a certain amount of data has been labelled, the system will then recommend labels and help users label the remaining data quickly and correctly. With each iteration, the model

makes better predictions, allowing the user to work more efficiently and ensuring labels are properly assigned to the data.

No less important is the risk of whether the data contains prejudices that can lead to, for example, data discrimination. For reliable applications of AI, it is important that the data with which an application has been trained is insightful, in order to be able to find out what a suggestion or decision is based on. Therefore, the origins of the data should be clearly traceable, and it is important that the composition of the data set is reliable and representative to the predictions it is trying to make.

6.2.4 Preconditions Aspects of IT General Controls (ITGCs)

Traditional IT measures are also preconditions for algorithms. Think of the management of access rights, continuity, and change management. Specifically for the control of the algorithm, it is important to have insight into the applications that are relevant to the algorithm and to have insight into the effectiveness of the relevant application controls and the underlying ITGCs. Specifically for algorithms, one can think of the logging information, the access rights, and the password management of the algorithm. The following GITC processes are important here: Logical Access Security (LTB), Change Management (WB), Operations (OPR), and IT Security (ITSEC). Since the GITC play a role in assessing the continuity and verifiability of the algorithm, the process surrounding Business Continuity Management (BCM) is also important.

6.3 Governance

As mentioned before, governance and ownership is an interesting issue in the field of AI. From a governance perspective, the business is responsible for the primary processes it serves. However, using ML usually is part of application functionality using an application. From the first moment that organizations start working on AI they work together with the IT department to install this new application, connect existing applications and set up an infrastructure on which the algorithm can safely perform processes. This raises the question who owns the ML algorithm and who is responsible for the algorithm's actions. It is important to have insight into the tasks and responsibilities of the algorithm. In addition to the benefits of cost reduction and process improvement, robotization also raises questions about its control. What does using the ML algorithm mean for internal controls in the process, now that the separation of functions, as we know it, cannot be realized in this way, for example? The regular risks also remain relevant, such as development, management and maintenance and access security of the algorithm. There are various IT governance frameworks that can provide guidance on this aspect. The perhaps best-known framework is COBIT.

When automating tasks, the general IT risks as we know them in the regular IT audit continue to apply. The difference is that these are now focused on a different object. The IT auditor will have to pay more attention to the management of risk associated with digitization. Internal audit professionals also have a responsibility to understand the risks introduced by ML algorithms and to ensure that their company's controls are well designed and working effectively to mitigate those risks. Unlike humans, who can skip a process step or be inconsistent in the way they process a transaction, an algorithm performs the task in a standard way, without bias or any variation, ensuring a high degree of accuracy. But ML algorithms can also involve risks if the proper controls are not in place and monitored. For example, because the actions an algorithm perform are consistent, any error becomes a systemic and widespread problem in that business process and data set. Or, if there is a business process change, but the ML algorithm has not been modified to reflect that change, it may not perform or introduce inaccuracy. Another potential risk is that if someone gains unauthorized access to an algorithm or the app it is integrated into, it can be modified or used to carry out unauthorized processing. Establishing AI governance and relevant controls in advance should help mitigate risks effectively. By embedding governance, risk management, and controls into the enterprise's mobilization and implementation of AI, organizations can catch problems before they arise. Doing right from the start is much more effective and cost-effective than putting together a patchwork of policies and controls later.

6.4 Human Aspect

Every development or (technological) progress in the past has had consequences for the available jobs pool. With the arrival of AI, employees may be concerned that their jobs are now at stake. It is more likely however, that man will have to work together with machines, whereby the strengths of the people are combined with those of the machines. This is also known as augmentation or collaborative intelligence. It is therefore important to include the human aspect in the process in order to experience development as positive and thus not to see development as a threat, but as an opportunity. The research by Wilson and Daugherty (2018) shows that greatest performance gains come when humans and smart machines work together, reinforcing each other's strengths. As a result, collaborative intelligence is optimally applied.

The human aspect, the culture is a factor to take into account, as is also recognized by Serrurier Schepper and Hiddink (2019). When implementing AI applications, there should be a collaboration at all levels of the organization, involving stakeholders from different disciplines and domains in order to achieve the best result. Collaboration is a key success factor. By involving the employee in the process, giving responsibility and a task, uncertainty can be removed, and the employee also sees the opportunities that this development entails.

Governance is much broader and includes other aspects, namely in the field of compliance, legal, and the human aspect. These other aspects of compliance and legal aspects are less relevant for the control of an algorithm. These aspects play a role in the creation of an algorithm, so they are not discussed in more detail here. For the sake of completeness, I would like to point out that if part of the process surrounding the control of an algorithm is outsourced (outsourcing), the organization remains responsible for the associated risks.

7 Role Auditor

The primary responsibility for quality and trust lies with the organization that develops the algorithm. An algorithm can sometimes become very complex, and as a result no one can fully understand how it exactly works. Sometimes it is possible that the algorithms start working in such a way that even its creators no longer understand why certain decisions are made, let alone that any of the end user can. This requires the auditor to adopt a proactive attitude by looking at the risk assessment, the design and implementation of controls aimed at controlling the algorithm early in the implementation process. Specific knowledge about the chosen application and the underlying programmed code is required, but also knowledge of the process concerned. This therefore requires a joint approach from the business and IT organization, but also from the auditor. Once deployed the algorithm can then be considered as a "black box," whereby it is not always clear which data a system contains and how algorithms work. As a result, it is not always possible to understand exactly how the output is created. Yet, transparency, comprehensibility, verifiability, and explainability are essential and one should always be able to see through afterwards or find out how certain decisions came about. To ensure transparency, comprehensibility, verifiability, and explainability, it is important to be able to answer the following questions when it comes to an algorithm:

- What rules has the model learned?
- How does the model think or reason?
- Who controls the algorithm?
- Who understands the algorithm (and the code)?
- What assumptions and choices were made when training?

To be able to make a well-founded statement about the reliability of an algorithm, an auditor will not be able to suffice with the traditional approach. The assumptions and/or choices made in the development of the algorithm are just as important. For example, about the data with which the algorithm is fed and whether it is sufficient for the purpose of the algorithm, the choice of the algorithm itself, and the methods used to test and optimize the correct operation of the result.

7.1 What Requirements Must an Algorithm Meet?

When auditing we test the performance of a system against a standard. This seems logical, but what is the standard against which to test? There is a certain fault margin that we can tolerate for an ML algorithm. However, this fault tolerance is arbitrary and needs to be put in perspective of a certain context. For instance, if a human life depends on the decision of the algorithm, we would tolerate less faults as when the decision would be for administrative purposes only. As Mona de Boer (2019) in her article it is people who devise, train, and feed algorithms with data. However, the involvement of humans in the design and creation process of an AI also introduces potential risks. The image that must be avoided is that supervision (and/or an audit) of algorithms offers 100% certainty. Just as the (human) civil servant was not flawless, an algorithm will not lead us into a flawless dream world.

European privacy legislation has been tightened further with the arrival of the GDPR. Among other things, the law requires that every decision made by a computer can be explained. This also sets requirements from European privacy legislation in the field of data and algorithms, where integrity and traceability are of great importance. However, the more systems become self-learning, start to feed themselves with data and select their algorithms themselves, the closer the moment comes that their functioning can no longer be understood by humans.

As indicated earlier, the actual use of AI for business processes takes place by means of an application. Just like other applications, these AI supported business processes also need to be adequately controlled. Likewise, for IT-components IT Governance controls should be implemented to ensure the continuous and proper working of the automated processes and to safeguard these processes against unauthorized changes or that hackers procure unauthorized access to the algorithm. The framework of standards is broader than just the IT perspective and will also address the management aspects of control, process, and content surrounding the control of algorithms.

7.2 Systems-Oriented Versus Data-Oriented Auditing

An audit of a ML algorithm can be both system- and data-oriented. Several sequential steps can be followed to audit the algorithm. The audit starts with a risk-based audit approach during which the auditor analyzes the risk that the financial statements are materially misstated. The auditor then adapts the approach to the outcomes of the analyses by planning system- and data-oriented activities. Using the system-oriented procedures, the auditor determines to what extent use can be made of the measures that the organizations themselves have put in place to prevent or discover a material misstatement in the financial statements. Depending on the outcome of this first step, substantive procedures are performed to obtain sufficient certainty about the quality of the accounting. The expectation is that in a mature organization in the

IT field you should be able to audit system-oriented, on realization where you focus on the process and not on the input/output. You also include other aspects and signals from other angles in this assessment, such as management information about complaints. Are there any signs that could indicate that the algorithm is not working properly? However, there is no standardized approach to address this question as it depends strongly on the context.

In the approach to assessing the mastery of an algorithm, a measurement moment will be: Can the process approach be applied, or will the data-oriented approach have to be applied? This also depends on the maturity level of the organization and the way in which the algorithm was created. The process approach will be chosen for an organization with a solid maturity level. Before the start of the research, this consideration must first be made, which options are available and on that basis the choice for a process- or data-oriented approach can be made.

7.3 Conclusion Role of the Auditor

As indicated, the primary responsibility for quality and trust in the control of the algorithm lies with the organization that develops the algorithm. Auditors can further strengthen this trust by checking whether the algorithm is doing what it is supposed to do and by asking critical questions that are in the public interest. The assessment of the (IT) organization and associated (IT) control measures has remained unchanged in all those years: there is always a person behind the (development of) systems and the auditor therefore focuses strongly on this. In a sense, you could say that AI—with a permanent feedback loop that provides learning capacity—is an extremely fast form of change management. In essence, algorithms are mainly about applying calculation rules yourself in order to also be able to make changes in order to make decisions. However, it is not just about checking the algorithm itself with the organization and the management measures surrounding it, but also paying attention to the data used, the methods used in the development and (continuous) optimization of the algorithm. These aspects of management, process, and content should therefore also be part of the assessment framework and thus the audit approach.

8 Conclusions

How do we make an algorithm reliable? This sounds difficult and complex. Control is part of one of the tools out there to manage the adverse effects of algorithms. These adverse effects are often reflected in the media, but of course many good things are also done with the help of algorithms. What risks do we see and how can we ensure that the AI application is created in a controlled manner and works reliably. This starts with having sufficient competences to understand how this works, both when it was created and afterwards how it should be investigated. Relevant control aspects

that are presented in this chapter are the minimum aspects that can be expected in the assessment framework aimed at assessing the control of an algorithm. This concerns aspects aimed at control, process (including the feedback loop), and content, but also aimed at preconditional aspects. Summarizing from my research, the following aspects are important that must be addressed in a testing framework aimed at controlling an algorithm. Control aspects aimed at:

- Control
- Process (including feedback loop)
- Contents
- Outsourcing

Preconditions aspects:

- ITGCs
- Governance

Other management aspects aimed at:

- Culture/human aspect
- Compliance aspect[1]
- Legal aspect (see Footnote 1)

It appears that the aspects discussed just now are most affected. As far as we are aware of, we did not identify any other control aspects that should be added to the testing framework aimed at assessing the control of an algorithm.

From the case study and other works it seems that enhancing knowledge within an organization about the inner workings of the algorithms is important. Therefore, a multidisciplinary approach is also important as it combines the knowledge of several disciplines (e.g., business and IT). Another finding is that it makes no sense to make a checklist and go through it in order to have an overview of all the risks. The risks associated with an algorithm depend on the context in which it is used. It is far more important that within the organization there is awareness and a basic level of knowledge about the algorithm. Knowing everything about the algorithm is virtually impossible, but organizations must be able to recognize the aspects, the level of consciousness, in order to hook up the right people from their specialism to the controlling process. These capabilities are required to ultimately be able to conclude that the application has been carefully developed, whereby the identified risks in the process have been thoroughly controlled in order to arrive at an algorithm that works sufficiently reliably. This is not only relevant for the organization itself, but also for supervisors of the algorithm, for example. It is therefore recommended to carry out the entire process from a multidisciplinary point of view, including drawing up the risk analysis. In this way there is timely insight into the risks in the various specialisms and this can be considered during the process.

[1] Not discussed in this chapter.

The study published by the Netherlands Court of Audit (2021) offers good frameworks for general control, reliability and safety, as well as model quality, data quality, and ethics, which are integrally interwoven with it. However, it can be noted that the assessment framework is generic in nature. It is a good solid foundation to be aware of the risks associated with an algorithm. The context must be leading for the interpretation of the assessment framework in practice, the general questions that must be answered prior to the application of the assessment framework help with this. The framework focuses on accountability afterwards, but also offers guidelines in advance in the field of quality aspects that are already relevant during the development and realization of the algorithms. Some of these have been identified separately, some of them have been included in the elaboration of the five perspectives. I have already noted that "outsourcing" is not specifically mentioned separately but is briefly mentioned under the perspective of management and accountability and does not appear explicitly in the other perspectives. Here too, the organization bears responsibility for the risks. In my opinion, this element could have been worked out emphatically in the assessment framework.

The "People/Culture" element is also not specifically mentioned in the assessment framework, but this can also partly be seen in conjunction with the multidisciplinary approach. However, the case study points out that involving people and assessing the culture is an important aspect. The cultural aspect should certainly not be underestimated. Ultimately people have to implement the algorithm and that is why it is important to involve them early in the development. Doing so will ensure that employees within the organization are not surprised by the change during the implementation, and consequently will resist it less.

We conclude that the assessment framework provides a broad basis for an audit. It is a generic framework that must be tailored to the situation and context of the algorithm. As mentioned, the testing framework serves as a practical instrument for the auditor and is a means of control afterwards. Of course, this framework can also be of great value and input at the front end for the quality requirements surrounding the creation and use of algorithms, at the front end of the process. It is important to involve the "People/Culture" element, not only in the development, but also the people in the organization who will be involved in the implementation, so that they are included in the change and are involved in the implementation. Don't be surprised by. This is partly reflected in the multidisciplinary teams. There is some overlap in this but is not mentioned separately as an aspect.

It is up to the organizations themselves to gain insight into the algorithms and their use and to realize how powerful and important the role of algorithms in a certain process can be. To subsequently deal with this in a good and controlled manner, focus should not only be on the opportunities and on the effectiveness and efficiency of the process, but also on the awareness and importance of the creation, implementation, and control of the process. Is the algorithm able to be 'accountable'? The management aspects recognized from my research offer the auditor guidelines for assessing the reliable operation of an algorithm that is relevant to the audit object of the IT auditor. These management aspects partly overlap in the available assessment framework published by the Court of Audit, which has been elaborated based on perspectives.

References

de Boer, M. (2019). Vertrouwen in een algoritmiserende samenleving: Hoog tijd om algorithm assurance op te pakken. *De IT-Auditor*. Retrieved from https://www.deitauditor.nl/wp-content/uploads/2019/04/Hoog-tijd-om-algorithm-assurance-op-te-pakken.pdf

Kohnfelder, L., & Garg, P. (1999). The threats to our products (Vol. 33). Microsoft Interface, Microsoft Corporation.

Liebchen, G. A., & Shepperd, M. (2008, May). Data sets and data quality in software engineering. In *Proceedings of the 4th International Workshop on Predictor Models in Software Engineering* (pp. 39–44).

NBA. (2020). *What if Wat als auditors een rol gaan spelen bij het temmen van algoritmes?*

Netherlands Court of Audit. (2021, January). *Understanding algorithms*. Retrieved from https://english.rekenkamer.nl/publications/reports/2021/01/26/understanding-algorithms#:~:text=The%20Court%20of%20Audit%20found,use%20and%20operation%20of%20algorithms

Serrurier Schepper, M. S., & Hiddink, T. (2019). *Artificial Intelligence in actie* (1st ed). van Duuren Management.

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal, 31*(2), 47–53.

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review, 96*(4), 114–123.