UNIVERSITY OF TURKU

ABSTRACT

- [ ] Bachelor's thesis
- [x] Master's thesis
- [ ] Licentiate's thesis
- [ ] Doctoral dissertation

Subject	Information Management	Date	31.05.2022
Author(s)	Tom de Vries	Number of pages	72 + 6 appendices
Title	ANOMALY DETECTION IN IT AUDIT: The possibilities and potential in the domain of IT Audit		
Supervisor(s)	Dr. Joris Hulstijn		

IT Audit is dealing with a continuous increase in complexity and work. Regulations get stricter, while IT plays an increasingly more important role in companies. New technologies like anomaly detection can play a role in supporting IT Audit decisions. Anomaly detection has recently seen use in many domains, including financial audit, for example in fraud detection. Yet IT Audit does not make use of this technology as of now. This research looks at the possible roles that anomaly detection can play in this domain.

This research starts by attempting to bring the existing literature on both domains closer together and then creating variables that influence successful anomaly detection implementation in IT Audit. Exploratory interviews led to different approaches to implementation. IT Audit currently works with random samples to offer reasonable assurance on a statistical basis. As anomaly detection requires more data than the samples can provide, the potential benefits and consequences of utilizing the entire data population in an audit are researched.

As controls are unique to each client, IT Audit tasks have been grouped per common IT risk. For each risk, the potential of anomaly detection is determined based on four variables: the impact of erroneous instances going undetected, the time spent on the audit task, the frequency of the task, and the external pressure. Interviews with IT Audit professionals have been used to go through the IT risks with the highest potential, and determine the challenges. For each challenge, solutions have been discussed, as well as their feasibility.

Two use-cases have been formulated based on the interviews. The first use-case aims to use anomaly detection to detect multiple manage change risks, by looking at the full data population of changes at big clients working in standardized systems. The second use-case aims to discover SoD concerns and could be combined with financial audit data to discover fraud. Unsupervised deep learning methods are most likely to succeed. Prior research indicates deep autoencoder neural networks as a suitable method.

The biggest challenges for implementation turned out to be in the current audit methodology, rather than development. The current sample approach is based on the notion that testing the full data population would not be possible while remaining within time and budget norms. New techniques, such as anomaly detection, might mean this notion is outdated, but the methods cannot be created and optimized due to the current restraints.

Keywords	IT Audit, anomaly detection, machine learning

ANOMALY DETECTION IN IT AUDIT

The possibilities and potential in the domain of IT Audit

Master's Thesis in
Information Management

Author(s):
Tom de Vries

Supervisors:
Dr. Joris Hulstijn

17.03.2022
Amsterdam

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AD	Anomaly Detection
AE	AutoEncoder
AI	Artificial Intelligence
ART	Accountability, Responsability, Transparency
CAATs	Computer Assisted Audit Techniques
CCM	Continuous Control Monitoring
DAD	Deep Anomaly Detection
DAENN	Deep AutoEncoder Neural Network
DL	Deep Learning
GL	General Ledger
GLAD	General Ledger Anomaly Detection
IPE	Information Produced by Entity
ISA	International Standards on Auditing
IT	Information Technology
ITAC	IT Application Control
ITDM	IT Dependent Manual Control
ITGC	IT General Control
JDM	Judgement and Decision-Making
m	Measured
MA	Manage Access
MC	Manage Change
ML	Machine Learning
MO	Manage Operations
M_t	Measured True
NN	Neural Network
PCA	Principal Component Analysis
Pr	Precision
Re	Recall
RP	Rank Power
SCOT	Significant Classes Of Transactions
ToD	Test of Design
ToE	Test of Effectiveness

1 INTRODUCTION

1.1 Background

The amount of data within organizations keeps increasing at a fast pace, and with it, the demand for companies to continuously analyze the data increases as well. For IT Audit the increase in data means that the way of working will have to be adapted constantly and innovation is a must to keep up with demand and competition. The amount of work in IT Auditing is increasing, and the growth in IT Auditors is unable to keep up with this pace. A solution to this problem would be to look for new technologies that can support the IT Audit practices. One such technology is anomaly detection, enabled by a Machine Learning (ML) algorithm. Anomaly detection.

> *"Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior." (Chandola et al., 2007)*

People tend to mix the definitions of narrow AI and Artificial General Intelligence (AGI). AGI refers to a "True" intelligent machine that can think for itself, solve a variety of problems, and is capable of having its own thoughts. (Goertzel & Pennachin, 2006) While AGI is more in line with the Hollywood depiction of AI, this research will be focused on narrow AI, which refers to an algorithm with a very specific task to excel at, while being virtually useless for any task outside of its scope. To further narrow down the scope, this research looks at a sub-domain of AI called Machine Learning. ML can make some tasks a lot more efficient, while it can even completely or partly re-place human interaction in other tasks. AI is by no means a new concept, but in recent years, the increased computing power has allowed for accelerated development in AI solutions. In the coming years, AI is expected to grow even more, and in doing so it will have an enormous impact on the way we work. This chapter will introduce the subject and make a connection between AI and IT Audit.

> *"Information technology (IT) auditing examines processes, IT assets, and controls at multiple levels within an organization to determine the extent to which the organization adheres to applicable standards or requirements." (Gantz, 2013)*

The definition of IT Audit provided by Gantz (2013) describes that processes and controls are compared to a standard. ML could potentially go through the data that comes from

these processes and find anomalies for the IT Auditor, which can potentially improve accuracy and quality while reducing human interaction on repetitive tasks. IT Audit follows strict rules and guidelines, which could provide a basis for the objective of an algorithm. While AI and automation are increasingly being recognized and integrated into the financial audit, the use of AI in IT Audit is still in its infancy. Recently, a team of researchers in EY Japan (EY, 2019) has managed to create an AI anomaly detection algorithm that can detect accounting fraud in General Ledgers(GL). While this application called EY Helix GLAD (General Ledger Anomaly Detection) is still in a testing phase, it demonstrates anomaly detection's potential for audit.

Anomaly detection can find outliers in data, which indicate a suspicious occurrence by defining a model of what is normal data in a process. (Rao et al., 2011) What is normal data in IT Audit can be determined by looking at existing auditing rules and practices. To pave the way for the implementation of AI-enabled anomaly detection, this research will delve deeper into the most prominent IT Audit tasks and processes, and compare the involved data with the data requirements of different machine learning techniques.

1.1.1 Company Introduction

This research is written as part of a thesis internship at Ernst & Young (EY) in the Technology Risk department. EY is an international firm that operates in four integrated lines of service: Assurance; Consulting; Strategy & Transactions; and Tax. The focus of this research lies on Assurance, or to be more precise: IT Audit. Spread across its service lines, EY served most of the companies in the Forbes Global 2000 in the year 2021 and is ranked number 1 worldwide in audit market share based on deal numbers, or number 2 based on proceeds. (EY, 2021)

1.1.2 Problem Statement

IT Auditing practices are work-intensive and require secure and precise measures. Anomaly detection has been used in many areas to reduce the workload and/or to provide extra measures. In many areas where data occurs in large quantities, AI has even proven to be more proficient in processing than humans themselves. Many AI experts believe that in the coming few years AI will outperform people in many tasks like driving, writing essays, and writing books, according to research by Grace et al. (2018).

Successfully implementing anomaly detection could offer a competitive advantage in a highly competitive setting where constant innovation and adaption is required, by potentially increasing efficiency and quality/accuracy of the IT Audit. Being able to look at more data (more cases) in the same timeframe means more assurance. A higher quality IT Audit could also mean improved security for the audited party.

Implementing anomaly detection brings its own set of problems to the table. There are many different approaches, and **each situation will require a matching approach** based on, among other things, the type of data (sequential or non-sequential), the quality of the data, and the required accuracy of the outcome. In general, an IT Audit has a **lower quantity of data** than financial audit. To add to the difficulties, IT Audit also has **less structure in its data**, as the numerical data for financial audit is easier to analyze than the mixture of textual data, diagrams, images, code, etc. involved in IT Audit. Therefore, the challenge for applying anomaly detection in this field lies in **selecting and examining the tasks and processes**.

1.1.3 Research Design

This research looks at an existing solution (anomaly detection) and tries to apply this in new territory (IT Audit). Both the anomaly detection perspective and the IT Audit perspective will be taken in this research to look for compatibilities and to try and resolve any potential conflicts.

The first part of the research will be a literature review (for both IT Audit & anomaly detection), followed by observations on the current situation at EY regarding the way of working and the tools and data that are being used. Next, the apparent opportunities for anomaly detection will be described, and variables for successful implementation will be determined based on interviews. By analyzing these variables for IT Audit tasks that show potential, challenges for implementation can be discovered.

1.2 Research question

This chapter will describe the main question that this research will try to answer in order to fulfill its purpose, as well as the sub-questions that will need to be answered before the central research question can be answered. The research questions aim to look at two different domains (IT Audit & anomaly detection) and cover both perspectives.

1.2.1 Central Research Question

Answering the central research question should provide EY with the necessary information on whether or not to proceed with anomaly detection in IT Audit, which methods to use, and which tasks and processes would benefit the most. The central research question has been defined as follows:

"How can anomaly detection be applied to the IT Auditing process to make a significant contribution to the efficiency and/or quality?"

1.2.2 Sub Questions

To answer to the central research question, more specific sub-questions will have to be answered. Answering these sub-questions should provide the necessary information to answer the central research question.

1. What are the most prominent ML techniques for anomaly detection and what are their (data) requirements?
2. How to select the appropriate machine learning technique for anomaly detection?
3. Which tasks and processes in IT Audit would benefit most from anomaly detection?
    a. What are the selection criteria to find promising tasks and processes?
    b. What are the specific benefits for those tasks and processes that fit the criteria?
    c. What are the challenges for those tasks and processes that fit the criteria?
4. Do IT Audit processes currently meet the data requirements and process requirements for anomaly detection techniques or can they be adapted to do so?

2 LITERATURE REVIEW

This chapter provides a compact overview of existing theory on the subject that this research will build upon. The Literature review has been divided into two parts. The first of which is this chapter about IT Audit, the second is a part about anomaly detection and ML.

This chapter provides a compact overview of existing theory on anomaly detection and ML techniques that this research will build upon. This report assumes that the reader has a basic understanding of AI machine learning. Basic concepts such as what is a neural network, what is machine learning, etc. will consequently not be explained. However, different types of specific architectures and subtle differences will be explained to justify the decisions that have been taken.

2.1 IT Audit

2.1.1 Definitions

While the term "audit" can be used to describe a wide variety of domains, it is commonly associated with financial audit. (Gantz, 2013) Information Technology Infrastructure Library (ITIL) defines an audit as:

> *"formal inspection and verification to check whether a standard or set of guidelines is being followed, that records are accurate, or that efficiency and effectiveness targets are being met."*

The International Organization for Standardization (ISO) guidelines on Audit describe audit as:

> *"systematic, independent and documented process for obtaining audit evidence and evaluating it objectively to determine the extent to which the audit criteria are fulfilled"*

Both ITIL and ISO maintain a definition of Audit that is non-specific to financial audit and leaves room for IT audit as a sub-category. IT Audit is a sub-category of audit that supports the client in realizing the risks that are connected to IT, as well as assessing the controls that are in place to mitigate these risks. (EY Technology Risk, n.d.)[1] In the

[1] Source from internal EY portal which is not publicly available

introduction, the definition of IT audit by Gantz (2013) from the book "***The Basics of IT Audit: Purposes, Processes, and Practical Information"*** was already provided.

2.1.2 Objectives

In the book "***Accounting Information Systems"*** (Romney, 2017) describes six objectives for typical information system audits:

1. **Objective 1: Overall Security**
   IT and information is protected from unauthorized access, modification, or destruction.
2. **Objective 2: Program Development and Acquisition**
   Development and acquisition is properly authorized
3. **Objective 3: Program Modification**
   Modifications are properly authorized
4. **Objective 4: Computer Processing**
   Processing of significant data is accurate and complete
5. **Objective 5: Source Data**
   Inaccurate data and improper authorizations are identified and resolved
6. **Objective 6: Data Files**
   Data files are accurate, complete, and confidential

The purpose of an IT Audit is to ensure that IT processes and applications can be evaluated as effective or reliable. This means that stakeholders can rely on IT to work for its intended (audited) purpose. (EY Atlas, n.d.)[2]

2.1.3 Controls

Controls play an important role in IT audit. Controls are used to ensure that IT applications and processes work as intended, and IT Audit involves making sure that the right controls are in place, and that they are effective. Controls provide efficiency, effectiveness, compliance, reliability, and assurance to the IT capabilities of a company. Gantz (2013) divides controls into three categories: preventive, detective, and corrective. Gantz continues to separate the controls further by their function, into administrative, technical, and physical.

---

[2] Source from internal EY portal which is not publicly available

Romney (2017) describes controls for many different components that can be the subject of an IT Audit, like access management, product versions, or path management. For access control, for example, controls could be reviewed regarding policies, procedures, and mechanisms that are in place to manage or restrict access to the subject.

2.2 IT Audit Process

IT Audit processes are difficult to put on paper as there is no standard process or procedure. IT Audit is dependent on both the way of working from the auditor and the circumstances of the audited party. A general overview of typical IT Audit processes can still be found in multiple sources. Hinson (2007) lists the following items as typical IT Audit work:

- **Operational computer system/network audits:** Review the controls
- **IT installation audits:** Review physical equipment and physical security
- **Developing systems audits:** Project management controls and information security controls
- **IT governance, management, and strategic audits:** Review governance (including strategies, visions & plans) and possible external IT relationships
- **IT process audits:** Review processes within IT
- **Change management audits:** Review changes that will be made or have been made to ensure previous controls will not be impacted
- **Information security and control audits:** Review controls regarding confidentiality, integrity, and availability of data
- **IT compliance audits:** Review of compliance with external constraints such as laws
- **Benchmarking**: Benchmark of IT systems and performance with competitors
- **Contingency planning:** Review of the plans and security measures that are in place to recover in case of emergencies
- **Special investigations:** Audit in unforeseen circumstances such as a possible security breach
- **Other**

Each of the items in the list described by Hinson has various tasks, processes, and controls that are considered part of an IT Audit. In the book ***"Accounting Information Systems"*** by Romney (2017), an overview of a typical IT Audit from beginning to end (figure 1) is provided.

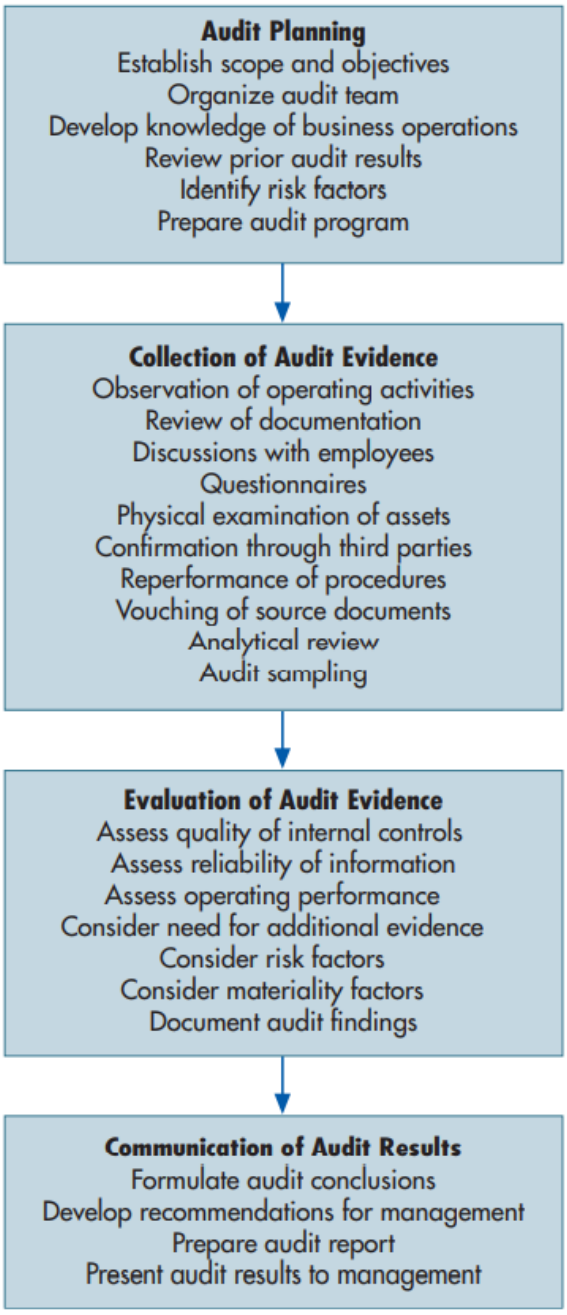

Figure 1 Overview of the Auditing Process (Romney, 2017)

Romney divides IT Audit into four stages: Audit planning; Collection of audit evidence; Evolution of audit evidence; and communication of audit evidence

2.3 Continuous Control Monitoring

In a typical IT audit, evidence is still gathered in many forms, including screenshots, emails, textual data, and Excel sheets. This evidence is sent to the auditor, usually on request of this specific data, and later analyzed by the IT Auditor. The evidence is requested to support a certain period, like three months for example, and is then repeated as many times as necessary to cover the full year. This approach is getting increasingly difficult to maintain, as the amount of data in organizations is growing while simultaneously the IT Audit requirements are getting more thorough.

CCM is a change in this process the data is monitored in real-time. The access to the system can be limited with access right to limit the IT Auditor to only the part that is being audited. In research from Brennan and Teeter (2010), CCM is described in three parts: first, establishing a foundation for the monitoring of controls. Second, based on the risks that the IT Audit is supposed to cover, procedures can be created and put into practice. And third, the results have to be communicated with the audited party so that corrective action can be taken.

2.4 Audit Risk Model

ISA 400 Risk Assessment and Internal Control is an auditing standard that is part of ISA (International Standards on Auditing). ISA 400 is based on the notion that an auditor does not always detect an error or misstatement. The risk that this occurs consists of a combination of inherent risk and control risk. Inherent risk is the chance that a misstatement (material or other) occurs, before taking into account the related controls. Control risk is described as the risk that the internal control systems do not detect an error or misstatement in a timely manner. (Blokdijk, 2004)

2.5 Anomaly detection

According to Mehrotra, K. G., Mohan, C. K., & Huang, H. (2017). anomalies or outliers can be defined as *"substantial variations from the norm"*. The norm refers to the normal (or expected) instance of a process, which will result in data that is comparable to other instances of the same process. Whenever an instance varies from the norm, this will be observable in the resulting data. When talking about anomaly detection algorithms, it generally refers to the use of AI in the form of machine learning algorithms to help the detection of said variations from the norm.

Helix GLAD, which was briefly mentioned in the introduction of this research, is an example of successful use of anomaly detection in audit. Helix GLAD is capable of going through General Ledger files with millions of entries and flagging the outliers that may be in there. The tool demonstrates that AI is not a replacement for auditors, but simply flags outliers that the auditors have to investigate further. Thus reducing repetitive tasks and providing more accuracy in the process. (EY, 2019)

Other areas where ML-based anomaly detection has been successfully implemented include insurance, healthcare, banking, telecom, and fraud detection. (Chalapathy & Chawla, 2019)

2.5.1 Measurements

The performance of anomaly detection algorithms is most commonly measured using precision, recall, and Rank-power. (Mehrotra, Mohan, & Huang, 2017)

2.5.1.1 Precision

Precision measures how many of the measured anomalies (m) are true anomalies (mt) using the formula:

$$Pr = \frac{m_t}{m}$$

The aim is to get precision as close to 1.0 as possible, meaning all measured anomalies are true anomalies. (Mehrotra, Mohan, & Huang, 2017)

2.5.1.2 Recall

Recall measures how many of the true anomalies (dt) are correctly identified by the algorithm as anomalies (mt) using the formula:

$$Re = \frac{m_t}{d_t}$$

The aim is to get recall as close to 1.0 as possible, meaning all true anomalies have been correctly identified by the algorithm. (Mehrotra, Mohan, & Huang, 2017)

2.5.1.3 Rank power

Precision and recall measure each outlier as equal, while in reality the outliers can be ranked based on how far they defer from the norm. Thus, only measuring precision and

recall would not give a fair indication of the performance of an algorithm. Rank power is used to measure the true anomalies sorted by degree of suspicion using the formula:

$$RP = \frac{m_t(m_t + 1)}{2\sum_{i=1}^{m_t} R_i}$$

(Mehrotra, Mohan, & Huang, 2017)

2.5.2 Datatypes In Anomaly Detection

Anomaly detection is used to detect variations from the norm, which means that a norm has to be defined beforehand. However, not all types of data can be measured in the same manner. Common needs for all data types to be used in anomaly detection are that there have to be rules or principles to the data, and the sample data should have a similar distribution to other data. (Mehrotra, Mohan, & Huang, 2017)

2.5.3 Types of Anomaly Detection

Mehrotra et al. (2017) describe three primary approaches to anomaly detection: distance-based; density-based; and rank-based. For each of these, the nature of the data can be supervised, semi-supervised, or unsupervised.

1. **Distance Based:** This approach considers data points to be more anomalous if they are far away from others.
2. **Density-Based:** This approach considers data points to be more anomalous if they are in a low-density area (not part of a cluster).
3. **Rank-Based:** This approach considers data points to be more anomalous if their closest neighbors have otherdata pointss closer to them.

The nature of the data that is used in anomaly detection can be either supervised, unsupervised, or semi-supervised. Supervised data involves a training set of data where the data is labeled, while unsupervised data does not require any labels, and data points are compared to all data in the set instead of the training data. (Mehrotra, Mohan, & Huang, 2017)

2.6 Machine Learning

ML is a subdomain of AI, that involves an algorithm that can learn and improve. Bonaccorso (2018) describes the main goal of ML as *"to study, engineer, and improve*

mathematical models that can be trained (once or continuously) with context-related data (provided by a generic environment) to infer the future and to make decisions without complete knowledge of all influencing elements (external factors)".

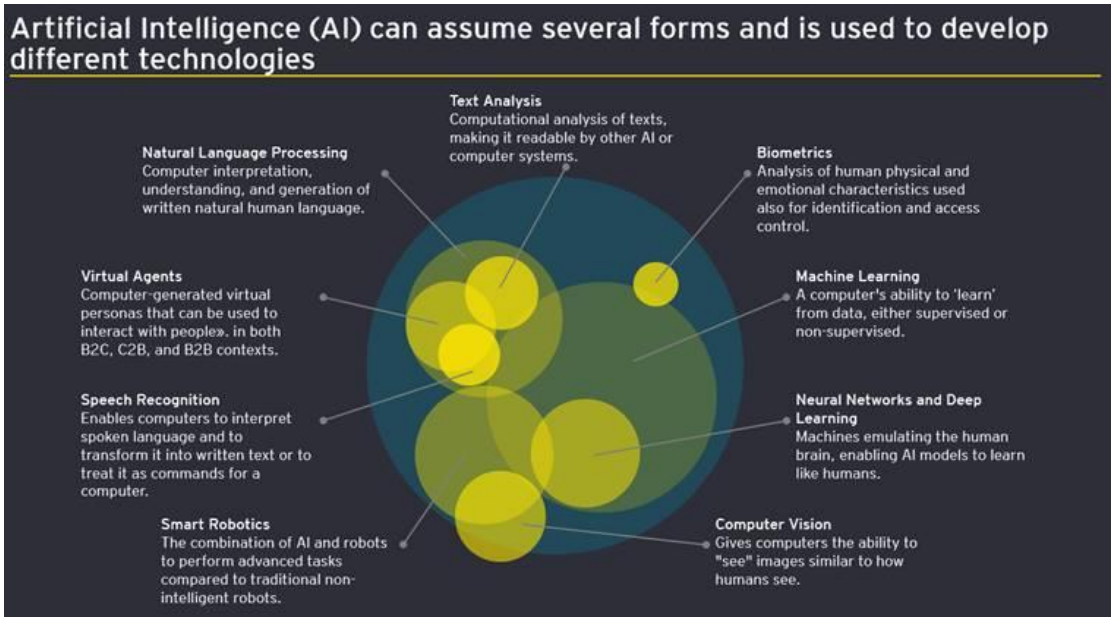

Figure 2 AI Subdomains (EY Atlas, n.d.)

Figure 2 AI Subdomains (EY Atlas, n.d.)" shows the different subdomains of AI and where they overlap. ML is one of the most important and widespread subdomains of AIorithm and is becoming increasingly more common in every field. (Bonaccorso, 2018)

2.6.1 Data Complexity

In research by L. Li and Y. Abu-Mostafa (2006), the statement is made that ML is about pattern extraction. This refers to the ability that ML has to recognize complex patterns. High complexity in data is not a problem for ML, and in fact, is where its strength lies. The pattern that is mentioned in the research refers to a hypothesis, rule, or structure.The simplicityy or complexity of a pattern can be measured by whether or not it can be generated by a program or compressed.

> *"It says that a pattern is simple if it can be generated by a short program or if it can be compressed, which essentially means that the pattern has some "regularity" in it." (L. Li and Y. Abu-Mostafa, 2006)*

The research continues by stating that the ideal complexity for ML is such that the patterns are infeasible for other solutions.

2.7 Machine Learning Methods

ML is a subcategory of AI and refers to self-learning algorithms. There are countless variants of ML algorithms that have been developed over the years, and generally, they have their own applications, strength,s and weaknesses. Anomaly detection makes use oan a ML algorithm to determine what is an outlier and what is normal data. Not all ML algorithms are suited for the purpose of anomaly detection, and even for those that are, the effectiveness will depend severely on the match between the method and the data. While there is no fool-proof method of selecting the ML algorithm that will have the best results, the algorithms can be filtered based on prior use in anomaly detection, intuition based on experience, and their known strengths.

In the book ***Machine Learning Algorithms: Popular algorithms for data science and machine learning*** (Bonaccorso, 2018) various ML techniques are explained. Table 1 ML Methods Divided by Category" contains a non-exhaustive list of common ML methods divided by category.

Table 1 ML Methods Divided by Category

Supervised	Dimensionality Reduction	Linear Discriminant Analysis
	Regression	Linear Regression
		K-Nearest Neighbors Regression (KNN)
		Random Forrest Regression
		Decision Tree Regression (CART)
		Support Vector Regression (SVR)
		Locally Weighted Scatterplot Smoothing (LOWESS)
		Multivariate Adaptive Regression Splines (MARS)
	Classification	Naive Bayes
		Logistic Regression
		K-Nearest Neighbors Regression (KNN)
		Extreme Gradient Boosting(XGBOOST)
		Gradient Boosted Trees
		Adaptive Boosting (AdaBoost)
		Random forest Classification
		Decision Tree classification (CART)
		Support Vector Machine (SVM)
Semi-super-vised		Label Propagation
		Label Spreading
		Self-training Classifier
		Uniform Manifold Approximation and Projection (UMAP)

Unsupervised	Dimensionality Reduction	Locally Linear Embedding (LLE)
		t-Distributed Stochastic Neighbor Embedding (t-SNE)
		Isomap Embedding
		Multidimensional Scaling (MDS)
		Principal Component Analysis (PCA)
	Clustering	K-Means
		Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
		Hierarchical Agglomerative Clustering (HAC)
		Gaussian Mixture Models (GMM)
	Association	Apriori
Reinforced		Deep Q Neural Network (DQN)
		Q-Learning
		SARSA (State-Action-Reward-State-Action)
		Proximal Policy Optimization
		Policy Gradient
		Temporal Difference
		Monte Carlo Methods
Neural Networks	Auto Encoders	Sparse Auto Encoder (SAE)
		Denoising Auto Encoder (DAE)
		Variational Auto Encoder (VAE)
		Auto Encoder (AE)
	Recurrent Neural Networks	Recurrent Neural Network (RNN)
		Long Short Term Memory (LSTM)
		Gated Recurrent Unit (GRU)
	Convolutional Neural Network	Deep Convolutional Network (DCN)
		Deconvolutional Network (DN)
		Deep Convolutional Inverse Graphics Network (DCIGN)
	Feed Forward Neural Networks	Deep Feed Forward Neural Networks (DFF)
		Feed Forward (FF)
	Generative Adversarial Networks	Generative Adversarial Networks (GAN)
Other	Probabilistic Graphical Models	Bayesian Belief Networks

Out of the methods listed in Table 1 ML Methods Divided by Category", many methods have already been successfully utilized for the purpose of anomaly detection. To narrow down the selection, this research will instead focus on recently conducted prior research in anomaly detection for closely related domains such as financial audit.

2.7.1 Method selection

In research about Deep Learning (DL) for anomaly detection by R. Chalapathy and S. Chawla(2019), the use of DL for anomaly detection is advocated. DL is a form of neural network that functions with a high amount of layers. DL for anomaly detection is also referred to as DAD (Deep Anomaly Detection). Chalapathy and Chawla provide a few factors to help in the metho selection. The nature of the data is the key factor in deciding on a DAD method. The data can be classified as either low-dimensional or high-dimensional, based on the amount features. DAD methods have been proven to function better on high-dimensional data.

A secondd factor in ML method selection is between supervised, unsupervised, and semi-supervised methods. While supervised ML methods can perform well for anomaly detection, it is less common than unsupervised and semi-supervised methods. This is mostly due to the lack of training data, as supervised methods require a large amount of labeled data that includes anomalies. Another weakness of supervised methods is that anomalies tend to change over time, which is more difficult to incorporate into supervised methods. (Chalapathy & Chawla, 2019) However, when supervised anomaly detection can be implemented, it performs better than unsupervised methods. (Görmitz et al., 2013)

2.7.1.1 Unsupervised method: Deep Autoencoder Neural Networks

The book ***"Deep Learning (Adaptive Computation and Machine Learning series)"*** (page 499 – 523) by Goodfellow and Courville, defines autoencoders as *"a neural network that is trained to attempt to copy its input to its output."* The goal of the autoencoder here is not to just copy the input to the output, but to learn the properties of the data. Typically, deep autoencodconsistists of two nearly identical neural networks. One for encoding and the other for decoding. Figure 3 Simplified Deep Autoencoder Neural Network Architecture" shows a simple example of an autoencoder architecture. The neural netwoconsistsist of the input layer, the output layer, and several hidden layers in between. The input layer and output layer contain the same number *n* of nodes. The hidden layers in between are builbuiltmetrically around a central hidden layer that acts as a bottleneck. After this bottleneck, the autoencoder is supposed to build the output layer to resemble the input as closely as possible. Since the autoencoder neural network will be trained on data that is considered "normal", on actual data the output should resemble the input more

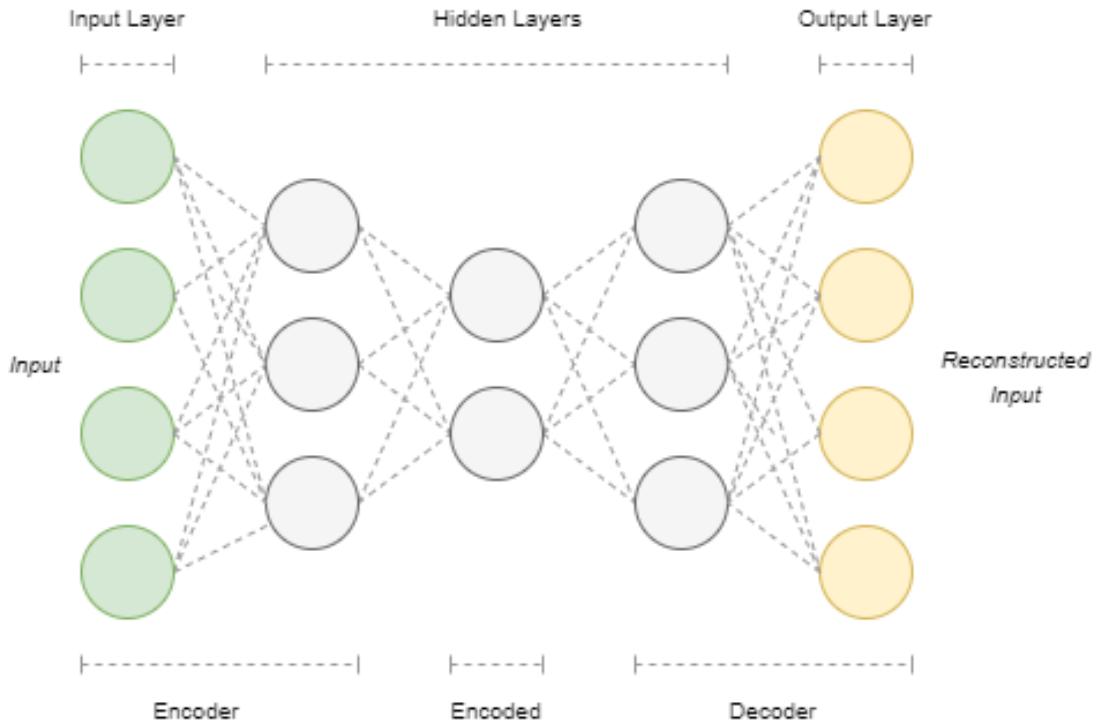

Figure 3 Simplified Deep Autoencoder Neural Network Architecture

closely if the data contained only "normal" data and resemble the input less if there are anomalies in the data. The loss in quality of the reconstructions determines how anomalous the instance is.

2.7.1.1.1 Supporting Research

In research by Schreyer et al. (2018), Deep Autoencoder Neural Networks are used for anomaly detection to detect accounting fraud in financial audits. One of the benefits of this method is that Deep Autoencoder NNs are a form of unsupervised learning, meaning unlabeled data can be analyzed without the need of human interaction. Unlike unsupervised learning, supervised algorithms require work upfront to label the dataset. An argument against Deep Autoencoder NNs is that supervised methods tend to have more accurate results. However, the added amount of work that is required to label the data makes supervised learning less desirable for the first implementations of anomaly detection. Another advantage of Deep Autoencoder NNs is that only one class of events is required in the training data. This means that the training set does not need to include a large number of examples of exceptions. In fact, no examples of exceptions need to be in the training data and Deep Autoencoder NNs will still be able to flag unexpected events (anomalies). This makes Autoencoders suitable for instances where the available training data does not include many instances of exceptions. (Lin & Jiang, 2021).

Another study that advocates for the use of Deep Learning in anomaly detection is the research by Bengio, Y. (2009) Deep Learning is a subset of machine learning that differentiates itself by the number of layers that are used in the neural network architecture. Deep Learning allows an algorithm to learn from multiple levels of abstraction (Bengio, Y., 2009)

2.7.1.1.2 Data Limitations

Earlier in this chapter, research by Schreyer et al. (2018) was mentioned for their use of Deep Autoencoder Neural Networks to detect accounting fraud. This research utilizes adaptions to the autoencoder architecture that have been made by Zhou and Paffenroth (2017). The adaptions include a filter layer and regularization penalty, which increase the effectiveness when clean training data is not readily available. As described earlier, an advantage of autoencoders is that they do not need to be trained on a combination of both "normal" data and outliers, but can instead function when trained on only one class of events.

2.8 Data Preparation

Before any ML algorithm can be created and implemented, the data will have to be prepared in a way that is compatible with the intended use. Data preparation is a large and important part of ML implementation, and there is a large amount of research on the topic. A recent book that touches on this topic is ***"Machine Learning Algorithms: Popular algorithms for data science and machine learning"*** by Bonaccorso (2018). The data

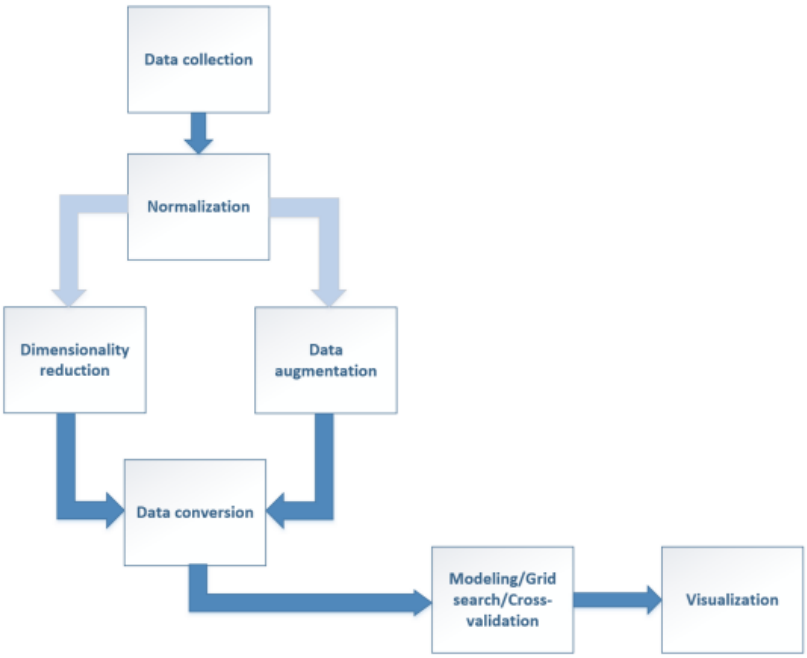

Figure 4 Data Preparation Diagram (Bonaccorso, 2018)

preparation is modeled into 7 steps, which can be found in Figure 4 Data Preparation Diagram (Bonaccorso, 2018)".

2.8.1 Data Collection

The first step is to collect the data. Bonaccorso mentions that a CSV file format or comparable is preferred, but in reality, it will often be a mix of multiple sources that will have to be combined. This step depends heavily on the situation, in IT Audit possibilities would be to request Excel/CSV extracts from audited systems like SAP that span over the period that is being audited, or to utilize CCM (see paragraph 2.3) to gain direct access to the required data.

2.8.2 Normalization

Normalization is about scaling the data. Different features in the data will have different scales, which can result in a disproportional effect on the outcome. To make the effect of features more proportionate, the data has to be scaled. (Bonaccorso, 2018) Ways to normalize the data include StandardScaler and RobustScaler, which are both part of the sklearn (Scikit-learn) library for Python.

2.8.3 Dimensionality Reduction

To reduce the computational time of the algorithm and avoid memory leaks, some features can potentially be removed to compress the data. Features can have a correlation without it being apparent at first thought, which means it is important to the effect this will have on the outcome. A common method is to use the Principal Component Analysis (PCA) to determine the features that can be removed without significantly impacting the outcomes. (Bonaccorso, 2018)

2.8.4 Data Augmentation

When the training data does not contain enough non-linear features, it can be difficult to understand the correlation between the data without overfitting on the training data. To combat this, new features can be created based on information from the original features. Bonaccorso (2018) refers to PolynominalFeatures, which is part of the sklearn library in Python.

2.8.5 Data Conversion

The data conversion step involves the encoding of the labels. Depending on the choice of conversion, focus can be adjusted to reduce noise and reduction errors, or efficiency. (Bonaccorso, 2018)

2.8.6 Modeling/grid search/cross-validation

Basing the decision for the choice of classification/clustering algorithm on a grid search can support the decision. The sklearn library in Python includes a method to test multiple models to determine the most suitable one for the data.

2.8.7 Visualization

The final step in the data preparation diagram by Bonaccorso (2018) is to visualize the results from prior steps by using plotting functions that are part of Python libraries such as matplotlib from SciPy.

2.9 Responsible AI

In a lecture about responsible AI, H. Weigand (2021)[3] named three parts of responsible AI: Accountability, Responsibility, and Transparency (ART). Accountability means that control over the AI has to be maintained and someone has to remain accountable for decisions that are made based on AI. Responsibility is described by Weigand as *"an ethical concept that refers to the fact that individuals and groups have morally based obligations and duties to others and to larger ethical and moral codes, standards and traditions."* Lastly, transparency refers to the ability to describe how the conclusions have been reached by the AI and how decisions have been made.

The link to (IT) Audit is that there is a high importance to being able to defend the choices that have been made and even to reproduce the outcome. There are three ways to achieve transparency: (1) simulatable models, decomposable models, and algorithmically transparent models. (Arieta et al, 2020)

[3] Lecture slides not publicly available

3 THEORETICAL FRAMEWORK

3.1 Judgement and Decision-Making

The judgment and decision-making(JDM) framework was created in research by Bonner (1999) as a way to determine whether improvements to judgement and decision-making in accounting was necessary. This is done by determining where the problem is in the current approach, what the solution could be, and whether or not this is possible in practice. According to Bonner, judgments *"tend to take the form of predictions or an evaluation of a current state of affairs"* while decisions *"refers to making up one's mind about the issue at hand and taking a course of action".* Figure 5 Framework for JDM Research (Bonner, 1999)" shows a visual representation of the framework.

The framework revolves around three categories of variables that influence performance: person, task, and environment. Person refers to the decision-maker in the process, in this case, the auditor, and what variables will affect performance. Task refers to the task that has to be completed and the complexity that comes with it. Environment refers to the circumstances that the individual or group has to work in, liketime pressuree or rules and requirements imposed by law or policy.

Bonner describes the ultimate goal of JDM research as "improving decision making". However, Bonner created the framework for auditing, and audit has some major differences from IT audit. Research by Charyyeva (2017) applied the JDM theory to IT Audit by suggesting that judgment and decision-making in IT audit usually means risk assessment. This research by Charyyeva created factors that influence an IT Audit and confirmed the factors with interviews with IT auditors from varying experience levels. Charyyeva concluded on several variables that influence decision-making in IT Audit and divided those variables in cintoegories that fit in the original model by Bonner (1999), these variables have been summarized in Table 2 Factors that affect JDM in IT Audit (Charyyeva, 2017)".

Table 2 Factors that affect JDM in IT Audit (Charyyeva, 2017)

Person	Knowledge	Business Processes
		IT Processes
	Expertise	Understanding of Processes
		Experience
	Prior Belief	Critical approach
		Reliability of various factors
	Decision-Aid	Guideline Framework
		Client JDM
Task	Risk	Risk of material misstatement
		Change of audit strategy
	Complexity	Different systems
		Extent of customization
	Presentation Format	User-friendly interface
Environment	Group Information Processing	Collective understanding
		Communication
	Corporate Governance and Control Framework	Control framework
		Audit regulation type
		company size
	Pressure	Time-budget pressure
		External inspection
	Accountability	Reputation

3.2 Framework for JDM Research

The research by Bonner (1999) included a framework that is aimed at deciding if the current JDM is in need of improvement. The framework consists of six steps (shown in Figure 5 Framework for JDM Research (Bonner, 1999)"): (1) Is it currently known if there is a need for improvement? If the answer is no, then find out the answer before moving on. (2) Can the performance of JDM be improved? For this research, this step involves finding out the potential benefits that anomaly detection can have on current JDM in IT Audit. After it has been established that anomaly detection could potentially

improve IT Audit JDM, variables can be determined for each of the categories (Person, Task, Environment). (3) In this step, this research deviated from the original model by looking at the obstacles that have to be overcome before anomaly detection can be successful. The original model looks at the deficiency of the current JDM method instead. (4) The fourth step in the framework is to look at possible solutions for all of the deficiencies. In this research, the deficiencies will be the challenges for anomaly detection implementation. (5) The final step is to confirm that the solutions are actually feasible.

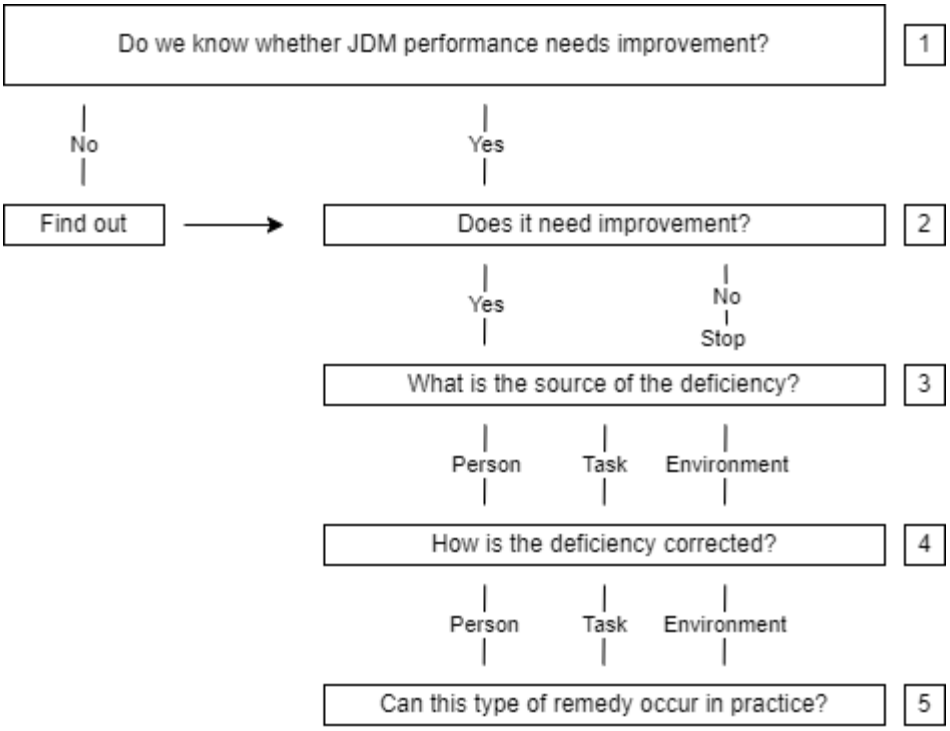

Figure 5 Framework for JDM Research (Bonner, 1999)

4 RESEARCH METHODOLOGY

This is a qualitative research using several different methods. The final product that this research produces is advice for implementation of anomaly detection in IT Audit, based on information that is collected through a systematic literature review, multiple rounds of interviews, and a case study of tasks within EY.

4.1 Study Design

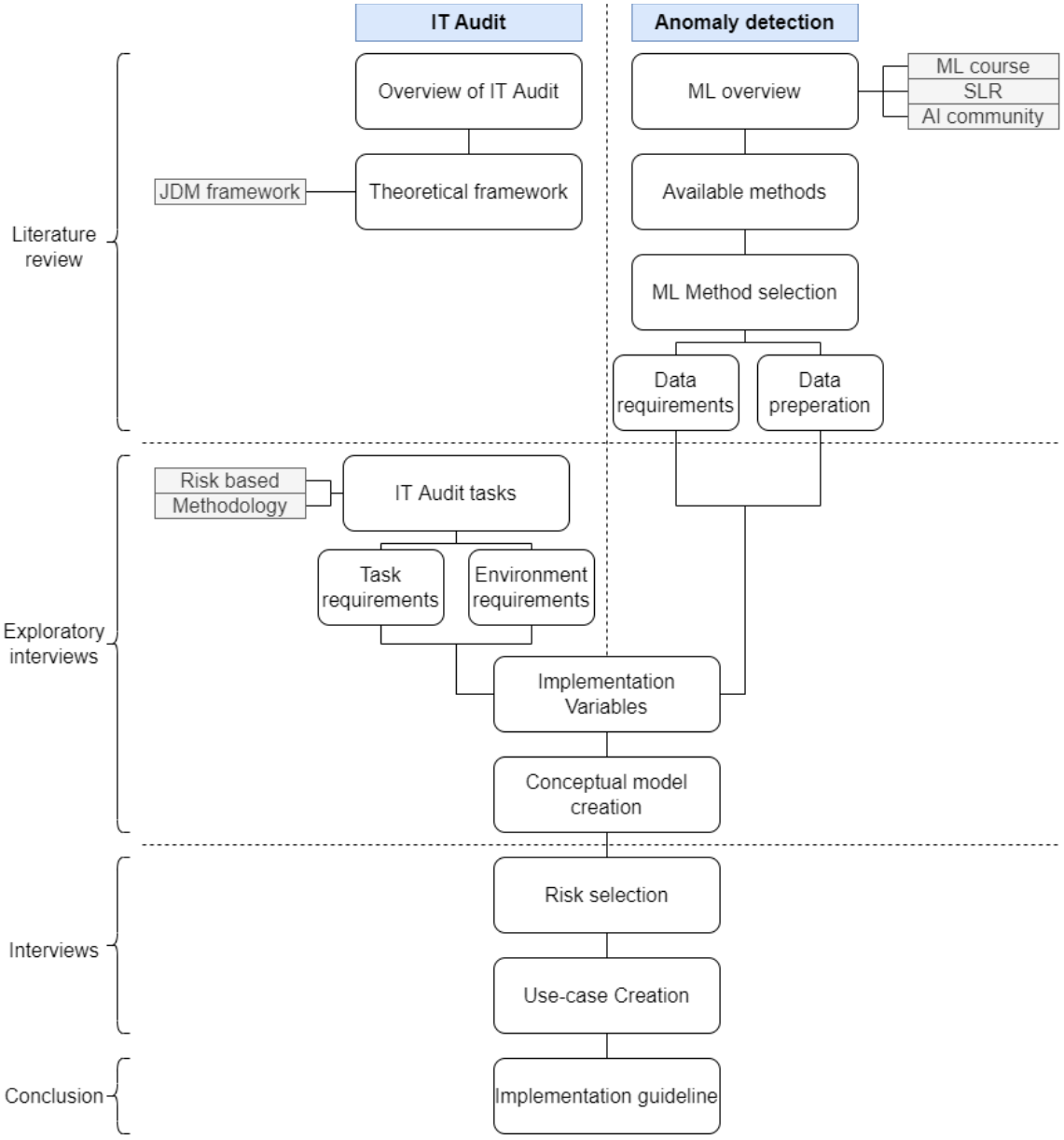

Figure 6 Study Design

Figure 6 Study Design" shows a visual representation of how the study has been conducted. The first step of the study was to bring the theory on IT Audit closer to the theory

on anomaly detection by comparing available ML techniques, their strengths, and their requirements to the goals for IT Audit and the tasks and data they have to offer.Systematicc literature has been conducted for both subjects by carefully choosing relevant keywords and reading studies on both subjects. Additionally, knowledge onofL techniques has been gained by the researcher by following courses in machine learning in Python.

The second part of the research consisted of exploratory interviews with IT Audit professionals within EY. Candidates have been selected based on experience with IT Audit and knowledge about AI/ML or other related subjects. The goal of the exploratory interviews was to gain initial knowledge on suitable tasks within IT Audit and potential challenges for implementation.

Another round of interviews has been conducted to verify findings, select potential tasks, and to discuss the challenges and solutions for implementation per task.

4.2 Design Science

The research was initially based on the design science research cycles by Hevner (2007), as can be seen in Figure 7 Adaption of Hevners Design Science Cycles". However, due to time constraints and scoping, early on in the research the decision was made to not produce a minimum viable product.

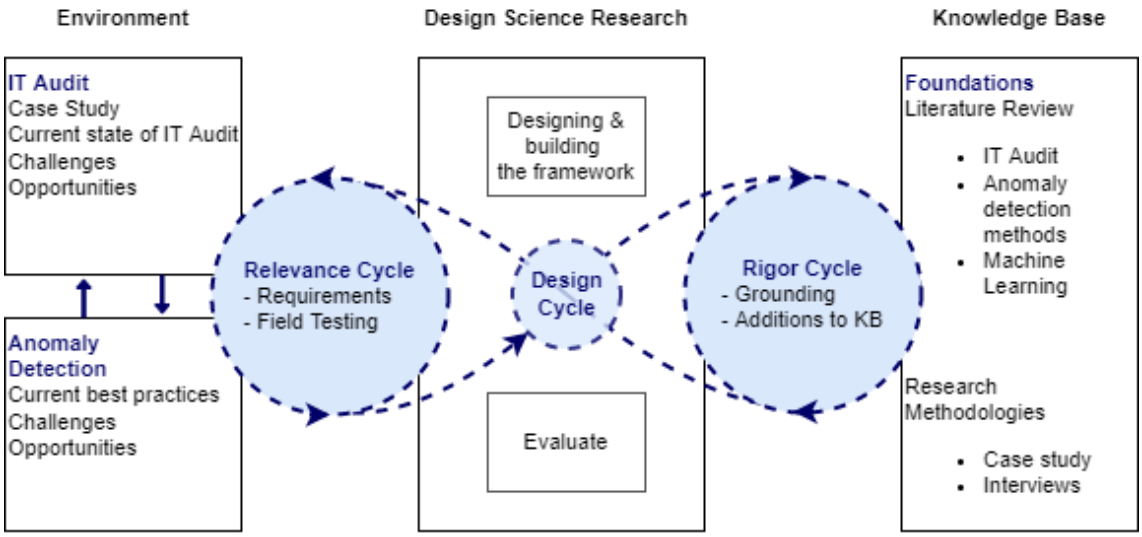

Figure 7 Adaption of Hevners Design Science Cycles

4.3 Interviews

Interviews for this research have been split up into multiple phases. To start, semi-structured exploratory interviews have been conducted with multiple IT Audit experts at EY in order to gain insights into the current situation, promising opportunities for anomaly detection, and available data. As these interviews rapidly influenced the course of the research, the snowball method was adopted so that new insights could be used in the following interviews. Based on these interviews and results from the literature review, the variables for successful implementation have been created.

Later in the research, structured interviews have been conducted among IT Audit professionals to go through the most common tasks and determine the potential for anomaly detection. For the tasks showing the most potential, the challenges have been discussed, as well as the solutions and their feasibility. The choice for structured interviews was taken based on the length and complexity of the interviews. The interviews are based on the structure described in chapter 3.2. The variables for implementation that are discussed can be found in chapter 6.2. The interview questions and the interview summaries can be found in the appendices.

5 IT AUDIT AT EY

This chapter is used to give a general overview of the current situation for IT Audit at EY.

5.1 Overview

While the theory by Hinson and Romney (as described in chapter 2.2) gives an idea of general tasks and processes that are part of IT Audit, it does not describe any task in detail. IT Audit consists of a broad variety of tasks and processes and can differ vastly between individual instances. While this variation between instances makes it difficult to describe tasks or processes in more detail, an overview of the path that IT Audit generally follows can be provided. In the case of EY, the general overview of IT Audit can be briefly described in 13 steps:

1. **Understand the entity`s use of IT:** Obtain an initial understanding of IT use and the complexity of IT at the entity.
2. **IT scoping decisions:** Decide which IT applications are relevant to the Audit and determine the preliminary strategy for each application.
3. **The risks of an entity using IT:** Determine the high-level risks for an entity using IT
4. **Determining our audit strategy for relevant IT applications:** Plan to rely on ITGCs or IT-substantive strategy
5. **Identify and understand IT processes, and identify risks:** Identify individual processes and their risks, divided into MA (Manage Access), MC (Manage Change), and MO (Manage Operations).
6. **Confirm our understanding of IT processes:** Confirm documentation, alignment with reality, and relevancy.
7. **Identify and confirm the design and implementation of ITGCs:** Test and confirm attributes of the ITGCs (only for ITGC-reliance strategy from step 4)
8. **Design IT-substantive procedures:** Design procedures to reduce audit risk to an acceptable level. (only for IT-Substantive strategy from step 4)
9. **Select ITGCs to test:** For each determined risk, select a control that addresses that risk. (only for ITGC-reliance strategy from step 4)
10. **Design tests of ITGCs:** Design tests for the selected controls (only for ITGC-reliance strategy from step 4)

11. **Execute tests of ITGCs:** Run tests, look for audit evidence and reliability, communicate any exceptions and deficiencies to the responsible party. (only for ITGC-reliance strategy from step 4)
12. **Evaluate IT processes:** Based on the tests of ITGCs it will be determined whether or not the controls provide reasonable assurance. (only for ITGC-reliance strategy from step 4)
13. **Update tests of ITGCs and IT-substantive testing for the remaining period:** When ITGCs have proven reliable, tests of ITGCs will be updated to fit significant changes, if any have occurred.

(EY Atlas, n.d.)[4]

Each of the 13 steps described above is part of the IT Audit process, and within each step, there can be a variation of tasks and processes that depend largely on the specific instance and the selected approach.

5.2 Structure Of IT Audit

It Audit at EY is divided into the following categories: Manage Access (MA); Manage Change (MC); Manage Operations (MO), and Cyber Security (CS).

5.2.1 Manage Access (MA)

MA is about ensuring integrity and confidentiality of data, and avoiding unauthorized or unintended access to systems, applications and data. Common subjects of IT Audit include the handling of access requests, admin access requests, access reviews, and Segregation of Duties (SOD) checks.

5.2.2 Manage Change (MC)

IT Audit can evaluate the current state of IT processes and applications. However, a system or application that has been deemed effective and reliable and which has effective controls in place can be impacted by unintended effects from changes and updates. MC is about safeguarding the continuity of controls. Typical parts of MC include controls and reviews of development or change to a system, changes to data, or changes to functionalities.

[4] Source from internal EY portal which is not publicly available

5.2.3 Manage Operations (MO)

MO is focused on Contingency planning (as described in IT Audit tasks by Hinson in chapter 2.2). An organization can have all of its controls and practices in place, but an event like a data center burning down can have disastrous consequences if data is stored in one place, or if no frequent backups are made. IT Audit activities in MO include a review of controls and procedureregardingut incident management, backups, and data storage locations.

5.2.4 Cyber Security (CS)

While CS is part of IT Audit at EY, due to time constraints and the relatively small portion of the work this involves, CS has been deemed out of scope for this research.

5.3 Controls Within EY

Controls at EY have been split into multiple categories: IT General Controls (ITGC); IT Application Controls (ITAC); IT Dependent Manual Controls (ITDM); Manual Controls and Substantive.

In each of these categories, a control can be preventive (to prevent something from happening), detective (with the goal of detecting when something happens), or corrective (To correct something when it happens). A control also falls under one of the categories described in chapter 5.2. Figure 8 Types and Objectives of Controls (EY Atlas, n.d.)" provides a visual represent of types of coationntrols at EY, as presented on an internal information portal.

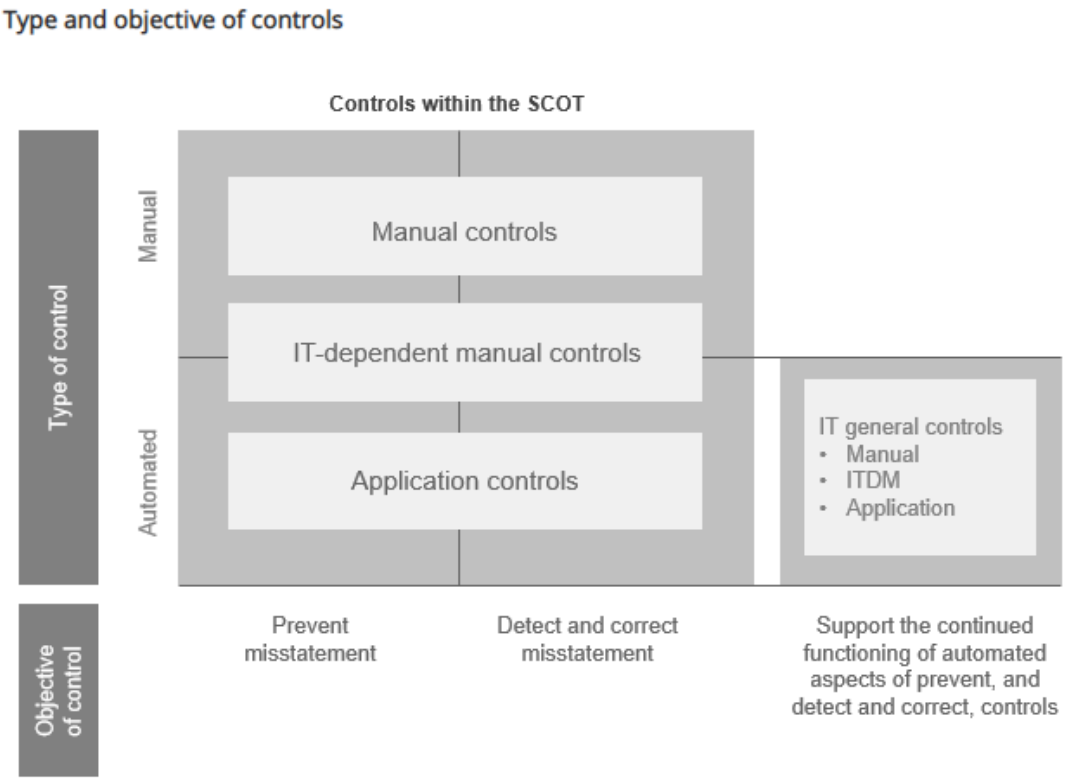

Figure 8 Types and Objectives of Controls (EY Atlas, n.d.)[5]

5.3.1 IT General Controls (ITGC)

ITGCs are controls over the general process and will usually be the first controls that will be tested for a client, as it would not matter if the applications and systems work or not when there are no effective controls on protocols and the use of IT. A test of ITGCs will happen on 5% of the instances, but with a lower limit of 5 and an upper limit of 25. Exceptions to this rule can be made, for example for days or months, where a sample size of under 5 is considered sufficient.

5.3.2 IT Application Controls (ITAC)

ITACs are controls over automated processes, to ensure that the results are reliable and effective enough. Processes where no human interaction is required (or where human interaction is only required to initiate the process) are dependent on the functioning of the system or program. ITACs are used to control that the use of the application or system (at

[5] Source from internal EY portal which is not publicly available

the time of testing) provides the expected outcomes. ITACs will be tested on the basis of a single sample.

5.3.3 IT Dependent Manual Controls (ITDM)

ITDMs are controls over processes that mostly automated, but still require human interaction at one or few points. For example, when an employee requests access, the process and changes in access rights can happen automatically, however, a manager will still have to authorize the request. ITDMs are treated almost identical to ITACs, and will similarly be tested based on a single sample.

5.3.4 Manual Controls

Manual Controls are controls on human actions and therefore are not part of IT Audit or this research.

5.4 Substantive

In rare instances, controls in an organization are considered ineffective or insufficient. In these cases, there will be a substantive test of the complete data set. This means that no samples are used for substantive tests.

5.5 Common IT Risks

The common IT risks are descriptions of the most common risks that companies try to cover with controls (see Table 3 Common IT Risks (EY, N.D.)"). The controls are created on the client side, and the clients might not describe the risks in the exact same words. In general, the IT Audit process revolves around those risks.

Table 3 Common IT Risks (EY, N.D.)

	Manage Change
1	*New IT application programs or changes to existing programs, including reports, configurations, and interfaces, do not function as described or requested because they are not adequately tested by appropriate persons.*
2	*New IT application programs or changes to the production IT application programs (including reports and interfaces) are not appropriate for the business or the IT environment.*
3	*Programs in production are not secured permitting developers to move unauthorized or untested changes into the production environment.*
4	*Configuration changes made by IT personnel are inappropriate or unauthorized.*
5	*Multiple instances of the same IT application that should be identical are not the same.*
	Manage Access
6	*Users of the IT environment aren't the intended users due to inadequate authentication and security settings.*
7	*Access rights risks:* *- Access granted to the IT environment (IT and Business) does not match the access approved* *- Access termination requests are not fulfilled timely* *- Access rights to the IT environment (IT and Business) do not remain appropriate over time.*
8	*Access requests for IT and business users of components of the IT environment are inappropriate*
9	*The access of IT users of the IT environment creates segregation of duties concerns.*
10	*Access to functions within the IT application is combined into roles. The access rights within the roles contain segregation of duties issues that could cause a material misstatement of the financial statements.*
11	*Direct data changes are made without authorization. (Of higher risk when there is routine use of direct data changes in the processing of transactions relevant to the financial statements.)*
	Manage Operations
12	*Hardware or software issues result in loss of data or the ability to accurately process that data.*
13	*Issues with programs that cannot process to completion are not addressed or are addressed inappropriately.*

6 RESULTS

6.1 Exploratory interviews Findings

This chapter summarizes the findings from the exploratory interviews which have been conducted with a number of IT Audit experts and preferably people with knowledge of AI/ML or automation on top of that. The exploratory interviews have been conducted in a semi-structured manner, where questions have been prepared beforehand but there was room for the interviewee and the interviewer to have an open discussion about the subject. As this research tries to bring two domains together in a new way, the exploratory interviews often led to new findings and directions to take the research. By utilizing the snowball method, all new information was immediately incorporated into the next interview to build further upon the findings.

6.1.1 tasks and processes

One of the first angles that this research took was to look into the upcoming role of CCM to provide a steady stream of data that could be advantageous for ML. Interview results from multiple sources (interviews A, B, E & F) quickly led to the conclusion that at the time of writing, CCM is not sufficiently implemented to form the data source it was thought to be.

Inquiring about the structure of the IT Audit processes at EY led to the conclusion that the categories MA, MC & MO are the most interesting for anomaly detection. There could be possibilities in both ToD and ToE, but the potential for ToE is deemed higher as it involves more work and more data. (Interview D)

Despite the rules and regulations around IT Audit, companies have a reasonable amount of freedom in designing and governing their controls. The risks that companies try to cover with their controls are more similar. While the risks can vary slightly per client and in description, a list with the most common risks can be created based on risks listed on GAM (Global Audit Methodology) in EY's internal Atlas platform. (Interview H) Based on the findings from the exploratory interviews, the decision was made to group tasks and processes together per common IT risk that clients wish to cover with controls. The most common IT risks have been taken from GAM (See chapter), and confirmed with multiple IT Audit managers and senior managers.

6.1.2 Data sources

Theory would suggest that more data improves the effectiveness of ML solutions. However, IT Audit work is performed on controls, and these controls are designed by the client companies, meaning that the work is not directly generalizable based on controls. Testing is performed on a sample basis, and evidence is usually requested after the samples have been randomly selected. Data is only collected for the instances that fall within the sample selection, which might be problematic for anomaly detection. The exploratory interviews provided four alternative ways of collecting sufficient data for anomaly detection:

The first suggestion is to look at instances of substantive testing. Substantive testing is performed when it is decided that the IT controls cannot be relied upon, or when the sample based test results give reason to perform more thorough testing. This approach would allow anomaly detection to be developed on-, and test its effectiveness on closed dossiers where substantive testing occurred. This way, the results could be compared to the results of the manual testing to determine the effectiveness and the potential added value for future use, without having to request more data from clients. (Interview A)

The second alternative for data collection is to combine data from multiple clients. If a process or task can be found that is performed for multiple clients, then a general algorithm could be trained on combined data, and later adjusted before use on a specific client. The difference in data that is used for each client is in theory not too big to allow this to work. Suggested categories to look for suitable processes are change management and incident management. (Interview C)

> *"A problem for IT Audit would be that you need a lot of data, and every client designs its processes slightly different. On the other hand, the Delta between these processes is not that big in general." (Interview C)*

A third alternative data source would be to look at the full data population, rather than a sample, in more instances. This approach would increase the amount of data to a point where large clients could potentially bring in enough data on a single issue to allow anomaly detection. On top of that, this could be combined with the second option and applied to multiple clients. All interview participants that were asked about this agreed that looking at the full data population instead of samples would mean a higher quality of the outcome. However, some potential downsides have been named, which are described in Paragraph **Error! Reference source not found.**.

> *"Statistically, the current approach might cover the risks. But it can be considered outdated when looking at the available technologies. When looking at the full population there is the risk that a lot of time has to be spend on analyzing what went wrong." (Interview E)*

The last potential data source is to look into CCM, despite the fact that this is currently not implemented sufficiently. CCM has good potential for the future and could replace multiple controls aonthe client side. (Interview F)

6.1.3 Benefits of Anomaly Detection

IT Audit struggles to recruit and retain enough employees to keep up with the high standards that keep increasing. If time can be saved on existing tasks by supporting the process with anomaly detection then it can allow IT Auditors to focus on complex issues. (Interview C)

Currently, IT Audit does not get as much technological developments as for example financial audit. (Interview E)

IT Audit aims to provide reasonable assurance by testing on a sample basis. If anomaly detection could enable the entire data population to be included, the degree of assurance could increase to nearly complete assurance.

6.1.4 Challenges

The interviews brought several potential challenges for the implementation of anomaly detection. Some issues might be specific to the approach or task it is used on, while others form a more general obstacle to overcome. When looking only at substantive testing, for example, this is something that is not frequent as of now.

To make a solution repeatable for multiple clients, both the IT Audit work and the data have to be similar across multiple (but not all) clients. Given that controls are designed by the clients, they are not a good basis for a repeatable task. Looking at the most common IT risks that clients try to cover with controls provides a more generalizable set of tasks. However, clients still mostly work in different systems and with different data structures.

While the interviewees agree with the notion that the quality of the IT Audit would (in most cases) improve when the full data population is considered, there are a few

downsides to this approach. The first argument against this approach is that with the current samples, "reasonable assurance" can be provided. Putting more effort into it to find more issues would not always add value, as the goal of the IT Audit is to provide reasonable assurance and not to perform the control for the client. A seconds argument against this approach is that it might translate into increased work for the clients, as they would have to deliver more data (all instances within a time period instead of just the instances from the sample), plus they would have to explain all extra findings that will result from looking at the full data population. Even though the consensus among interviewees was that testing without samples was an improvement, it requires a change of mindset for both IT Auditors and clients. Another argument is that the relationship with the client could deteriorate if more "findings" are communicated back to the client. Especially if the new findings are explainable but for example result from a lack of documentation of the event/exception. (Interview A, E & G)

> *"Say we have 250 instances, we would take a sample of 25. Based on this sample we can come to a reliable enough conclusion, however, if we look at all 250 instances, there is ten times more chance that we will find something that is going wrong. Which would also imply more work for the IT Auditors, and more work for the client." (Interview A)*

The last challenge that was brought to light in the interviews is the need for algorithm assurance and transparency. When an algorithm is used to find anomalies, the metrics behind the decisions have to be known and the outcome must be re-creatable to comply with regulations. (Interview F) Currently, when the samples bring any findings regarding the client, then the sample is extended. If more findings are made, there might be need for substantive testing. When anomaly detection is used to look at all the data, of cour,se there will be more findings. This does not fit in our current regulations, as the step to substantive testing would have to be made. The anomaly detection algorithm is unlikely to be reliable enough to replace the substantive testing, so applying anomaly detection might result in the need for substantive testing. (Interview H)

6.2 Variables For Successful Implementation

6.2.1 Data Requirements

In the book "***Anomaly Detection Principles and Algorithms"*** (Mehrotra et al., 2017) a few assumptions are described concerning data that is used for analysis purposes. The data should be part of a process, and there have to be rules, guidelines or principles behind the data, as without this, the data would be random and thus not yield any valuable insights. A second assumption is that the data should have a similar distribution over the entire data set. If any large part of data is analyzed, it should show the same patterns as any other significantly sized portion of the data.

In general, ML requires a large quantity of data. Not only does more data improve the results for ML, it also means that traditional methods would be more difficult, which presents even more motive for an ML solution. For implementation, it would be easier if the datatype is consistent for the whole dataset. While it is possible to have an algorithm include a variety of different datatypes, it would take away from the simplicity and make the development much more cumbersome.

6.2.2 Other Requirements

Other than data requirements, there are variables that impact the successful implementation of anomaly detection on which process selection can be based. These criteria have been based on the exploratory interview findings in Chapter 6.1. Given the high initial investment cost in both time and money that is required to implement an anomaly detection algorithm, there needs to be appropriate arguments to defend this investment. A process that occurs frequently will earn back this investment quicker than a process that happens irregularly. Similarly, a process that is currently time-consuming and requires a lot of manual labor, will have more to gain from anomaly detection than a relatively small task. Lastly, there is an argument to be made for the degree of accuracy that this task or process requires. Anomaly detection and ML can improve the quality of execution of a process by being able to look go through large datasets without overlooking details or making other "human" mistakes.

6.2.3 Process Selection Variables

The previous two subchapters about requirements for anomaly detection can be split up into more precise variables that can be used to determine whether or not a task or process would be likely to benefit from an automated solution like anomaly detection, and whether or not it could succeed. Table 4 displays the mentioned variables.

Table 4 Variables for Successful Anomaly Detection Implementation

JDM	Variable	Role
Person	Knowledge	Dependency
Task	Impact of misstatement	Indicates potential
	Degree of standardization	Dependency
	Complexity	Dependency
	Data quantity	Dependency
	Frequency	Indicates potential
	Duration	Indicates potential
Environment	Audit methodology	External variable
	Company size	Dependency
	Pressure	Indicates Potential & dependency
	Explainable AI	External variable
	Relationship with client	Dependency

6.2.3.1 Knowledge

The knowledge of an IT Auditor can play different roles in anomaly detection implementation depending on the approach. Anomaly detection could require knowledge of ML to fully utilize the potential of the tool. This is mostly the case if the tool would require heavy changes to the algorithm before it can be used in a new situation. Even in this situation, it does not impact the knowledge requirements for most of the IT Auditors, but merely a single person or a select few that are responsible for the changes. In most cases however, the anomalies will only be used as a starting point or an initial filter on the data,

and the final decision would still depend on the IT auditor in the same way as it does in the current situation. If the anomaly detection solution would reach a level of accuracy, reliability, and transparency on which it could perform an IT Audit almost independently, then this would mean a dramatic shift in the required knowledge. At least for the foreseeable future this is ,not a reasonable expectation, which means that the final decision will still largely depend on the knowledge and expertise of the IT Auditors about the process.

6.2.3.2 Impact of misstatement

This variable encompasses the possible impact of a misstatement or mistake in the documentation. If the goal of anomaly detection is to find more issues, then finding more issues should make a significant difference. This variable is aimed at the importance of finding anomalies in the clients' data. This relates to the question: *"How important is it that each and every mistake or misstatement is discovered in the IT Audit?"* For some risks, an anomaly could be indicative of fraud, while for others it can simply be a mistake in documentation without much consequence. A higher impact of misstatement or mistakes means more incentive to use anomaly detection.

6.2.3.3 Degree of standardization

The degree of standardization refers to the standardization of the process at the client side, and consequently the data/evidence that is received from different clients. Some common IT Risks are generally covered in the same or a similar way across most of the bigger clients. This would mean that an anomaly detection algorithm has more chance to be applicable in multiple instances. From the exploratory interviews, it can be concluded that the controls are not standardized, but the common IT risks that clients try to cover are relatively similar. Although not every risk will have the same degree of standardization across companies.

6.2.3.4 Complexity

This variable refers to the complexity of the task. If most of the tasks could be automated using an RPA solution, then developing an anomaly detection algorithm might be less desirable. On the other hand, the task needs to be simple enough to allow data to be classified as either anomalous or normal. ML is ideal to find complex patterns in data that are

difficult for people to see or put into words. Data complexity in ML is further explained in chapter 2.6.1.

6.2.3.5 Data quantity

ML needs data to learn. Depending on the method, anomaly detection can require examples of normal data and anomalous data. A high data quantity is thus a requirement for anomaly detection.

6.2.3.6 Frequency

Tasks that occur more frequently can benefit from an anomaly detection solution more frequently. Which results in more potential value for the solution.

6.2.3.7 Duration

Tasks with a longer duration have more potential for anomaly detection as there is more time to be gained. Anomaly detection could make the work more efficient by reducing the amount of manual work that has to be done.

6.2.3.8 Audit Methodology & Laws

As described in the exploratory interview findings in chapter 6.1.4, one of the challenges of anomaly detection in IT Audit is that it does not fit the current IT Audit methodology. If findings are made during the test of samples, the test will be extended, and if more findings are made then substantive testing might be necessary. More findings as a result of anomaly detection would in the current methodology require more testing to be done.

Currently, samples have to be taken randomly. Using anomaly detection as a first filter, where only anomalous results are tested instead of a random sample does not comply with the current regulations.

6.2.3.9 Company size

When looking at an anomaly detection that functions for a single client, or requires heavy changes before functioning for another client, the size of the client will start to play a role. Bigger clients will not only bring the necessary amount of data more easily but also usually require bigger time investments that could justify the use of anomaly detection.

6.2.3.10 Time pressure

Time pressure is the value that clients put into the time spent on the Audit of a certain IT risk. This is closely related to the impact of misstatement variable, as clients are likely to allow more time to be spent on a risk that can have a large impact on the organization. High time pressure could make it more difficult to convince clients to change the process, deliver more data, and spend more time on explaining anomalies. At the same time, high time pressure can be a good argument for anomaly detection if it could make the process more efficient.

6.2.3.11 Explainable AI

In IT Audit it is important that all decisions and actions are explainable and repeatable. Chapter 2.9 describes the different ways of achieving transparency.

6.2.3.12 Relationship with client

If anomaly detection could negatively impact the relationship with the client, this would be an argument against the implementation. Several interviewees indicated that the relationship with the client could deteriorate for several different reasons. First, clients currently provide evidence after it is requested by the IT Auditor, and provide only the data that is necessary. Requesting more data could result in more work for the client, depending on the way the data is structured and how the IT Auditor can access it. The second is that clients are getting reasonable assurance with the current approach and might *not always* appreciate too many findings being communicated back to them, even if they are deserved.

6.3 Conceptual Model

The conceptual model is based on the JDM-framework as described in paragraph 3.1. The three categories that make up the JDM framework are person, task, and environment. Research by Charyyeva (2017) has determined variables that influence JDM in IT Audit (see Table 2 Factors that affect JDM in IT Audit (Charyyeva, 2017)"). The variables by Charyyeva have been taken into consideration while creating variables for successful anomaly detection implementation, for the reason that this connects the research by Bonner with the IT Audit domain.

The JDM framework by Bonner (1999) and the adaption to IT Audit by Charyyeva (2017) have determined factors that influence successful decision-making in IT Audit. To repurpose this model for anomaly detection implementation, the variables that

influence JDM in IT Audit can be compared with variables that influence successful anomaly detection implementation. Anomaly detection can be compared to decision-making by looking at its role in the process. Algorithms for anomaly detection will make a "decision" on whether or not something is anomalous. Besides determining what is anomalous and what is not, the algorithm will most likely not make any final decisions as it would be difficult to get an outcome that is reliable enough to comply with regulations. Which makes the algorithm nothing more than a tool to support decision-making by the IT Auditor. This leaves two alternatives to utilize the JDM framework: (1) Look at how different variables affect successful anomaly detection implementation. (2) Add variables for successful anomaly detection to the framework and look at how anomaly detection affects IT Audit decision-making. In this case, the algorithm is regarded as the subject making the decisions.

This research focuses on the potential of anomaly detection development and whether or not it is feasible, rather than the use of an anomaly detection solution when it is working. Theories and models like TAM, TAM2, or UTAUT could be used to research the acceptanceofn anomaly detection and have been considered for this research. As measuring technology acceptance is not the main goal of this research, it was decided that the use of these models does not fit in the research scope.

The conceptual model used in this research is based on the JDM framework by Bonner (1999) (chapter 3), and takes into account the variables from the research by Charyyeva (2017) to connect the model to IT Audit. The variables from paragraph 6.2.3 have been put in a visual representation of the framework in Figure 9 Conceptual Model". As this research is attempting to combine two domains in a new way, the models that are used are not intended for this purpose. Models have been adapted to fit to the circumstances.

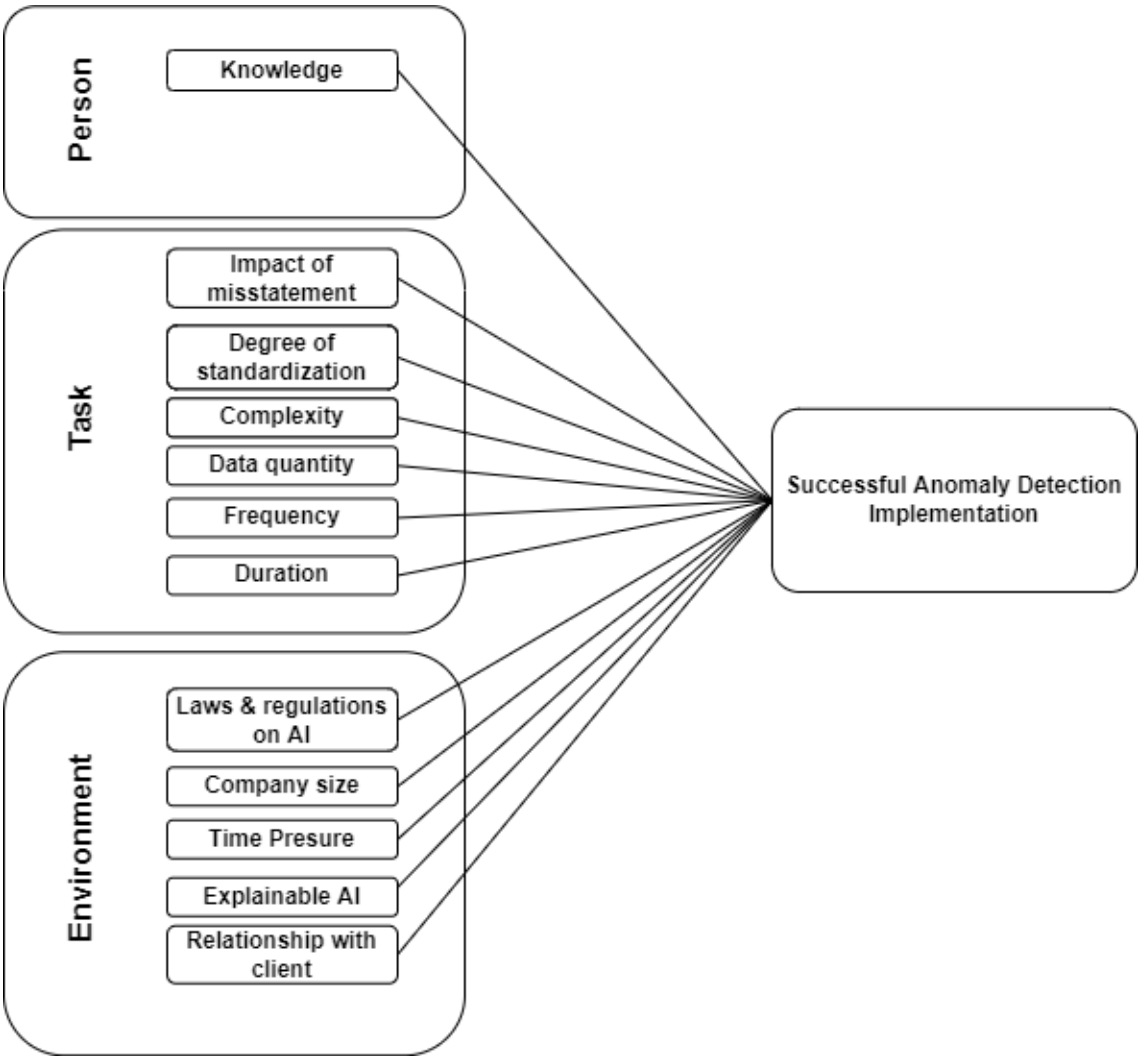

Figure 9 Conceptual Model

6.3.1 Variables For Potential

The potential of anomaly detection on a given task depends on only a few of the determined variables. The variables "impact of misstatement", "duration", "frequency", and "pressure" can be used to determine the potential of a task or process. Tasks with higher duration and frequency have more to gain from anomaly detection than tasks that do not require as much time. The impact of misstatements indicates the added value of the increasing amount of findings in an IT Audit.

6.3.2 Task Dependencies

Some variables can form obstacles to successful anomaly detection implementation, depending on the task or process, and the data that comes with it. The degree of standardization, complexity, and data quantity that come with a certain task can have a big impact

on the chances of success. From the environment category, the company size and relationship with the client can make either a significant or an insignificant impact ton he implementation depending on the approach, task, and client.

6.4 Approaches for Task Selection

The exploratory interviews provided a few distinct approaches to the implementation of anomaly detection, and they are not completely mutually exclusive. Several choices have to be made before settling on an approach. Those choices can be supported by looking at the variables from chapter 6.2.3, and analyzing how they would affect the process.

6.4.1 Multiple clients

In an ideal world, one solution would work for many different clients without too many adaptions. Whether or not this is possible depends on a few variables like the degree of standardization. On the other hand, a solution that only works for a single client or needs many changes to work for a different client will only be reasonable if the client brings enough data, and the solution can bring enough value.

The first choice is between a solution for a single client or multiple clients. One solution for multiple clients is more desirable as it increases the repeatability and utility of the solution. However, a solution that works for multiple clients requires a higher degree of standardization of the process and data. A solution that is designed for a single client or that would require a large number of adjustments for each client would depend on other variables. A high frequency and duration of the task would be required, as well as a high data quantity for the specific task for the given client. The variables “company size” and “time pressure” could also influence the potential for anomaly detection for a single client. Figure 10 Variable Requirements Per Approach” gives an overview of the variables that impact the decision.

6.4.2 Machine Learning Method

A second choice to make would be the choice between ML algorithms. In the literature review, different methods have been provided. While no method can be called the best in all situations, certain factors can help to make a decision. To narrow down the choice between algorithms, the choice between supervised and unsupervised learning can be made based on selected variables (See chapter 2.7). Supervised learning is generally more

accurate as the data is labeled. However it requires a ,higher degree of standardization, and ideally a lower complexity. Further choices regarding the ML method are explained in chapter 6.7.

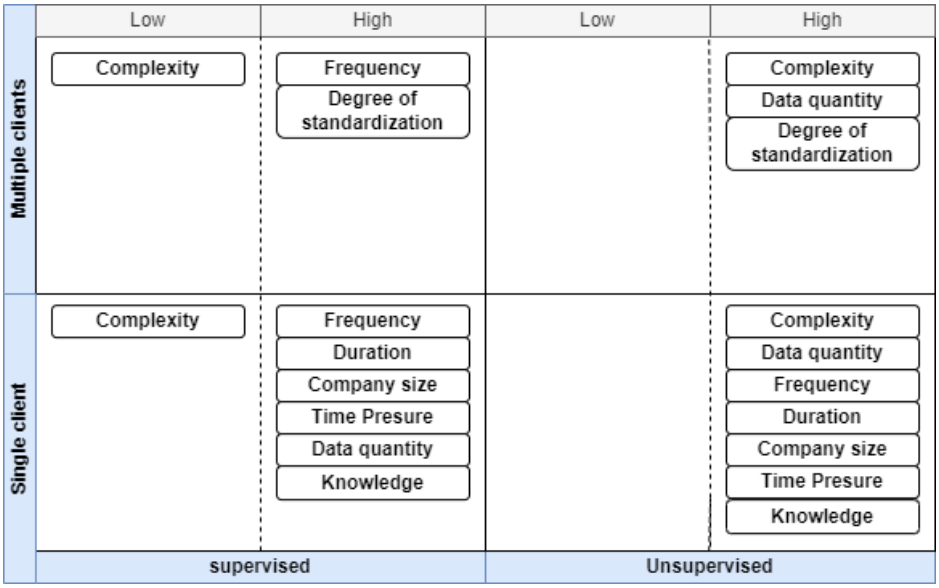

Figure 10 Variable Requirements Per Approach

6.4.3 Random Samples Vs. Anomaly Based Samples

The use of the solution will depend on the situations in which it can be applied. Currently, substantive testing is not the most common way of performing an IT Audit, but a lot of data is often involved, and a lot of time can potentially be saved. Sample-based testing does not provide enough resources for anomaly detection to work, but applying anomaly detection over the full data population instead of samples would mean that the utility of the solution is much broader than only considering the substantive testing when it would occur now. Replacing the sample-based approach in certain instances brings its own set of limitations and dependencies, however, as this would not fit within the current IT Audit methodology, and would require a higher standard of transparency/explainability.

6.4.4 Continuous Control Monitoring

CCM is currently not utilized enough within IT Audit to apply anomaly detection. Looking at the future, CCM might hold potential for clients to take the place of several controls, and give a warning when an issue occurs. Despite the potential for the future, CCM would take place on the client side and is currently not far enough to be ihe focus of this research.

6.5 Interview Results

6.5.1 Interview structure

The goal of this round of interviews can be described in five parts: (1) confirm variables, conceptual model, and risks. (2) Determine IT risks with the highest potential for anomaly detection. (3) Determine the variables that hold back anomaly detection implementation for each risk (dependencies). (4) Find solutions for the dependencies for each risk (5) Confirm the feasibility of the solutions. Figure 11 Interview Structure Based On Banner (1999)" shows a visual representation of the structure.

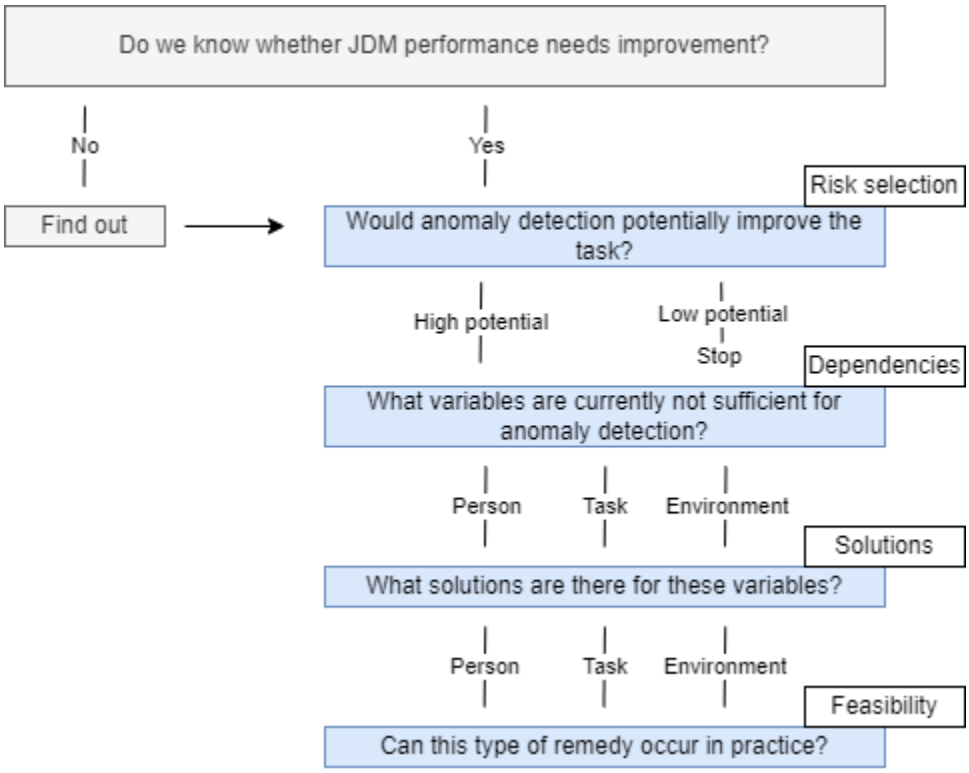

Figure 11 Interview Structure Based On Banner (1999)

The five steps described above are based on the framework for JDM research in accounting (chapter 3.2). The interviews have been conducted in several parts, as discussing all the steps in depth involves too much information to fit in a single interview. The interview questions can be found in Appendix III: Interview Questions.

The risks that show the highest potential in the first phase of the interview have been discussed more in-depth. For each chosen risk, questions have been asked to determine the dependencies (variables that can form a challenge for successful implementation).

After the dependencies have been determined for a risk, questions have been asked about possible solutions. The final step for each risk is to look at the feasibility of the solutions.

6.5.2 Confirmation of Findings

The first step of the interview involved a presentation of the variables that influence anomaly detection implementation, the conceptual model, and the most common IT Risks provided by GAM. After discussing the subject with the participating interviewees, minor adjustments have been made to the structure and naming of the conceptual model and some variables.

6.5.3 Risk Selection

The first part of the interviews included the confirmation of the findings as described above and an extensive discussion about the risks that offer the highest potential for anomaly detection. Given the time required to discuss a risk iin-depth the variables determined in chapter 6.3.1 have been used as a first filter. Based on these four variables, the interviewees have ranked the IT risks (chapter 5.5) in order of potential. The findings of these discussions can be found in Interview I1 and Interview J1 in the appendices. The appendices include a summary of the rankings in tables, which can be found in Table 6 Potential Per Risk (interview I)" and Table 7 Potential Per Risk (interview J)". The risks that show the most potential according to the interview will be researched in more depth in the following rounds of the interview.

The first conclusion from this part of the interviews is that some risks can be merged. The risks in question are named separately in the literature as there are differences between them, however the IT Audit work surrounding these risks is combined in a single process. Risks 1, 2, and 4 from MC have been indicated to all belong to the same process. *"For risk number 4, the same logic can be followed as for risk 1 and 2, as they are reviewed together." (Interview J1)*

Risks 1, 2, and 4 have been discussed in the following rounds of the interview as a single group (Interview I3 & Interview J2). The other risks that have been indicated as possessing the highest potentialare: risk 9 (Interview I2), risk 7 (Interview I4 & Interview J4), rand isk 12 (Interview J3).

6.5.4 Risks 1, 2 & 4

Risk description:

1. New IT application programs or changes to existing programs, including reports, configurations, and interfaces, do not function as described or requested because they are not adequately tested by appropriate persons.

2. New IT application programs or changes to the production IT application programs (including reports and interfaces) are not appropriate for the business or the IT environment.

4. Configuration changes made by IT personnel are inappropriate or unauthorized

6.5.4.1 Dependencies

The tasks related to this risk are highly standardized when compared to other risks. Clients often work in systems SAP of Microsoft Dynamics. These systems have standardized modules with pre-defined data fields, which makes both the task and the data structure more standardized. A problem with standardization would be in the use of emails or ticketing systems that clients often use for part of the process.

> *"The difference can be that small companies might use email for requests and confirmation while bigger clients typically use a ticketing system like ServiceNow." (Interview I3)*

The tickets and emails in the process make the task rather complex. RPA could potentially be used to automate the first few steps, but not much more. While this leaves room for anomaly detection to add value, the lack of structure in the tickets and emails across clients could make it difficult.

The data quantity connected to this risk is high when compared to other risks. All companies keep changelogs, and especially bigger companies can have thousands of changes in a year. The high number of changes that can occur also mean that the current approach using samples covers only a fraction of the data, while the importance of discovering mistakes has been deemed high. This is a good argument to use anomaly detection on the full data population, and then analyze the anomalies in more detail.

> *"For example, in half a year a client makes 800 changes. In this example we might test around 18 samples over this half-year period, at the end of the year, we will have tested 25. The chance that you will*

spot the changes that were not approved properly or where the testing evidence was not documented, is really small." (Interview I3)

The relationship with the client might be an issue. While more insights will often be useful for the client, and the change manager would be happy to know what is going on, there is also a political game that is being played.

6.5.4.2 Solutions

The standardization issue with the tickets and emails could be partially resolved if clients would be willing to use a standard template and unique identifier for the change. For the ticketing systems, this is rather standard and does not require changes. Changing the email communication is a simple change that many clients are likely willing to make.

The complexity challenge comes from the emails and tickets as well. NLP has been suggested as a means to analyze and categorize the emails. However, this is a complex solution and would not be desirable for a first-time implementation of anomaly detection. The positive side is that not all clients require the contents of an email or ticket to be analyzed. Simply the subject of the mail or the type of ticket could indicate what the communication concludes. Clients, where the process is designed like this, could be a good first target, while NLP could be useful in a later stage.

To gather enough data, this risk is standardized enough to combine data from multiple clients. Even when considering just a single large client, there is so much data and so many hours spent on the IT Audit of this risk that it could be worth developing an anomaly detection solution. There are already RPA tools available that can support the data collection from different systems.

The relationship with the client could be affected if too many insignificant findings are relayed to the client. Insignificant, explainable mistakes in the documentation are not interesting for the client but require extra work to analyze and explain.

6.5.4.3 Feasibility

The proposed solutions are deemed feasible. However, the current It Audit methodology might still be a limiting factor. Samples have to be random, and replacing random samples with a selection based on anomaly detection would require a high explainability and a change in regulations.

6.5.5 Risk 7

Risk description:

Access Right Risks:

- Access granted to the IT environment (IT and Business) does not match the access approved

- Access termination requests are not fulfilled timely

- Access rights to the IT environment (IT and Business) do not remain appropriate over time.

6.5.5.1 Dependencies

The tasks related to this risk are less standardized than for most risks, as it is very common for the process to occur mostly via email. Although some clients use specific systems for this. The data that is used is mostly email interactions and some Excel extracts. There are more structured rules for this risk, which would allow an RPA solution to support with a bigger portion of the work. The main problem is the structure of the data in emails. Anomaly detection for this risk is not interesting for smaller clients. For bigger clients there is often a system with all employees, roles and functions. The quantity of data is likely still smaller than with other risks. The high likelihood of mistakes in this process makes it an interesting target to look at the full data population rather than the samples.

6.5.5.2 Solutions

The tasks and data can be more standardized by requesting clients to work with email templates when this is not yet the case. As this is a rather small change, many are likely to comply. Outlook has simple features that enable certain automated actions based on email content, which could help with documentation if all emails in the process have to be accessible.

6.5.5.3 Feasibility

This risk is regarded as important by clients, which make them more likely to make changes in order to fit the new requirements that come with anomaly detection. There are some arguments against the use of anomaly detection however. The relatively low complexity and data quantity would allow a simpler solution (like RPA) to function. While anomaly detection might be possible for this risk, it is likely better to look for other solutions first.

6.5.6 Risk 9

Risk description:

The access of IT users of the IT environment creates segregation of duties concerns.

6.5.6.1 Dependencies

The tasks related to this risk are often performed in SAP GRC by bigger clients. This module makes the task and data rather standardized among clients that use it. Small companies might not use this module. The classification of profiles that are considered critical or important is determined by the client, which might make the data slightly different per client. The data is usually delivered in Excel files.

The task is deemed too complex for RPA, as there are indicators that can for example indicate fraud. Those indicators are not easy to figure out and program into an RPA solution. ML happens to be strong at finding complex patterns, which might allow anomaly detection to add value.

The data quantity connected to this risk is already significant when only considering samples. Looking at substantive testing or the full data population would increase this even more. Requesting more data should not be a problem for the relationship with the client, as this data is well documented in the systems and thus requires minimal extra effort.

6.5.6.2 Solutions

The only point against anomaly detection implementation is that it could be argued that the tasks related to it are not frequent, as this is usually performed once per year. The infrequency of the task can be partly balanced out by the high level of standardization that could allow multiple of the bigger clients to be supported by a single solution with minimal tweaks required. On top of that, tasks related to this risk are performed for practically every client in each iteration. (Interview I1)

6.5.6.3 Feasibility

The high impact of issues that can be detected make it more likely that clients are willing to participate in new methods. There is a high degree of standardization and enough data available. The complexity of the task makes it unlikely to be solved with a solution like

RPA, which leaves potential for ML, which happens to excel at discovering complex patterns.

6.5.7 Risk 12

Risk description:
Hardware or software issues result in loss of data or the ability to accurately process that data.

6.5.7.1 Dependencies

Tasks related to this risk happen in relatively standardized systems. Data is often in the form of emails, tickets, and reports about the backups. The emails and tickets might form a problem for anomaly detection. The report of the backup could be automated with RPA, but the rest of the process would likely be too complex.

According to interview J3, testing over the full data population will not have an impact on the quality as obvious as it is for other risk.

Addition data requests/extractions would not pose a threat for the relationship with the client. And the high impact of findings for this risks make it less likely that a client would be bothered by increased findings.

6.5.7.2 Solutions

Clients are likely willing to make changes to their processes to comply with requests. For example in the form of more structure in the email or ticket communication. If the data can be standardized enough then the complexity of the decision-making should not form an issue.

6.5.7.3 Feasibility

An advantage for this risk is that companies care highly about it, and are consequently more likely to adapt if it can improve the findings. The emails and tickets that are part of the evidence might form a problem for data collection and complexity.

6.6 Use-Cases

This chapter describes the use-cases for ML-based anomaly detection in IT Audit that show the highest potential and feasibility. The results from chapter 6.5 have led to two use-cases, based on risk group 1, 2 & 4, and risk 9. Risk 7 has shown less potential and has been left out of the use-cases as a result. While risk 12 shows potential, a lot of the evidence is recorded in email and tickets, which makes it a difficult task to start on.

6.6.1 Use-Case 1: MC Anomaly Based Sampling

6.6.1.1 Goal

The first use-case is based on common IT risks regarding Manage Change. The goal is to create an algorithm that can detect anomalous data entries that may indicate that changes to a system (1) have not been properly tested by appropriate people, (2) are not functioning correctly, (3) are not appropriate, (4) configuration changes are inappropriate or unauthorized. The anomalies found can be used to create a list of changes that will be part of the IT Audit, to replace the random sample that is currently the common approach. This ensures the same quality of IT Audit over each change, but with a more targeted approach to changes where findings are most likely.

6.6.1.2 Data Collection

The most desirable outcome is a model that can be trained on multiple clients, to create repeatable value. The standardized systems and modules that are used by clients make this a possibility. If this proves impossible during data collection and preparation or during the training/testing of the algorithm, then data from a single large client can be considered as a simpler yet less effective alternative. The amount of data for MC at a single large client appears to be sufficient to consider this as an alternative.

The clients for the pilot of this use-case should preferably work in the same system (SAP) for this process to keep data as uniform as possible. A good starting point would be to look at clients where the **contents** of tickets/emails are not important to the process. Looking through written text would require NLP which would add to the complexity of the solution, and should be avoided for the pilot.

6.6.1.3 Challenges

A challenge for this use-case would be the tickets and emails that are usually part of the process to approve of changes. As suggested in the data collection part of this use-case, this can be circumvented by starting with companies where the written contents of the emails/tickets are not impactful on the process.

The biggest challenge lies in the audit methodology. Samples have to be random and cannot usually be influenced by the auditor. A change in methodology and mentality is required to make this use-case usable in practice, even if it proves effective.

If an algorithm turns out to be effective and it is applied to the bigger clients, then a high number of anomalies might be found. Imagine a data set with 10.000 changes for one client. 200 anomalies might be found, out of which most can probably be explained. 200 is still significantly more work than the 25 samples which is now the upper limit. Even if the anomalies can be ranked on how anomalous they are, testing the 25 most anomalies entries might result in many findings. This raises several questionsaboutf decision-making based on these findings. Are 20 mistakes out of 10.000 changes cause for concern? For a client it might be interesting to know what is going wrong, but what does it mean for the outcome of the IT Audit?

The last challenge is about the explainability of the algorithm. For an audit, every choice has to be explainable, and the algorithm results have to be reproducible. As ML algorithms can continuously improve over time, using a single starting state of the algorithm for every time it is used would potentially hinder its growth and effectiveness. An alternative would be to save a copy of the version of the algorithm that is used for each audit so that it can be used for reference and reproduction of the audit if necessary.

6.6.2 Use-Case 2: MA: SoD concerns and fraud detection

6.6.2.1 Goal

The second use-case is based on SoD concerns resulting from access to the IT environment, which comes from IT risk 9. The goal is to look at collective anomalies to see if patterns exist that might indicate fraud, and look for other SoD concerns. Fraud is often hidden by multiple small transactions that are individually no reason for concern. In financial audit, anomaly detection is already being used to detect fraud by looking at collective or group anomalies in transactions (Chalapathy & Chawla, 2019). In IT Audit, the

authorizations of individuals to IT environments, and the function roles they have, can potentially point towards fraud. Requests, authorizations, or transactions are often made by a small number of people and could be identified by role. (Interview I2) A step further would be to include financial audit data to look for patterns between authorizations and transactions.

6.6.2.2 Data Collection

The large amount of data connected to SoD concerns, and the standardized systems and data structures that are commonly used make this a promising target for anomaly detection. The initial data collection should ideally be from clients that work in the same system. SAP GRC has been indicated as a commonly used module among big clients. Since most of the required data can easily be extracted from the system in Excel files, the data collection is likely not a challenge for this use-case.

6.6.2.3 Challenges

The audit methodology regarding the number of findings is a general challenge among the use-cases. As explained in the challenges for use-case 1, more findings in the IT Audit would in the current methodology result in implications. Another general challenge across the use-cases is the explainability of the algorithm. Decisions in IT audit have to be explainable, and the audit has to be reproducible.

A challenge that is unique to this use-case lies in the combination of SoD data with data regarding financial transactions to detect possible fraud. The potential of fraud detection makes this a desirable solution. However, the financial audit and IT Audit are separated, and combining the two would require a collaboration. The source of the potential for this use-case also forms its weakness.

6.7 Next Steps For Implementation

While the use-cases that have been created and described in the previous chapter still have a few obstacles to overcome, the first few steps towards implementation can already be taken.

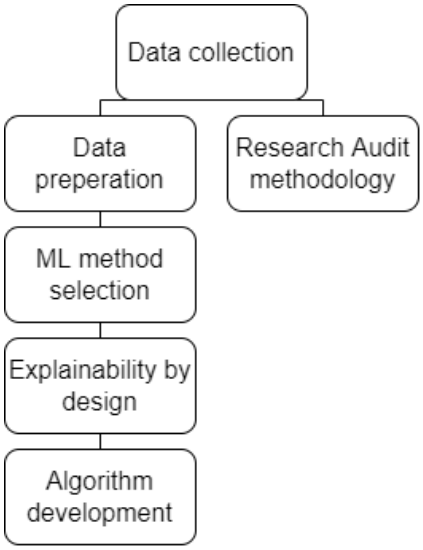

Figure 12 Next Steps For Anomaly Detection Implementation

6.7.1 Data Collection

The first choice for data collection for both use-cases would be to look at big clients that work in similar, commonly used systems so that data from multiple clients can be used in the pilot. However, bigger clients working in highly standardized systems are less likely to be the subject of substantive testing under the current approach, which means that access to the data would have to be requested. If this is not possible, the alternative is to look for closed dossiers where substantive testing has been performed. The advantages of this approach iarethat the outcome of the anomaly detection can be compared to the findings that have originally been done, no additional data has to be requested, and no damage can be done to the relationship with the client. The downsides to using closed dossiers where substantive testing has been performed is that substantive testing in the current approach usually does not provide ideal circumstances for anomaly detection, as it is performed on exceptional cases. Anomaly detection would likely see better performance in situations with more standardized data, structure ,and controls. At least for testing purpose,s this would be a good target.

6.7.2 Data Preparation

The collected data can be prepared by following the seven steps provided by Bonaccorso (2018), which can be found in chapter 2.8. The data preparation step is important because it provides valuable information that is required for the ML methods selection.

6.7.3 Machine Learning Method Selection

After the data preparation step, several decisions can be made to narrow down the ML method selection. The first decision is between supervised, unsupervised, semi-supervised, or hybrid methods. As described in chapter 2.7.1, the best results can generally be achieved with supervised learning. Obtaining labels for anomalous data however, can b,e problematic, as anomalies are by default uncommon. As explained in chapter 2.7.1, Chalapathy and Chawla recommend deep anomaly detection methods, and suggest he use of unsupervised or semi-supervised methods like different types of autoencoders or RNNs. A condition to DL is that it performs better on high-dimensional data, which is why the decision for the method is made after data collection and data preparation. The choice for unsupervised methods like AEs is further supported by research from Schreyer et al. (2018), where the use of deep autoencoder neural networks is advocated.

6.7.4 Explainability

Explainability of the algorithm will be required before use in IT audit can be realized. As ML algorithm can improve over time, it is not recommended to limit the use of the algorithm to a single version that can be used for every Audit. Additionally, it is likely required to partly retrain the model for use on each client. In order to still allow for the reproduction of the audit, a copy of the version used can be saved. Chapter 2.9 provides information about how transparency of AI can be achieved through different means.

6.7.5 Research Audit Methodology

A general challenge for anomaly detection in IT Audit is that the current methodology does not support the use of anomaly detection in the way this research suggests. However, research into the implications of audit methodologies would warrant a complete thesis in and of itself. While a complete solution to the issue cannot be provided in this research due to time and scope limitations, the reresearcher'spinion (based on the results and interactions with IT Audit professionals) can be provided.

Current IT Audits aim to provide reasonable assurance that IT controls are effective. This is done using random samples, that have a statistically sound basis. The argument against testing the full data population (and thus anomaly detection as suggested in this research) is that the current method provides reasonable assurance while fitting in the financial and time restraints. Testing the full data population is not reasonable with the currently available tools, and more findings might result in an unreasonable increase in audit time and budget. What these arguments fail to consider is the added value that anomaly detection could bring to the IT Audit in new information, the quality of the audit, and the correlation between features.

Even if anomaly detection would be able to provide concrete results, it would be impossible to use the new findings within the current methodology. Using anomaly detection on the full data population would often result in aa numberof anomalies that is much larger than the current sample would be. Manually going through all anomalous instances could increase the amount of work and the audit budget by an unreasonable amount. Reaching a conclusion based on the anomaly detection results alone would require a level of accuracy and reliability that is unlikely to be reached. Replacing the current random sample by a sample from the most anomalous entries would result in a solution that works in theory. The sample size does not increase, so the amount of work and required budget stays the same, but the quality will improve as erroneous data entries will be more likely to be in the sample. In practice, however, the findings would lead to an extension of the samples to see if more is found, or likely even result in a negative outcome of the audit.

Testing all anomalies might result in a temporary increase in findings, and thus an increase in audit budget and time. On the other hand, the insights that this provides to the auditor and the auditee could be used to reduce the number of anomalies that will be found in the following audits. Over time this could stabilize and provide an environment where the full data population can be tested, drastically fewer erroneous instances slip through the audit, and the time and budget limits are not exceeded. The insights mentioned referring to the auditee that will adapt its processes and documentation to the new standards, and the auditor who can adapt the algorithm. For example by recognizing categories of anomalies that can speed up the manual audit work by more specifically indicating what is causing the anomaly.

7 DISCUSSION

7.1 Conclusions

This research managed to answer the research questions. The answers to the questions and other conclusions made during the research are summed up in this chapter.

The most prominent techniques for anomaly detection depend on the context and the available input data. Supervised methods have better overall results, but require training data on both classes of events, and anomalies are uncommon by definition. Unsupervised and semi-supervised methods are more common, and recent research supports the use of autoencoders. Specifically, deep autoencoder neural networks are recommended based on their ability to find complex patterns in unlabeled data, without the need for anomalies in the training data. While DAENNs are the general recommendation, the final choice depends on task and data requirements for the method, and task and data availability of the processes. It can be concluded that the ML method should be chosen after the task is selected (based on potential and feasibility), and the data has been collected, as the available data dictates the possibilities.

Tasks in IT Audit are difficult to generalize as a consequence of varying IT controls. Common IT risks can be used to group common tasks together across multiple clients. Based on the potential impact of increased audit findings, frequency and duration of the task, and pressure from clients, the IT risks that would benefit most from anomaly detection have been determined to be: (1) a combination of risks 1, 2 & 4 from manage change. (2) Risk 7 from manage access. (3) Risk 9 from manage access. (4) Risk 12 from manage operations. The challenges for anomaly detection concerning these risk, and the feasibility of solutions have been researched by formulating variables for successful implementation. The variables have been classified as indicators for potential, requirements for success, or general dependencies. This research led to the conclusion that for two of the aforementioned risks, the chance of success is higher, leading to the formulation of two use-cases. The first use-case aims to use anomaly detection to detect multiple manage change risks, by looking at the full data population of changes at big clients working in standardized systems. The second use-case aims to discover SoD concerns, and could be combined with financial audit data to discover fraud.

The biggest challenge for anomaly detection in IT Audit does not lie in getting useful results, but rather in the required changes to the methodology and mindset before it can

be used in practice. The current approach is based on random samples and offers reasonable assurance on a statical basis. This sample approach is based on the notion that testing the full data population would not be possible while remaining within time and budget norms. New techniques, such as anomaly detection, might mean this notion is outdated, but the methods cannot be created and optimized due to the current restraints.

The main research question is defined as follows: *"How can anomaly detection be applied to the IT Auditing process to make a significant contribution to the efficiency and/or quality?"* Making the final IT Audit decision rely solely on anomaly detection without human confirmation is unlikely, as it requires professional judgment, high reliability, cand omplete explainability, and it does not conform with audit regulations and methodology. That being said, anomaly detection can make a significant contribution to IT Audit by playing a supporting role.

The described use-cases for anomaly detection would be able to make a positive impact on the IT Audit quality. The steps that have to be taken before implementation involve two parts. On one side are the practical steps to a functioning solution. The data has to be collected and prepared, a method has to be selected and an algorithm has to be developed. On the other side is the audit methodology that currently does not suit the use of anomaly detection. Additional research is required to see if the methodology can be adapted to the use of new technologies. At the moment, Audit is based on providing reasonable assurance based on a mathematical function behind the samples that have to be taken. This methodology is based on the notion that testing the full data population does not fit time and budget restrictions.

The practical side of applying anomaly detection in IT Audit can be concluded to be feasible. For both use-cases, the data is available and standardized enough among some of the bigger clients. In the current context, it would already be possible to add value to the audit. The methodology is what is holding back the success for now, and requires additional research.

7.2 Limitations

By attempting to introduce anomaly detection in a new domain, this research has been unable to rely fully on available models and theories. Adaptations to models and theories had to be made in the research, which means that effectiveness has not been proven, and selected models were not always intended for the exact purpose they have been used for.

As is often the case with qualitative research, tresearcher'sh's personal op,inion and beliefs could have an unintended effect on the outcome.

Time constraints have limited both the scope of this research, and the possibility to go further in-depth into the two research domains. The time constraint also resulted in a limited number of interviews, which means the research quality relies partly on the expertise and knowledge of the participants. To combat this limitation, interviews have been conducted with employees with different functions, mostly highly experienced in their domain, and when possible with knowledge of both IT Audit and AI/ML or other relevant experience.

The thesis is written during an internship at EY in the Netherlands and contains many terms and methods that are exclusive to EY. Consequently, the results of this research are limited in generalizability. That being said, the research isbuiltd to be applicable in as many situations as possible. The variables for successful implementation apply to any IT Auditor. The common IT risks, while based on information from EY, are relevant in any situation. The use-cases and the suggested steps for implementation should by large be generalizable. As the results are based on interviews that have been conducted exclusively within EY, there is no guarantee that the findings would be the same when the research is reproduced in a different setting.

7.3 Future Research

This research attempted to find ways in which anomaly detection could offer value to IT Audit, looked at the challenges for the tasks with the highest potential, researched solutions to overcome these challenges, and discussed the feasibility. The final chapters of this research leave a clear opportunity for future research in two distinct directions. The first possibility for future research lies in the two use-cases that have been created, which could be used as starting point for research to initiate development and test practical effectiveness. The second direction for future research is to look at the challenges in the audit methodology that would prevent solutions from being used in real-life situations.

The third option for future research is less obvious from the research conclusions. The focus of this research was on the use of anomaly detection in external audits. Some of the challenges regarding the audit methodology might not apply or might be less relevant when the focus is shifted towards internal audit.

7.4 Relevance

7.4.1 Business Relevance

The intent of this research has a clear business relevance in the sense that it attempts to provide a solution for a practical issue. IT Audit is a constantly changing domain, with a rapid increase in complexity in the form of new standards, and a rapid increase in workload as the role of IT systems for companies is growing. Anomaly detection has proven to contribute to other, similar domains such as financial audit. Extending this principle to IT Audit could allow for a higher degree of assurance.

The contribution that this research delivered toward this goal, is that the eendpointof this research serves as a starting point for future research and anomaly detection implementation. The provided use-cases can be tested to gauge the practical effectiveness and the potential. Continuing in the direction of the suggested use-cases could provide support in determining erroneous entries in MC audits, and could help monitor SoD concerns and potentially detect fraud.

7.4.2 Scientific Relevance

The gap in scientific research that this research tries to fill is in the application of anomaly detection in the domain of IT Audit. Anomaly detection has been applied in many domains, but the exceptional variation in audit evidence and the difference in IT controls between companies make this a tough subject. Not much information can currently be found regarding the initiation of anomaly detection, or the potential it offers for It Audit.

The contribution this research made to the scientific knowledge base can be found in the variables that influence successful implementation in this domain, the multiple approaches that have been considered, and the tasks that show the most potential.

REFERENCES

AXELOS Limited. (2011). *ITIL® glossary and abbreviations* (1.0). https://www.axelos.com/corporate/media/files/glossaries/itil_2011_glossary_gb-v1-0.pdf

Benbasat, I., Goldstein, D. K. and Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly*

Bengio, Y. (2009). *Learning deep architectures for AI*. Now Publishers Inc.

Blokdijk, J. H. (2004). Tests of Control in the Audit Risk Model: Effective? Efficient? *International Journal of Auditing*, *8*(2), 185–194. https://doi.org/10.1111/j.1099-1123.2004.00089.x

Bonaccorso, G. (2018). *Machine Learning Algorithms: Popular algorithms for data science and machine learning, 2nd Edition* (2nd Revised edition). Packt Publishing.

Bonner, S. E. (1999). Judgment and Decision-Making Research in Accounting. *Accounting Horizons*, *13*(4), 385–398. https://doi.org/10.2308/acch.1999.13.4.385

Brennan, G., & Teeter, R. A. (2010). Aiding the Audit: Using the IT Audit as a Springboard for Continuous Controls Monitoring. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1668743

Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*.

Chandola, V., Banerjee, A., & Kumar, V. (2007, August). Anomaly Detection : A Survey. https://www.researchgate.net/publication/220565847_Anomaly_Detection_A_Survey

Charyyeva, M. (2017, May). *Professional Judgment and Decision-making in assessment of highly customized IT*. Mahri Charyyeva.

EY. (n.d.). Technology Risk. Ernst & Young Global Ltd. Retrieved February 21, 2022, from https://www.ey.com/en_in/consulting/technology-risk-services

EY Atlas. (n.d.). *IT: IT Processes [effective for audits of periods ending on or after 15 December 2020].* live.atlas.ey.com. Retrieved March 9, 2022, from https://live.atlas.ey.com/#document/1707578/SL314828937-1707578?pref=20043/19/102&crumb=8/638336

EY. (2019, January 7). *How an AI application can help auditors detect fraud*. EY - US. https://www.ey.com/en_us/better-begins-with-you/how-an-ai-application-can-help-auditors-detect-fraud

Gantz, S. D. (2013). *The Basics of IT Audit: Purposes, Processes, and Practical Information (Basics (Syngress))* (1st ed.). Syngress.

Goertzel, B., & Pennachin, C. (2006). *Artificial General Intelligence*. Springer Publishing.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning (Adaptive Computation and Machine Learning series)* (Illustrated ed.). The MIT Press.

Glorot, X. & Bengio, Y.. (2010). *Understanding the difficulty of training deep feedforward neural networks*. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research. 9:249-256 Available from https://proceedings.mlr.press/v9/glorot10a.html.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts. Journal of Artificial Intelligence Research, 62, 729–754. https://doi.org/10.1613/jair.1.11222

Hall, J. A. (2015). *Information technology auditing*. Cengage Learning.

Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*.

Hinson, G. (2007). The State of IT Auditing in 2007. *EDPACS*, *36*(1), 13–31. https://doi.org/10.1080/07366980701547065

Li, L., & Abu-Mostafa, Y. S. (2006). Data complexity in machine learning.

Lin, T. H., & Jiang, J. R. (2021). Credit Card Fraud Detection with Autoencoder and Probabilistic Random Forest. *Mathematics*, *9*(21), 2683. https://doi.org/10.3390/math9212683

Mehrotra, K. G., Mohan, C. K., & Huang, H. (2017). Anomaly detection principles and algorithms (Vol. 1). New York, NY, USA:: Springer International Publishing.

Moeller, R. R. (2010). IT audit, control, and security (Vol. 13). John Wiley & Sons.

Nico Gornitz, Marius Kloft, Konrad Rieck, and Ulf Brefeld. Toward supervised anomaly detection. ¨ Journal of Artificial Intelligence Research, 46:235–262, 2013.

Paris, A. (n.d.). *Continuous Control Monitoring (CCM) – What is it, and Why is it so important?* Metricstream. Retrieved February 21, 2022, from https://www.metricstream.com/blog/continuous-control-monitoring-why-it-so-important

Rao, K. H., Srinivas, G., Damodhar, A., & Krishna, M. V. (2011). Implementation of anomaly detection technique using machine learning algorithms. *International journal of computer science and telecommunications*, *2*(3), 25-31.

Romney, M. S. P. J. B. (2017). Accounting Information Systems, Global Edition (14th edition). Pearson Education Limited.

Schreyer, M., Sattarov, T., Borth, D., Dengel, A., & Reimer, B. (2018, August). *Detection of Anomalies in Large-Scale Accounting Data using Deep Autoencoder Networks*.

Teeter, R. A., Brennan, G., Alles, M. G., & Vasarhelyi, M. A. (2008*). Aiding the audit: using the IT audit as a springboard for continuous controls monitoring.* Unpublished working paper, Rutgers business school.

van Veen, F. (2017, April 1). *Neural Network Zoo Prequel: Cells and Layers.* The Asimov Institute. https://www.asimovinstitute.org/neural-network-zoo-prequel-cells-layers/

van Veen, F. (2019, April 27). *The Neural Network Zoo*. The Asimov Institute. https://www.asimovinstitute.org/neural-network-zoo/

Zhou, C., & Paffenroth, R. C. (2017). Anomaly Detection with Robust Deep Autoencoders. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/3097983.3098052

APPENDICES

Appendix I: IT Related Audit Activities EY

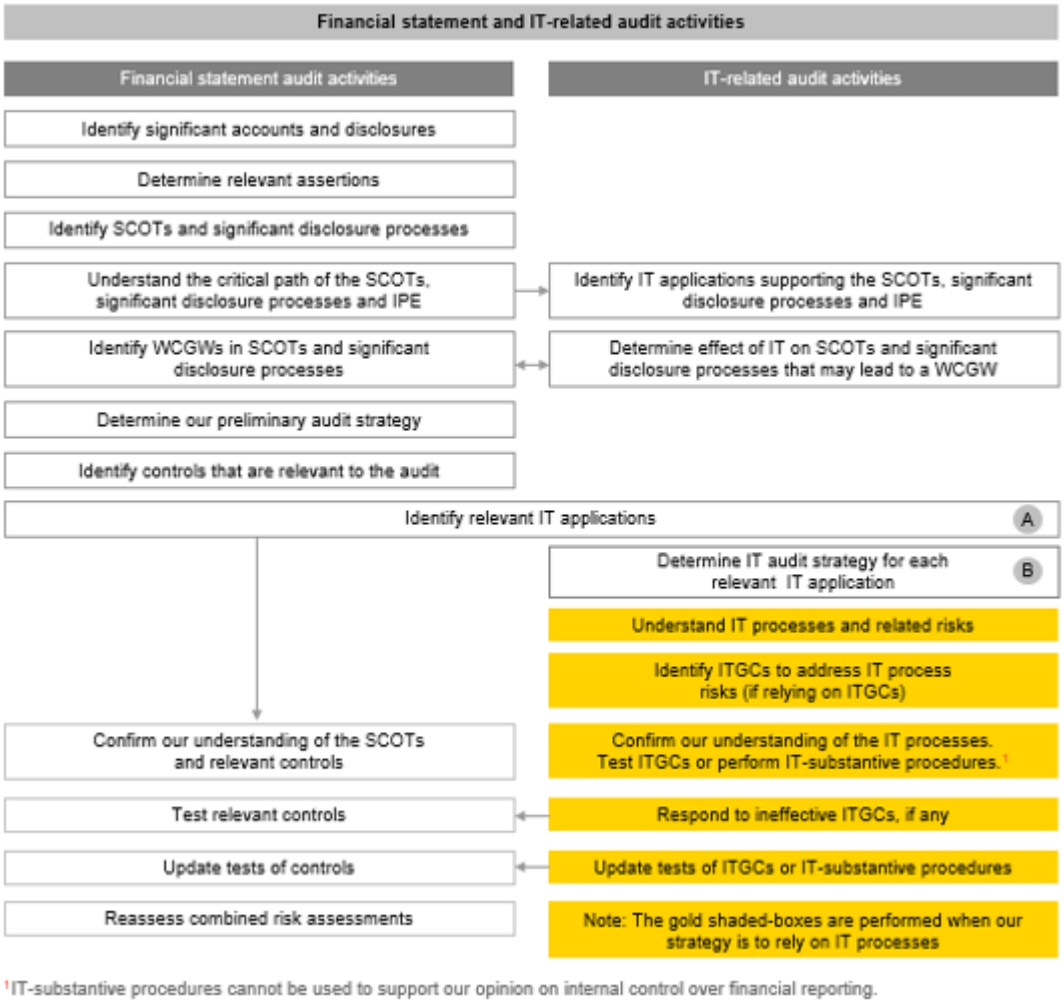

Figure 13 Appendix 1: IT Related Audit Activities (EY Atlas, n.d.)[6]

[6] Source from internal EY portal which is not publicly available

Appendix II: Exploratory Interviews Summary

Table 5 Interview Candidates & Roles

Interview	Role
Interview A	Senior Assurance, Risk ASU, ASU Risk FAIT IT Auditor
Interview B	Partner/Principal Assurance, Risk ASU, ASU Risk FAIT
Interview C	Manager Assurance, Risk ASU, ASU Risk FAIT
Interview D	Manager Assurance, Risk ASU, ASU Risk FAIT Senior IT Auditor
Interview E	Senior Manager Assurance, Risk ASU, ASU Risk FAIT Manager Data Analytics FSO
Interview F	Senior Manager Assurance, Risk ASU, ASU Risk FAIT
Interview G	Senior Manager Assurance, Risk ASU, ASU Risk FAIT
Interview H	Manager Assurance, Risk ASU, ASU Risk FAIT

Do you think Continuous Control Monitoring can be used as a source of data for anomaly detection using ML?

Continuous control monitoring is likely not useful in its current state for machine learning as there is not enough data coming in. (Interview A)

There are currently not many clients that make use of CCM, and the ones that do only bring in a minimal amount of data. Not enough for ML as of now. (Interview B) (Interview E)

CCM has a good potential for the future. At the client side it could replace multiple controls with one solution that can notify the client as soon as something goes wrong, so it can be fixed. When looking at how much it is implemented now, I would say it is very few. (Interview F)

Are there any other sources of data in IT Audit that could be sufficient for anomaly detection?

Yes, data does not necessarily have to be a problem for anomaly detection and ML. One option would be to look at client dossiers from the past to train an algorithm on this data. The dossiers have already been closed so there will be no findings that can be used for the IT Audit for this dossier, but it also means that no additional data has to be requested at the client side. This means there will be no risk of negative impact on the relationship. The general idea here is to see if ML and anomaly detection can be used on closed dossiers, and to then compare the results with the results from the actual IT Audit too see if this would have led to new insights, or if it could have helped to reach the same conclusions with less work. This would mean too use old data to test the technique for research purposes. (Interview A)

A problem for IT Audit would be that you need a lot of data, and every client designs its processes slightly different. On the other hand, the Delta between these processes is not that big in general. Combining data from different clients could get an algorithm to 90% functionality. Then for every client it just has to be adjusted for the remaining 10%. This could be done in for example change management, or incident management. (Interview C)

Data is currently collected based on the sampling. This means that after the random selection has been made, the evidence is requested at the client. In order to get more data than just the samples, more has to be requested at the client, which might not always be appreciated. (Interview E)

IT Audit currently works with samples, however for anomaly detection to be successful this would likely have to change. Do you think this is an option?

When we look at the way IT Audits are performed now, we base a conclusion on samples. Say we have 250 instances, we would take a sample of 25. Based on this sample we can come to a reliable enough conclusion, however if we look at all 250 instances, there is ten times more chance that we will find something that is going wrong. Which would also imply more work for the IT Auditors, and more work for the client. Especially the partners here are still in the mindset of "why would I do much more, if the current approach gives a reliable enough answer?". This means that a shift in the way of thinking is required to start to get rid of samples. Often, the extra cases that will be found where something appears to be wrong, will not have a big impact other than extra work, as they turn out to be exceptions that can be explained. It would be great if there was a way to take this into the analysis beforehand. If you end up with a large number of anomalies, but the user of the results cannot trust that any of them will actually have an impact, then there will be little use for the tool. The challenge here is that if you have to take the context of the organization too much into account, there will be a lot of extra work to any analysis, which could reduce the usefulness. But if you do not take the context into account enough there might be a large number of false positives. (Interview A)

Statistically the current approach might cover the risks. But it can be considered outdated when looking at the technologies that are available. When looking at the full population there is the risk that a lot of time has to be spend on analyzing what went wrong. (Interview E)

If there would be a way to look at more data and get more results, without too much extra work compared to the current methods, then this would definitely have potential. This could be interesting from an efficiency standpoint testing could come closer to substantive testing over ITGCs without having as much work as substantive testing. (Interview G)

A problem I run in to now, is that there appear to be few tasks and processes in IT Audit that make use of enough data to make ML useful AND are structured enough to make an algorithm fit more than one single instance. Do you have any ideas for specific tasks or processes?

This is a challenging issue, there are some processes that we have started using RPA for about ten years ago. This works for processes that can be repeated the exact same way. If you want to look at AI/ML instead, then there is very little data available on this subject so far. (Interview A)

Looking at the changes that companies make, ML could be trained on a set of "right" changes and "wrong" changes so that predictions can be made on changes for a client. Currently EY is no where near the point where it is understood if this is possible and how it would work. The problem is that Neural networks need a lot of data. But there is not that much Delta over the clients, so maybe you can combine data from multiple clients to get a general model (90% complete) and then finetune the algorithm on specific clients when you need it. This could be tried on for example MC data, or data on how incidents are processed in the IT environment. It would be good to train a NN to make sure there are no security incidents that slip past the process. (Interview C)

In MC there are certain new systems at the client side that work on micro updates. This can cause thousands or tens of thousands per year to occur. Not every change is interesting for IT Audit. AI could play a role in determining which changes are interesting to IT Audit and which are not. (Interview F)

Is there reason to look for anomaly detection solutions in IT Audit?

IT Audit has a big problem with recruiting and retaining employees. In the Netherlands we hire employees even when they do not speak Dutch as the demand is bigger than the supply. Employees used to stay in the job for around six or seven years, now this has been decreasing for a few years and it is often three or four. (Interview C)

Can you explain more about the processes within IT Audit at EY?

IT Audit at EY is split into four parts MA, MC, MO and CS. For each group, there are ITGCs, ITACs/ITDMs and substantive testing. Controls can be preventive, detective or corrective, and are designed by the client. In IT Audit we provide trust that the controls work by testing the design and the effectiveness. When looking for anomaly detection solutions, the focus should be on MA, MC and MO. (Interview D)

There seem to be a lot of developments in EY in different countries in financial audit (e.g. EY Helix GLAD), are there developments like this in IT Audit?

It is interesting that there are developments like this (EY Helix GLAD) that are not really known in the Netherlands. It is a shame that not all information is shared, but it is not strange, as the focus is more on short term issues with pressure on them. (Interview C)

In financial Audit there are more developments regarding new technologies and methods than in IT Audit. Although we recently experimented with analytics on MA. In MA, clients often work in their own structure with a lot of emails or tickets, so we had to create a data structure they could use instead. For the data analytics the view is on the entire data population rather than just samples. This was simply a pilot but has not been used in practice a lot. There is still room for a ML solution for MA ass this RPA solution does not make complex findings, but mostly saves time in the acquisition of data through standardization. (Interview E)

An argument that has been given against more testing (full population rather than samples) is that it might cost more time to solve, while the current approach is sufficient. What is your opinion on this?

The change from samples to testing the full data population would increase the quality of the IT Audit. This can be a slow change however. (Interview E)

It depends on how efficient it can function. If the tool is good enough, then it could almost be considered as a more efficient way of substantive testing. (Interview G)

A solution that has been proposed before is to look at multiple clients and combine the data. Is there a process within IT Audit where the data is similar enough between clients?

This is a tough question. For this the data could maybe be put in a template after acquiring it. (Interview E)

Are there any expected problems for anomaly detection in IT Audit?

Working in IT Audit with AI would require a form of algorithm assurance. There needs to be a way to guarantee that the algorithm does what can be expected of it. (Interview F)

Another argument against changing the process to look at more data then samples is that the client would have to do more work in delivering data and explaining findings. What is your view on this?

If data would be requested for example monthly or every half year, and the data can retrieved relatively easy, then this should not form a problem. In many cases retrieving all data from a system in a given period is not more work than extracting only a few specific instances that have to be extracted one by one. If more findings come to light based on the anomaly detection algorithm and this happens every few months, then clients will learn from this process and adapt to get less anomalies in the future. (Interview G)

Appendix III: Interview Questions

Part I: Potential

(Filter risks, closed questions)

1. [Impact] Are there severe consequences to misstatements?
2. [Frequency] Is the task linked to this risk frequent in your work?
3. [Duration] Is the task linked to this risk time-consuming?
   (Or would it be if the full population would be tested?)
4. [Pressure] Is there time pressure on the task linked to this risk?

Part II: Dependencies

For the tasks linked to this risk:

5. [standard.] Are the tasks standardized over multiple clients?
   a. Are there common systems/structures that this usually happens in? (that would allow an algorithm to learn from data from other clients, or allow the algorithm to be repurposed for other clients)
   b. Are similar data fields used (even if names differ) for multiple clients?
6. [complexity] Is the task complex?
   a. Can this task be solved with RPA?
   b. Can data be classified as anomalous or normal?
7. [Data quant.] Does is task involve large data quantities?
8. [Pressure] Would a test on the full population add quality over testing samples?
9. [Relation] Would the relationship with the client be at risk?
   a. Would requesting more data deteriorate the relation?
   b. Would false positives deteriorate the relation?

Part III: Solutions

(Only look for variables that have been determined as deficiency)

Look at the four approaches to see if they would solve the deficiency.

[standard.] Data is not standardized:

10. Can data be standardized?
    a. Would clients be willing to work in a unified way if it could enable better audit?
    b. Could a template be created that the IT Auditor could put the data in before using the anomaly detection tool?
    c. Is the task big/important enough to warrant an algorithm that can detect an unknown pattern (unsupervised) for a single client?

[complexity] Data is not complex enough:

11. Could anomaly detection add something over RPA?

Data is too complex:

12. Is there a way data can be divided into normal or anomalous?

[Data quant.] Task does not involve enough data:

13. Can enough data be gathered by testing the entire data population instead of samples?
14. Is task standardized enough to enable combination of data from multiple clients?

[Pressure] No improvement in quality would occur from testing full population instead of sampling:

15. Could anomaly detection be used to increase the efficiency instead?

[Relation] Requesting more data would bring the relationship at risk:

16. Could a solution like CCM reduce the amount of work that a client would have from delivering extra data?

False positives would bring the relationship at risk:

17. Would the clients be able/willing to quickly adapt to new requirements in logging and delivering evidence so that the false positives would decline over time?
18. Would a different balance in recall/accuracy be a solution? (The balance between false positives and misstatements that slip through the algorithm)

[Open]

19. Do you have any other solutions that have not been mentioned?

Part IV: Feasibility

20. Are the solutions mentioned realistic/feasible?
    a. Why (not)?

Appendix IV: Interview I

Interview I1

Role: Staff/Assistant
Assurance, Risk ASU, ASU Risk FAIT
Staff IT Audit

Discussing the potential use-cases based on common risks led to a few insights about the potential of anomaly detection in IT Audit. Three risks have been indicated as having a high potential for this research. Table 6 Potential Per Risk (interview I)" below shows the ratings given in the interview.

Risk number 9 (from manage access) was given the most potential for successful anomaly detection implementation. This is a risk that is covered in practically every IT audit, so it is highly recurrent. The work related to this risk takes a lot of time, and the potential impact is extremely high as there is risk for fraud and abuse of access.

Risk number 7 (access management) received just a slightly lower score overall. With a lower impact then risk 9, but a higher frequency. The lower score comes from a slightly lower duration and pressure.

The risks 1, 2 & 4 received average scores when taken individually. The tasks related to these risks are often performed together as they are closely related. When looking at these risks as a group, they are much more promising. As the tasks related to the risks are combined in an IT Audit, the decision was made to look at these risks as a group instead of individual risks.

Table 6 Potential Per Risk (interview I)

Manage Change		Impact	Frequency	Duration	Pressure	Total
1	*New IT application programs or changes to existing programs, including reports, configurations and interfaces, do not function as described or requested because they are not adequately tested by appropriate persons.*	3	5	3	3	14
2	*New IT application programs or changes to the production IT application programs (including reports and interfaces) are not appropriate for the business or the IT environment.*	3	3	3	2	11
3	*Programs in production are not secured permitting developers to move unauthorized or untested changes into the production environment.*	4	4	3	3	14
4	*Configuration changes made by IT personnel are inappropriate or unauthorized.*	4	4	2	3	13
5	*Multiple instances of the same IT application that should be identical are not the same.*	3	3	2	2	10
Manage Access		Impact	Frequency	Duration	Pressure	Total
6	*Users of the IT environment aren't the intended users due to inadequate authentication and security settings.*	4	2	2	2	10
7	*Access rights risks:* *'- Access granted to the IT environment (IT and Business) does not match the access approved* *- Access termination requests are not fulfilled timely* *- Access rights to the IT environment (IT and Business) do not remain appropriate over time.*	4	5	3	3	15
8	*Access requests for IT and business users of components of the IT environment are inappropriate*	4	4	3	3	14
9	*The access of IT users of the IT environment creates segregation of duties concerns.*	5	4	4	4	17
10	*Access to functions within the IT application is combined into roles. The access rights within the roles contain segregation of duties issues that could cause a material misstatement of the financial statements.*	4	4	3	3	14
11	*Direct data changes are made without authorization. (Of higher risk when there is routine use of direct data changes in the processing of transactions relevant to the financial statements.)*	4	4	2	2	12
Manage Operations		Impact	Frequency	Duration	Pressure	Total
12	*Hardware or software issues result in loss of data or the ability to accurately process that data.*	4	2	1	2	9
13	*Issues with programs that cannot process to completion are not addressed or are addressed inappropriately.*	3	2	2	2	9

Interview I2

Risk number: 9

Risk description: *The access of IT users of the IT environment creates segregation of duties concerns.*

Dependencies Risk 9

Variable: degree of standardization

1. *Are the tasks standardized over multiple clients?*
    a. *Are there common systems/structures that this usually happens in? (that would allow an algorithm to learn from data from other clients, or allow the algorithm to be repurposed for other clients)*
    b. *Are similar data fields used (even if names differ) for multiple clients?*

It is usually the same or a similar approach. A system that is often used for this by clients is SAP GRC (Governance, Risk & Compliance). This system has a module that can be used to make this analysis. Smaller clients will not always use this module from SAP GRC but might instead do a review on the authorization of a person and whether or not it matches the role they have.

There is a difference here in the use of preventive and detective controls. Bigger clients will generally have detective controls on top of their preventive controls, where they check if someone has accidently received access to a "transaction" that was unintended. The degree of standardization in the end is relatively big here, as a lot of companies use SAP for this.

What kind of data is used in these tasks?

Often Excel files. There can be differences per client, as companies can determine for themselves what profiles they consider critical, high importance or something similar. And only relevant profiles are included in the control. The analysis itself is often done by a manual check in Excel.

Variable: complexity

2. *Is the task complex?*
    a. *Can data be classified as anomalous or normal?*

The tasks performed at EY for this risk are also often performed in Excel, which is a very time-consuming job. EY performs sort of a "reperformance" of what the client has done

in the internal audit. What makes it difficult is that there can be exceptions to what authorizations a person can have based on their profile. Often this is explainable and there is a valid reason for the exception. The problem is that the evidence is often in the form om email, which makes it more complex to work with. Most of the work for the IT Audit is in Excel files however, which is also due to the large volume of the files. The situation before describes a substantive test of a client, which means this is not tested based on samples. In other cases this can be based on samples instead. With the client described before, mistakes were found which led to the decision for substantive testing, where more mistakes were found. This is exactly why solutions like anomaly detection are needed in IT Audit. Even with just samples it takes long, but when testing everything a lot more is found.

b. Can this task be solved with RPA?

No I do not believe this can be solved with RPA. This is also one of the tasks that is easily combined with financial aspects. If you can find a pattern in which people make transactions and when and for how much, you could use this to discover potential fraud. Fraud is nowadays often done using a large number of small transactions as a way to hide it. Usually it will be the same people or people from the same role. The data can probably indicate this somehow, but it is not something RPA could solve.

Variable: data quantity

3. Does is task involve large data quantities?

The work related to this risk is already intense the way it is done now using samples. When looking at all the data involves there would definitely be enough data available.

Variable: pressure

4. Would a test on the full population add quality over testing samples?

If anomaly detection could be used to look at the full data population, and send back a shortlist with anomalous instances that have to be tested instead of the usual samples, would this improve audit quality?

Using a solution like this over the full population would enable us (IT auditors at EY) to discover more questionable instances and provide better assurance. I believe this would definitely improve quality.

Variable: relationship with the client

5. Would the relationship with the client be at risk?

a. Would requesting more data deteriorate the relation?

Requesting a lot of data from the client will not be a problem. This is simply getting data from a system and would not be a struggle point. When the argument is presented to the client in a way that shows that this new way of working would provide better results, then it is my expectation that the clients will even be excited for this. Especially with this risk, as there is a lot at stake for the clients. This is where you can discover fraud especially.

b. *Would false positives deteriorate the relation?*

No not necessarily. Not everyone in a company might be happy with feedback about what is found. Most clients would however be interested to know what is going on.

Solutions Risk 9

Variable: degree of standardization

6. *Can data be standardized?*
   a. *Would clients be willing to work in a unified way if it could enable better audit?*

Yes clients would be willing to work in a more unified way if the purpose is properly explained. Most companies would be happy to have a way that can offer more assurance, especially if it only requires minor changes like a standard template for emails that they use.

   b. *Could a template be created that the IT Auditor could put the data in before using the anomaly detection tool?*

Yes the data could be standardized for a lot of companies. The controls are made by the clients themselves, but the work and execution is similar enough to try to standardize more.

   c. *Is the task big/important enough to warrant an algorithm that can detect an unknown pattern (unsupervised) for a single client?*

This risk is absolutely important enough to think about new and innovative solutions. And the amount of work that it is for an IT Auditor even when working on sample basis means that right now there is no way to look at more data.

Variable: complexity

7. *Could anomaly detection add something over RPA?*

The data is already too complex for RPA.

8. *Is there a way data can be divided into normal or anomalous?*

The reason there is no standard solution for this at the moment is that there are still differences between the clients. The data can be classified as anomalous or not, but the differences made it difficult to do so before.

Variable: data quantity

9. *Can enough data be gathered by testing the entire data population instead of samples?*

The task already contains enough data.

10. *Is task standardized enough to enable combination of data from multiple clients?*

The data from multiple clients could possibly be combined since the task is reasonably similar across big companies, especially when working on the same system like SAP GRC. It does not even matter much if clients are using Excel or not, as there are only very few change management applications where you cannot get an export to Excel. They all include information like the day of the change, time of the change, and the system.

Variable: pressure

11. *Could anomaly detection be used to increase the efficiency instead?*

-

Variable: relationship with the client

12. *Could a solution like CCM reduce the amount of work that a client would have from delivering extra data?*

Getting the data from the client would not be a problem for this risk if they know what the reason is. Getting all data directly from the system could even be less work.

13. *Would the clients be able/willing to quickly adapt to new requirements in logging and delivering evidence so that the false positives would decline over time?*
14. *Would a different balance in recall/accuracy be a solution? (The balance between false positives and misstatements that slip through the algorithm)*

-

Variable: open question

15. *Do you have any other solutions that have not been mentioned?*

-

Feasibility Risk 9

16. *Are the solutions mentioned realistic/feasible?*
    b. *Why (not)?*

They appear to be realistic.

Interview I3

Risk number: 1, 2 & 4

Risk description:

1. *New IT application programs or changes to existing programs, including reports, configurations and interfaces, do not function as described or requested because they are not adequately tested by appropriate persons.*
2. *New IT application programs or changes to the production IT application programs (including reports and interfaces) are not appropriate for the business or the IT environment.*
4. *Configuration changes made by IT personnel are inappropriate or unauthorized.*

Dependencies Risk 1, 2 & 4

Variable: degree of standardization

1. *Are the tasks standardized over multiple clients?*
    a. *Are there common systems/structures that this usually happens in? (that would allow an algorithm to learn from data from other clients, or allow the algorithm to be repurposed for other clients)*

The level of standardization in systems that clients use is really high. Most common is the use of SAP, but another option that is often used is for example Microsoft Dynamics. About 60 to 70% of the clients use SAP for this. This is a highly standardized module in SAP for with even the fields have been pre-determined. The resulting table looks the same across all clients, even if you use different systems the files will look similar. The general level of standardization is judged as remarkably high.

    b. *Are similar data fields used (even if names differ) for multiple clients?*

The data looks the same for most clients ass they use the highly standardized systems like SAP. The difference can be that small companies might use email for requests and confirmation while bigger clients typically use a ticketing system like ServiceNow. A ticketing system will provide a higher level of standardization as it ensures that every instance will have the same structure and look.

What kind of data is used in these tasks?

The data for the IT Audit is a combination of change logs, tickets and emails.

Variable: complexity

2. Is the task complex?
    c. *Can data be classified as anomalous or normal?*

Using ML this task could be useful for anomaly detection. The tickets or emails now difficult to classify. A NLP solution could be used to look through text in emails and

tickets, but this would make the task complex to implement, as the NLP tool would have to gather data from emails and tickets, and then an anomaly detection algorithm has to include this data. This might increase complexity for this risk.

d. Can this task be solved with RPA?

It depends on the way the textual evidence is noted. Sometimes this can include incredibly long tickets with evidence included somewhere in the text. This would be difficult to automate when the tickets can be different.

Variable: data quantity

3. Does is task involve large data quantities?

Relatively speaking, the tasks related to these risks is where there is data for every client, and the bigger the client the more data will be available. Compared to other risks this includes a large quantity.

Variable: pressure

4. Would a test on the full population add quality over testing samples?

If anomaly detection could be used to look at the full data population, and send back a shortlist with anomalous instances that have to be tested instead of the usual samples, would this improve audit quality?

Yes. The problem with the current approach of samples is that there can be unapproved changes going to the production environment. For example, in half a year a client makes 800 changes. In this example we might test around 18 samples over this half year period, at the end of the year we will have tested 25. The chance that you will spot the changes that were not approved properly or where the testing evidence was not documented, is really small. At bigger companies there might be special projects where there are 200 or 300 changes going to the production environment in just one day. To make sure we do not only sample from this project, it is often taken separate from the rest of the changes. This means that from this project only a miniscule amount of changes are tested. While in practice there is often at least a few changes in a project like this that go wrong or that is not documented properly. Being able to test everything, or to test more targeted would make a big difference.

Variable: relationship with the client

5. Would the relationship with the client be at risk?

c. Would requesting more data deteriorate the relation?

Requesting more data would not be a problem at all. These are all examples where you can explain the reasoning behind the request for more data. Only companies that currently

do not properly document their changes might have difficulties as they would need to look for the data. For companies where is it documented well it would not mean much extra work.

d. Would false positives deteriorate the relation?

If ten examples would be send back to the client and eight of them turn out to be perfectly explainable then the client would not appreciate that at all. However, if we look at the results of the anomalies as a list as a selection tool on what to include in the IT Audit instead of the random sampling, then there will still be a lot more issues that will be brought to light. These issues might all be explainable and simply a mistake in documentation, but this is something the client will not necessarily mind. In fact, the responsible person for changes, like the change manager, might be interested to know that there are many mistakes in documentation. A client would likely be happy to know that there are

Solutions Risk 1, 2 & 4

Variable: degree of standardization

6. *Can data be standardized?*
   d. *Would clients be willing to work in a unified way if it could enable better audit?*

The biggest issue with standardization is in the emails and tickets that are part of the work. If clients could help to improve the IT Audit with a minimal change like using a template for the email or ticket and including a number so it can be automatically recognized with change it is part of, then this should be very doable. The clients would likely not be willing to change their way of testing. They might be willing to change the way of documentation.

e. Could a template be created that the IT Auditor could put the data in before using the anomaly detection tool?

f. Is the task big/important enough to warrant an algorithm that can detect an unknown pattern (unsupervised) for a single client?

Variable: complexity

7. *Could anomaly detection add something over RPA?*

Not Applicable

8. *Is there a way data can be divided into normal or anomalous?*

The main reason for the complexity is the emails and tickets, for which NLP might be necessary if they are not standardized enough. Other then standardizing this at the client

side, this complexity can be avoided by changing the approach slightly. Not all client currently require to look into the email or ticket contents, as the fact that a ticket or email to confirm has been sent is enough in a lot of cases. This might be more common than having to look through the content of the email in detail. NLP might be a step further that can still be included later on.

Variable: data quantity

9. *Can enough data be gathered by testing the entire data population instead of samples?*

There is already enough data. However, when implementing a solution like this, the aim is always to make it work for multiple clients with only minor adjustments.

10. *Is task standardized enough to enable combination of data from multiple clients?*

Yes, even for the ticketing systems that can increase the complexity there are only very limited different options (such as Service Now and TOPdesk). On top of that the solution should be aimed at the common situation and not on the few exceptions to the rule.

When applying this (or any ML anomaly detection solution) for a single client, you still have to make it worth the effort. For example by making it functional for other clients with adjustments. These risks are good examples of where t could work to combine data from multiple clients though, because it is so standardized.

Variable: pressure

11. *Could anomaly detection be used to increase the efficiency instead?*

For instances where substantive testing is performed now it could help if this would prove reliable enough.

Variable: relationship with the client

12. *Could a solution like CCM reduce the amount of work that a client would have from delivering extra data?*

For companies that have their documentation in order it should be no problem to request more data. If data can be taken directly from the client's system then it would turn out to be even less work for them.

13. *Would the clients be able/willing to quickly adapt to new requirements in logging and delivering evidence so that the false positives would decline over time?*

Not applicable

14. *Would a different balance in recall/accuracy be a solution? (The balance between false positives and misstatements that slip through the algorithm)*

Not applicable

Variable: open question

15. Do you have any other solutions that have not been mentioned?

No.

Feasibility Risk 1, 2 & 4

16. Are the solutions mentioned realistic/feasible?

c. Why (not)?

The solutions seem feasible because it will ultimately help the client, which means they will be willing to make changes. The use of anomaly detection seems to offer benefits if it can be implemented.

Interview I4

Risk number: 7

Risk description:

Access Right Risks:

- Access granted to the IT environment (IT and Business) does not match the access approved
- Access termination requests are not fulfilled timely
- Access rights to the IT environment (IT and Business) do not remain appropriate over time.

Dependencies Risk 7

Variable: degree of standardization

1. *Are the tasks standardized over multiple clients?*
   a. *Are there common systems/structures that this usually happens in? (that would allow an algorithm to learn from data from other clients, or allow the algorithm to be repurposed for other clients)*

Some companies may use systems for this, but it is common for this to happen over email.

   b. *Are similar data fields used (even if names differ) for multiple clients?*

The tasks are relatively standardized, access request are usually done in a uniform way.

What kind of data is used in these tasks?

A combination of (Excel) extracts from systems and mostly email interactions.

Variable: complexity

2. *Is the task complex?*
   a. *Can data be classified as anomalous or normal?*
   b. *Can this task be solved with RPA?*

The possibility that this can be approached with an RPA solution is bigger then for the other risks discussed in the interview. Especially when looking at the first risk description for risk 7. There are more structured rules for this risk that an RPA solution might help with. However, the last description for risk 7 about risks remaining appropriate over time is more difficult for RPA. There is usually a review to check whether or not the access someone has are still appropriate. This is usually done by the manager of this individual. Finding a pattern in the available data might be challenging however.

Variable: data quantity

3. *Does this task involve large data quantities?*

A solution for this risk would not likely be applied to a small client. For bigger clients there is often a system with all employees, roles and functions that is easy to extract. The amount of data is still probably less then for the other risks that have been discussed, but for big companies there can be a huge turnover rate.

Variable: pressure

4. *Would a test on the full population add quality over testing samples?*

If anomaly detection could be used to look at the full data population, and send back a shortlist with anomalous instances that have to be tested instead of the usual samples, would this improve audit quality?

Yes, for this risk it is especially interesting to look at the full population as it is more common that something is amiss. For example a new hire that accidently gets the same access as an existing employee while the existing employee had another role that was not required for the new hire. This happens relatively regularly.

Variable: relationship with the client

5. *Would the relationship with the client be at risk?*
    a. *Would requesting more data deteriorate the relation?*

For some companies, requesting more data might be problematic. If emails are not documented in a structured way, requesting information about more processes would require a lot of extra work from the client. Other companies that use a ticketing tool like ServiceNow might have no problem with the request for more data.

    b. *Would false positives deteriorate the relation?*

As the anomalies are not directly used as feedback for the client but first go through the "normal" IT Audit process the extra feedback will be of value. This should not be an issue.

Solutions Risk 7

Variable: degree of standardization

6. *Can data be standardized?*
    a. *Would clients be willing to work in a unified way if it could enable better audit?*

Yes, especially the emails that are often used in this process can easily be standardized by creating a template that requires minimal change for each task. If this helps the client

they will be willing to make these changes. This might be easier to change for even the small companies then the other risks, as for small companies this would be a small change for them.

b. *Is the task big/important enough to warrant an algorithm that can detect an unknown pattern (unsupervised) for a single client?*

Big companies with a high turnover rate have a lot of data available for this process. And big companies with a high turnover rate are not uncommon at all. However, again it is desirable to create a solution that can be used on multiple clients with small adjustments.

Variable: complexity

7. *Could anomaly detection add something over RPA?*

RPA could potentially be used for parts of this risk. ML might be better suited to work with data from emails, but if the process can be standardized RPA might be possible.

8. *Is there a way data can be divided into normal or anomalous?*

There should be a pattern in the data that can help to determine whether something is an anomaly.

Variable: data quantity

9. *Can enough data be gathered by testing the entire data population instead of samples?*

Yes, as discussed before especially at bigger companies there is a lot of data and testing the entire population would improve the process.

10. *Is task standardized enough to enable combination of data from multiple clients?*

For a lot of companies yes.

Variable: pressure

11. *Could anomaly detection be used to increase the efficiency instead?*

-

Variable: relationship with the client

12. *Could a solution like CCM reduce the amount of work that a client would have from delivering extra data?*
13. *Would the clients be able/willing to quickly adapt to new requirements in logging and delivering evidence so that the false positives would decline over time?*

The problem about the structure in the documentation of emails regarding this risk could be solved with relatively simple changes. A standard template in outlook combined with built-in options to perform certain actions with mails about this risk could solve the problem. Whether or not companies would be willing to make changes like these is not certain.

14. Would a different balance in recall/accuracy be a solution? (The balance between false positives and misstatements that slip through the algorithm)

Not applicable

Variable: open question

15. Do you have any other solutions that have not been mentioned?

No the questions covered it.

Feasibility Risk 7

16. Are the solutions mentioned realistic/feasible?

a. Why (not)?

This seems like an opportunity for anomaly detection to work as companies see this risk as highly important. If the risk is more important for clients there is more incentive to make changes to their way of working, and there is usually more documentation. The solutions seem possible.

Appendix V: Interview J

Interview J1

Role: Manager
Assurance, Risk ASU, ASU Risk FAIT

The risks that can potentially be addressed using anomaly detection have been discussed extensively. The pressure from the client is not regarded as an influence that can be judged for any risk, as this should not have much impact on the decision making. Therefore, it has been given an equal score among all risks.

Manage Change

The impact of the testing of changes (risk number 1) is the biggest out of all the risks in change management, as in this process the validation is performed on whether or not the change works as expected and whether or not there is a negative impact on other objects in the application. A risk analyses beforehand is important to know what to test.

Risk number 2 is more about the approval before a change goes to production, which included a validation on all different facets (including testing) are present. It is about challenging and taking responsibility of the change. The impact for risk 2 is slightly lower than for risk 1, frequency and duration are the same.

Risk number 3 is about the possibility that developers can circumvent the process for risk 1 and 2, and put a change into the production environment. The potential impact here is high, but the likelihood is quite low. The reason for the high impact is that it often indicates conscious actions that have a negative impact (fraud). The frequency here is lower, and duration is lower as well as it is often based on merely two user lists.

For risk number 4, the same logic can be followed as for risk 1 and 2, as they are reviewed together.

Risk number 5 is not frequent, it is possible that companies have the same application running in different countries and on different updates. This has to be checked to ensure that the reviews on other controls are actually effective on all different versions.

In general the tasks for risk 1, 2 and 3 are audited in the same way. After determining the population of changes and configuration changes, a sample is taken and test are performed for controls on the three mentioned risks combined.

Manage Access

Risk 6 is about the authentication method. Depending on the method there can be a password setting validation or a LDAP/AD validation. These validation generally happen about once or twice per year. The duration is also not that long.

Risk 7 is about the user access being approved, reviewed and withdrawn when the user leaves the company. Mistakes can have a high impact, especially for leavers where the access right are not withdrawn.

Risk 8 is about starters or movers in an organization that get an account, which will be approved by the right people, and with access rights relevant for the job function. Giving someone the wrong access rights can have a very high impact, and even more so since people can be assigned access rights outside of the standard process/protocol. Frequency depends on the requests for access, which can be high or low. Duration however is quite high for us as there are up to 25 samples that all have to be examined from request to approval.

Risks 9 & 10 are about segregation of duties (SOD) concerns. The frequency for these is lower then for other risks as this is usually reviewed once per year.

A change directly made in the database can have a high impact, which is why there has to be a protocol and approval for this. Applications can contain controls, and with direct changes in the database those controls can be circumvented. Duration can be high as the population has to be determined first and the process has to be understood.

Manage Operation

Risk 12 is about making back-ups. This is done frequently, often daily or real-time. On top of that testing the restoration of back-ups happens around once or twice per year. The impact can be very high but the likelihood of issues is less high. If an issue occurs it can mean loss of data. The duration is high because of the many samples.

Risk 13 is about job monitoring. The impact can be high if data is not being processes between systems. Since it often includes real-time monitoring there can be a high frequency. The duration is high because of the many samples. It can be hard to test if it has been resolved in a timely manner, as the data and documentation does not always show what happened in reality.

Table 7 Potential Per Risk (interview J)

Manage Change		Impact	Frequency	Duration	Pressure	Total
1	*New IT application programs or changes to existing programs, including reports, configurations and interfaces, do not function as described or requested because they are not adequately tested by appropriate persons.*	5	4	4	3	16
2	*New IT application programs or changes to the production IT application programs (including reports and interfaces) are not appropriate for the business or the IT environment.*	4	4	4	3	15
3	*Programs in production are not secured permitting developers to move unauthorized or untested changes into the production environment.*	5	3	2	3	13
4	*Configuration changes made by IT personnel are inappropriate or unauthorized.*	4	4	4	3	15
5	*Multiple instances of the same IT application that should be identical are not the same.*	3	3	3	3	12
Manage Access		Impact	Frequency	Duration	Pressure	Total
6	*Users of the IT environment aren't the intended users due to inadequate authentication and security settings.*	4	2	2	3	11
7	*Access rights risks:* *'- Access granted to the IT environment (IT and Business) does not match the access approved* *- Access termination requests are not fulfilled timely* *- Access rights to the IT environment (IT and Business) do not remain appropriate over time.*	5	4	4	3	16
8	*Access requests for IT and business users of components of the IT environment are inappropriate*	4	4	4	3	15
9	*The access of IT users of the IT environment creates segregation of duties concerns.*	5	2	4	3	14
10	*Access to functions within the IT application is combined into roles. The access rights within the roles contain segregation of duties issues that could cause a material misstatement of the financial statements.*	5	2	4	3	14
11	*Direct data changes are made without authorization. (Of higher risk when there is routine use of direct data changes in the processing of transactions relevant to the financial statements.)*	5	3	4	3	15
Manage Operations		Impact	Frequency	Duration	Pressure	Total
12	*Hardware or software issues result in loss of data or the ability to accurately process that data.*	5	5	4	3	17
13	*Issues with programs that cannot process to completion are not addressed or are addressed inappropriately.*	4	5	4	3	16

Interview J2

Risk number: 1, 2 & 4

Risk description:

1. *New IT application programs or changes to existing programs, including reports, configurations and interfaces, do not function as described or requested because they are not adequately tested by appropriate persons.*
2. *New IT application programs or changes to the production IT application programs (including reports and interfaces) are not appropriate for the business or the IT environment.*
4. *Configuration changes made by IT personnel are inappropriate or unauthorized.*

Dependencies Risk 1, 2 & 4

Variable: degree of standardization

1. *Are the tasks standardized over multiple clients?*
    a. *Are there common systems/structures that this usually happens in? (that would allow an algorithm to learn from data from other clients, or allow the algorithm to be repurposed for other clients)*

There are systems that are reasonably standardized that are often used like SAP. Some companies use ticketing systems like ServiceNow. Most tasks can be reasonably standardized.

    b. *Are similar data fields used (even if names differ) for multiple clients? What kind of data is used in these tasks?*

The data and structure is very similar in most cases. The difference is in the wording and the IT maturity of the company. Some companies have for example the approval assigned to a certain person in a ticketing system which already enforces the right person to approve. Most companies are not at this stage however.

Variable: complexity

2. *Is the task complex?*
    a. *Can data be classified as anomalous or normal?*

For ML you could even think about an NLP solution that can easily access the emails and tickets.

    b. *Can this task be solved with RPA?*

Especially the email and ticketing approvals would be difficult to utilize with an RPA solution because of the slight differences in structure. The first few steps in the process might be possible with RPA. So getting a change log from Excel and performing standard

actions in Excel. A problem could be that in the email contact there are replies that simply say “ok” as a response, which means it needs context to be understood. This kind of data can make it difficult to be automated.

Variable: data quantity

3. *Does this task involve large data quantities?*

Companies have change logs, tickets and email interactions all in their systems. Especially at bigger companies this will be a large quantity. Data can also come from the clients vendors that send in reports that can be relevant to the situation. These exceptions are difficult to include in the data for anomaly detection. Purely data that can be retrieved from a system should be sufficient from bigger clients.

Variable: pressure

4. *Would a test on the full population add quality over testing samples?*

If anomaly detection could be used to look at the full data population, and send back a shortlist with anomalous instances that have to be tested instead of the usual samples, would this improve audit quality?

Definitely. In terms of quality you could go from a high degree of certainty to maybe even 100% certainty or near that. The current methodology however, dictates that a sample is taken. While the quality would definitely go up, it does not completely fit in the current methodology. When a mistake is found in the current methodology, the number of samples is increased. If nothing else is found, it can be regarded as an isolated incident. When you want to look at all data you first have to determine the population. This is simply a change log.

Variable: relationship with the client

5. *Would the relationship with the client be at risk?*
    a. *Would requesting more data deteriorate the relation?*

Depends on the way of extraction. If data can be extracted by simply connecting to the system and doing the extraction, it will be completely fine. At the moment, clients often have to connect to the system to extract evidence for every sample individually. If there is no option to extract everything at once it would mean a lot of extra work for the client. If it can be extracted directly then it could potentially even save the client time compared to the current approach. A lot of clients would simply give access to the system to avoid having to do the extraction of data themselves.

    b. *Would false positives deteriorate the relation?*

If you look at more data, it only makes sense that you will find more mistakes. Clients will not be happy if more mistakes in their reporting or processes are discovered. It might be useful feedback that something is not going well, but there is also sort of a political game that is played. If there are five instances where something goes wrong and all of them are discovered and relayed to the client, it will be experienced as less positive then when the positive instances are highlighted as well.

Solutions Risk 1, 2 & 4

Variable: degree of standardization

6. *Can data be standardized?*
    a. *Would clients be willing to work in a unified way if it could enable better audit?*

A lot of clients would likely be willing to make changes to the process concerning for example the email templates and tickets. Some clients could be unwilling and would rather decide for themselves how they approach the tasks. Also take into account that a large number of big clients are already quite standardized, so they can already be included.

    b. *Is the task big/important enough to warrant an algorithm that can detect an unknown pattern (unsupervised) for a single client?*

Yes, but with some clients being standardized already it might be good to look at a solution for multiple clients first. Looking at a client like Shell, there is so much work in this process just for Shell right now that it would definitely still be worth a dedicated solution.

Variable: complexity

7. *Could anomaly detection add something over RPA?*
8. *Is there a way data can be divided into normal or anomalous?*

The emails that can form a problem in complexity because of a lack of structure could be resolved by getting clients to use a standard template if there is not enough structure in the current process.

Variable: data quantity

9. *Can enough data be gathered by testing the entire data population instead of samples?*

Yes, the amount of data even with sampling is already high.

10. *Is task standardized enough to enable combination of data from multiple clients?*

Yes, for a large number of bigger clients.

Variable: pressure

11. Could anomaly detection be used to increase the efficiency instead?

There are already RPA tools being created and in use that can help to gather the relevant data from the client systems. This can already help with efficiency. The RPA solution extracts the relevant data from for example SAP where different information has to come from different systems or tables and can then be merges into one excel sheet. ML could be used similarly but go a few steps further.

Variable: relationship with the client

12. Could a solution like CCM reduce the amount of work that a client would have from delivering extra data?

CCM would be more looking at an implementation at the client side. This would likely require more something that runs on the IT Auditor side. If this could be implemented in something that can run on the client side and give feedback then it might still fit in the current way of working.

13. Would the clients be able/willing to quickly adapt to new requirements in logging and delivering evidence so that the false positives would decline over time?

This is dependent on the situation and client, but if it is in the interest of the client to improve and adapt then it will happen. If it is about something that is considered as insignificant, like a change in documentation that would be extra work to make more reliable while it would not bring much benefit to solve it, then clients might be less willing to adapt.

14. Would a different balance in recall/accuracy be a solution? (The balance between false positives and misstatements that slip through the algorithm)

Using only the anomalies that are significant enough could help to convince clients of the benefits of this new approach. A part of IT Audit is professional judgement. There can be something slightly off about the evidence, while a client will simply explain it by saying it was an exceptional situation and there was a reason why it had to be done this way. An algorithm will likely look at this data and see that something is off, but might lack the professional judgement to decide when to look past that.

Variable: open question

15. Do you have any other solutions that have not been mentioned?

When tools are developed for use in IT Audit, it will have to be certified before it can be used. How to handle in audit methodology that you now find five mistakes in a thousand

instances? Does this mean you have to switch to substantive because five mistakes have been found?

Feasibility Risk 1, 2 & 4

16. Are the solutions mentioned realistic/feasible?

a. Why (not)?

The data quantity should not be a limiting factor. Standardization can be improved by asking clients to make small changes, and even without it is already reasonably standardized. Whether or not it fits current IT Audit methodology is not sure.

Interview J3

Risk number: 12

Risk description: *Hardware or software issues result in loss of data or the ability to accurately process that data.*

Dependencies Risk 12

Variable: degree of standardization

1. *Are the tasks standardized over multiple clients?*
    a. *Are there common systems/structures that this usually happens in? (that would allow an algorithm to learn from data from other clients, or allow the algorithm to be repurposed for other clients)*

It is relatively similar. There is a tool that shows if there has been a completed backup in either a ticket or report. The tools usually have the option to extract data in a standard file-type like Excel, CSV or PDF.

    b. *Are similar data fields used (even if names differ) for multiple clients? What kind of data is used in these tasks?*

The data is either a ticket, or the outcome of a report containing whether or not there has been a completed back-up and if not then an explanation. The data is mostly the same.

Variable: complexity

2. *Is the task complex?*
    a. *Can data be classified as anomalous or normal?*

Yes, but it can be difficult to check if whether the backup has actually been checked or not.

    b. *Can this task be solved with RPA?*

The report of the backup can be RPA, but the documentation about a check that has been performed on the backup is less likely to work with RPA.

Variable: data quantity

3. *Does this task involve large data quantities?*

As this happens multiple times per day at big clients there is a lot of data.

Variable: pressure

4. *Would a test on the full population add quality over testing samples?*

It is not as obvious here whether quality would improve. It would give more insights perhaps, but it is about absolute questions like: “has the check been performed?”.

Variable: relationship with the client

5. *Would the relationship with the client be at risk?*
    a. *Would requesting more data deteriorate the relation?*

In most cases it would be no problem. If tools do not allow for easy data extraction or if everything is documented in emails that are not structured it might be more difficult.

    b. *Would false positives deteriorate the relation?*

Not as much on this risk, as it is easy to convince a client that this is an important subject with high impact. This makes it easier to convince clients to provide more insights. The employee who has to do the work might not appreciate it, but the IT manager would like to know when something goes wrong.

Solutions Risk 12

Variable: degree of standardization

6. *Can data be standardized?*
    a. *Would clients be willing to work in a unified way if it could enable better audit?*

Yes they would be willing to make changes if the tools allow for it. Not all clients would be in this situation.

Variable: complexity

7. *Could anomaly detection add something over RPA?*

If the data is standardized enough, the conclusion would be: If I can do it, then ML could do this too.

8. *Is there a way data can be divided into normal or anomalous?*

Yes. If something is handled in a timely manner and follows the protocol for example it would be normal.

Variable: data quantity

9. *Can enough data be gathered by testing the entire data population instead of samples?*

It is better to look to combine data from clients.

10. *Is task standardized enough to enable combination of data from multiple clients?*

In principle yes. The way of performing the review that the client follows is key here. But if they follow a similar way of working it is possible.

Variable: pressure

11. Could anomaly detection be used to increase the efficiency instead?

This is not very relevant here. It will not change the amount of work by much.

Variable: relationship with the client

12. Could a solution like CCM reduce the amount of work that a client would have from delivering extra data?

See interview B2 question 12.

13. Would the clients be able/willing to quickly adapt to new requirements in logging and delivering evidence so that the false positives would decline over time?

Yes, as this risk is important and clear to clients then may be more willing to adapt.

14. Would a different balance in recall/accuracy be a solution? (The balance between false positives and misstatements that slip through the algorithm)

No as the findings will be absolute.

Variable: open question

15. Do you have any other solutions that have not been mentioned?

No.

Feasibility Risk 12

16. Are the solutions mentioned realistic/feasible?

a. Why (not)?

The solutions for this risk seem feasible because this risk is highly cared for by companies.

Interview J4

Risk number: 7

Risk description: Access Right Risks:

- Access granted to the IT environment (IT and Business) does not match the access approved
- Access termination requests are not fulfilled timely
- Access rights to the IT environment (IT and Business) do not remain appropriate over time.

Dependencies Risk 7

Variable: degree of standardization

1. *Are the tasks standardized over multiple clients?*
   a. *Are there common systems/structures that this usually happens in? (that would allow an algorithm to learn from data from other clients, or allow the algorithm to be repurposed for other clients)*

The tasks for this risk can be viewed in separate parts. Most of the work is rather standardized and works in a similar way. A user list with roles and access information will be sent to the person responsible for those people to perform a check on the roles and authorization. Since it is often in emails or Excel files it is not really in the same systems.

   b. *Are similar data fields used (even if names differ) for multiple clients? What kind of data is used in these tasks?*

Yes, it is usually user lists with comparable information and structure, and emails or tickets to communicate the outcomes.

Variable: complexity

2. *Is the task complex?*
   a. *Can data be classified as anomalous or normal?*

Yes, but for some parts of the process it requires professional judgement. In comparison to other risks, for this one the first few steps can be improved with ML, then in the middle there is need for professional judgement, and the last few steps could be supported again.

   b. *Can this task be solved with RPA?*

A small part of the task could maybe be supported by RPA. Most of the task is reviewing what already happens which is difficult with RPA.

Variable: data quantity

3. *Does this task involve large data quantities?*

This process happens maybe once or twice per year at the client side. One instance is not much data compared to other risks. Big clients can have a high number of systems and users with access right however. The amount of emails can be high in these cases.

Variable: pressure

4. *Would a test on the full population add quality over testing samples?*

If anomaly detection could be used to look at the full data population, and send back a shortlist with anomalous instances that have to be tested instead of the usual samples, would this improve audit quality?

Right now, concessions have to be made in what can be reviewed.

Variable: relationship with the client

5. *Would the relationship with the client be at risk?*
    a. *Would requesting more data deteriorate the relation?*

The email contact might be difficult. If it can be organized in a way where it is not too much extra work it should not be a problem.

    b. *Would false positives deteriorate the relation?*

-

Solutions Risk 7

Variable: degree of standardization

6. *Can data be standardized?*
    a. *Would clients be willing to work in a unified way if it could enable better audit?*

Yes, creating templates for emails is rather small change.

    b. *Is the task big/important enough to warrant an algorithm that can detect an unknown pattern (unsupervised) for a single client?*

For some clients probably.

Variable: complexity

7. *Could anomaly detection add something over RPA?*

Yes RPA can only help with a few select steps where anomaly detection can go further.

8. *Is there a way data can be divided into normal or anomalous?*

There are most likely patterns or indicators that can be used, yes.

Variable: data quantity

9. *Can enough data be gathered by testing the entire data population instead of samples?*

Yes.

10. *Is task standardized enough to enable combination of data from multiple clients?*

Yes most likely for bigger clients.

Variable: pressure

11. *Could anomaly detection be used to increase the efficiency instead?*

Yes there would probably be a way to use anomaly detection to make the current process more efficient.

Variable: relationship with the client

12. *Could a solution like CCM reduce the amount of work that a client would have from delivering extra data?*

It depends on the perspective. For internal audit the purpose would be different than for external audit.

13. *Would the clients be able/willing to quickly adapt to new requirements in logging and delivering evidence so that the false positives would decline over time?*

Not applicable

14. *Would a different balance in recall/accuracy be a solution? (The balance between false positives and misstatements that slip through the algorithm)*

Not applicable

Variable: open question

15. *Do you have any other solutions that have not been mentioned?*

No

Feasibility Risk 7

16. *Are the solutions mentioned realistic/feasible?*
    a. *Why (not)?*

-

Appendix VI: Data Management Plan

The data management plan has been provided by the University of Turku as a mandatory inclusion in the thesis.

1. Research data

Research data refers to all the material with which the analysis and results of the research can be verified and reproduced. It may be, for example, various measurement results, data from surveys or interviews, recordings or videos, notes, software, source codes, biological samples, text samples, or collection data.

Research data type	**Contains personal details/information***	**I will gather/produce the data myself**	**Someone else has gathered/produced the data**	**Other notes**
Data type 1: Literature review		X		
Data type 2: *Exploratory interviews*				Participants have been anonymized
Data type 3: *Interviews*				Participants have been anonymized
Data type 4: *Internal documentation/information from company*				Not all sources may be publicly available

* Personal details/information are all information based on which a person can be identified directly or indirectly, for example by connecting a specific piece of data to another, which makes identification possible. For more information about what data is considered personal go to the Office of the Finnish Data Protection Ombudsman's website

2. Processing personal data in research

If your data contains personal details/information, you are obliged to comply with the EU's General Data Protection Regulation (GDPR) and the Finnish Data Protection Act. For data that

contains personal details, you must prepare a Data Protection Notice for your research participants and determine who is the controller for the research data.
I will prepare a Data Protection Notice** and give it to the research participants before collecting data ☒

The controller** for the personal details is the student themself ☐ the university ☐

My data does not contain any personal data ☒

** More information at the university's intranet page, Data Protection Guideline for Thesis Research

3. Permissions and rights related to the use of data

Find out what permissions and rights are involved in the use of the data. Consult your thesis supervisor, if necessary. Describe the use permissions and rights for each data type. You can add more data types to the list, if necessary.

3.1. Self-collected data

You may need separate permissions to use the data you collect or produce, both in research and in publishing the results. If you are archiving your data, remember to ask the research participants for the necessary permissions for archiving and further use of the data. Also, find out if the repository/archive you have selected requires written permissions from the participants.
Necessary permissions and how they are acquired

3.2 Data collected by someone else

Do you have the necessary permissions to use the data in your research and to publish the results? Are there copyright or licencing issues involved in the use of the data? Note, for example, that you may need permission to use the images or graphs you have found in publications.
Rights and licences related to the data

4. Storing the data during the research process

Where will you store your data during the research process?

In the university's network drive ☐
In the university-provided Seafile Cloud Service ☐
Other location, please specify: On the company network drive ☒

The university's data storage services will take care of data security and backup files automatically. If you choose to store your data somewhere other than in the services provided by the

university, please specify how you will ensure data security and file backups. Remember to make sure you know every time where you are saving the edited/modified data.
If you are using a smartphone to record anything, please check in advance where the audio or video will be saved. If you are using commercial cloud services (iCloud, Dropbox, Google Drive, etc.) and your data contains personal data, make sure the information you provide in the Data Protection Notice about data migration matches your device settings. The use of commercial cloud services means the data will be transferred to third countries outside the EU.

5. Documenting the data and metadata

How would you describe your research data so that even an outsider or a person unfamiliar with it will understand what the data is? How would you help yourself recall years later what your data consists of?

5.1 Data documentation

Can you describe what has happened to your research data during the research process? Data documentation is essential when you try to track any changes made to the data.
To document the data, I will use:
A field/research journal ☐
A separate document where I will record the main points of the data, such as changes made, phases of analysis, and significance of variables ☐
A readme file linked to the data that describes the main points of the data ☐
Other, please specify: ☒

5.2 Data arrangement and integrity

How will you keep your data in order and intact, as well as prevent any accidental changes to it?

I will keep the original data files separate from the data I am using in the research process, so that I can always revert back to the original, if need be. ☒

Version control: I will plan before starting the research how I will name the different data versions and I will adhere to the plan consistently. ☒

I recognise the life span of the data from the beginning of the research and am already prepared for situations, where the data can alter unnoticed, for example while recording, transcribing, downloading, or in data conversions from one file format to another, etc. ☒

5.3 Metadata

Metadata is a description of you research data. Based on metadata someone unfamiliar with your data will understand what it consists of. Metadata should include, among others, the file name, location, file size, and information about the producer of the data. Will you require metadata?

I will save my data into an archive or a repository that will take care of the metadata for me. ☒

I will have to create the metadata myself, because the archive/repository where I am uploading the data requires it. ☐

I will not store my data into a public archive/repository, and therefore I will not need to create any metadata. ☒

6. Data after completing the research

You are responsible for the data even after the research process has ended. Make sure you will handle the data according to the agreements you have made. The university recommends a general retention period of five (5) years, with an exception for medical research data, where the retention period is 15 years. Personal data can only be stored as long as it is necessary. If you have agreed to destroy the data after a set time period, you are responsible for destroying the data, even if you no longer are a student at the university. Likewise, when using the university's online storage services, destroying the data is your responsibility.
What happens to your research data, when the research is completed?

I will store all data for 1 year.

I will destroy all data immediately after completion, because:

I will destroy part of the data, but store part of it for 1 year1, because: Some data might be required if any changes have to be made to the document or when questions are asked.

If you will store the data, please identify where: Company network drive

Remember to keep the data management plan updated throughout the research project.