

Towards the Future of Work: Managing the Risks of AI and Automation

By

James Man

B.Sc. Computer Engineering, University of Alberta, 2005
M.Eng. Electrical and Computer Engineering, University of Alberta, 2007
MBA, HEC Paris, 2013
MBA, Tsinghua University, 2013

SUBMITTED TO THE MIT SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MANAGEMENT OF TECHNOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MAY 2022

©2022 James Man. All rights reserved.

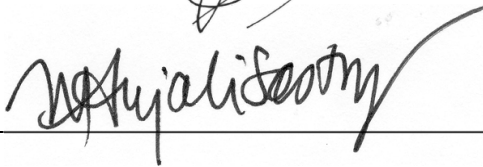
The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part
in any medium now known or hereafter created.

Signature of Author: _____



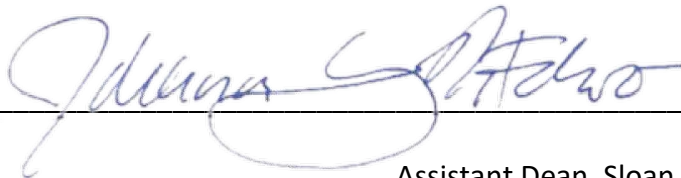
MIT Sloan School of Management
May 6, 2022

Certified by: _____



Anjali Sastry
Senior Lecturer
Thesis Supervisor

Accepted by: _____



5/5/22
Johanna Hising DiFabio
Assistant Dean, Sloan Fellows and EMBA Programs
MIT Sloan School of Management

THIS PAGE INTENTIONALLY LEFT BLANK

Towards the Future of Work: Managing the Risks of AI and Automation

By

James Man

Submitted to MIT Sloan School of Management
on May 6, 2022 in Partial Fulfillment of the
requirements for the Degree of Master of Science in Management of Technology.

ABSTRACT

Many believe in a vision of the future where almost all work is automated. A first step already underway involves Robotic Process Automation (RPA) technology, which firms use to automate standardized computer work. The larger step that needs to be taken towards this vision lies in connecting RPA to AI, so that Machine Learning (ML) algorithms can be used to automate human “intelligence” and decision making in companies.

Management research surrounding the concept of Intelligent Automation (IA) is nascent and spans multiple domains. This thesis consolidates the fragmented research landscape through a Systematic Literature Review to address four research questions: 1) What use cases are IA fulfilling? 2) Which ML algorithms and technologies are employed? 3) What risks are associated with IA? and 4) What risk mitigation techniques are there? The findings paint a picture of what is needed to advance the value that IA delivers to firms and shore up professional practices.

Results show that the bulk (66%) of cases centered on document processing and chatbots. ML models, tended to be uninterpretable, posing transparency and risk challenges. The systematic coding of 77 key sources yielded 36 risks that fell into eight clusters that are explored in depth. Corresponding risk mitigation measures covered far less ground, leaving many risks unaddressed. The risk registry derived in this thesis offers a starting point for a structured approach to managing emergent risks necessary for IA to deliver on its promise to improve work.

Thesis Supervisor: Anjali Sastry

Title: Senior Lecturer

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgements

I would like to thank my thesis supervisor, Anjali Sastry for her guidance throughout this research. In addition, I would like to thank Jonathan Ruane for his valuable input regarding possible research methodologies and academic rigor leading up to this thesis. Finally, I would like to thank my program advisor Becca Souza and my former work supervisor Bindi Basan for their support before and throughout the thesis writing process.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

Chapter 1: Introduction	11
1.1 The Automation of Work	11
1.2 Robotic Process Automation (RPA)	11
1.3 Intelligent Automation (IA) and Risks	13
1.4 Research Questions	14
1.5 Research Goal	15
Chapter 2: Research Methodology	15
2.1 Literature Selection Process	16
2.2 Search Term Keyword Selection	16
2.3 Search Results	17
2.4 Qualitative Coding Scheme	18
2.4.1 Relevance Scores	18
2.4.2 Text Highlighting Method for Free-Text Capture	18
2.4.3 Research Question 1 (IA Use Cases) Coding	19
2.4.4 Research Question 2 (ML Algorithms and Technologies) Coding	19
2.4.5 Research Question 3 (IA Risks) Coding	20
2.4.6 Research Question 4 (IA Risk Mitigation Techniques) Coding	20
2.5 Chapter 2 Summary	20
Chapter 3: Data Analysis	20
3.1 Basic Summary Statistics	21
3.2 Research Question 1 (IA Use Cases)	23
3.3 Research Question 2 (ML Algorithms and Technologies)	26
3.4 Research Question 3 (IA Risks)	29
3.5 Research Question 4 (IA Risk Mitigation Techniques)	32
3.6 Chapter 3 Summary	34

Chapter 4: Data Discussion	34
4.1 IA Use Case Riskiness	34
4.2 Machine Learning Algorithms	35
4.2.1 Measuring Performance of Classifiers – Confusion Matrices.....	36
4.2.2 Confidence Intervals	37
4.2.3 Interpretability	38
4.3 Machine Learning Technologies.....	39
4.4 Risks.....	41
4.4.1 Socio-Organizational Risks	41
4.4.2 Operational Risks	47
4.5 Risk Mitigation Techniques	53
4.5.1 Planning and Due Diligence	54
4.5.2 Algorithm Selection.....	57
4.5.3 Human Interaction Design	58
4.5.4 Operations	60
4.6 Chapter 4 Summary.....	62
Chapter 5: Risk Register Development	63
5.1 Risk to Risk Mitigation Mappings	63
5.2 Governance	63
5.2.1 Define IA Specific Process Selection Criteria	64
5.2.2 Leverage Existing Policies	64
5.2.3 Measure Employee Impact	65
5.2.4 Set Baselines and Monitor Data	65
5.2.5 Document.....	66
5.3 Mitigating Unaddressed Risks	66

5.3.1 Unaddressed Risk 1 - Departmental Resistance	66
5.3.2 Unaddressed Risk 2 - Loss of Job Meaning	67
5.3.3 Unaddressed Risk 3 - Loss of Job Security	67
5.3.4 Unaddressed Risk 4 - Mistrust in Management	68
5.3.5 Unaddressed Risk 5 - Reduced Work Preparedness.....	68
5.3.6 Unaddressed Risk 6 - Transfer Learning Bias	68
5.4 Final Risk Register.....	69
5.5 Chapter 5 Summary.....	69
Chapter 6: Conclusion and Final Discussion	69
6.1 Research Summary.....	70
6.2 Research Limitations	71
6.3 Future Research	72
6.4 Final Discussion	73
Appendix	75
A. Exact Search Terms used in Document Search Databases	75
B. Summary of SLR Search Results	75
C. Research Question 1 (IA Use Cases) Coding Scheme.....	75
D. Research Question 2 (ML Algorithms and Technologies) Coding Scheme.....	77
E. Research Question 3 (IA Risks) Coding Scheme	77
F. Research Question 4 (IA Risk Mitigation Techniques) Coding Scheme.....	78
G. Coded Data	78
G.1 Overall Relevance, Publication Year and Publication Type.....	78
G.2 RQ1 Coded Data	81
G.3 RQ2 Coded Data	84
G.4 RQ3 Coded Data	87

G.5 RQ4 Coded Data	89
H. Number of Research Papers Addressing Specific Combinations of RQs	93
I. Risk Levels of Real IA Use Cases (Relevance Score of 2)	93
J. Implied Risk to Risk Mitigation Mappings in the Literature	94
K. Risk to Risk Mitigation Mappings.....	95
L. IA Risk Register	97
References	100

Chapter 1: Introduction

“The Robots are Coming for Phil from Accounting. Workers with college degrees and specialized training once felt relatively safe from automation. They aren’t.” – the New York Times [1].

For the past ten years, two automation trends have been quietly revolutionizing the ways that firms produce their work. The first was the automation of back-office computer work through Robotic Process Automation (RPA) technology. The second trend, currently underway, looks to automate work once thought to be safe from automation. This is the automation of worker intelligence and decision-making, through combining automated computer work with Artificial Intelligence. Firms will need a clear understanding of the risks involved as they push towards achieving the “intelligent automation” of their enterprise.

1.1 The Automation of Work

Businesses have been trying to automate their work for over 100 years [2], to gain operating efficiencies and improve revenues. Manufacturing has become more computerized and mechanized, resulting in better business outcomes, such as faster production, fewer manufacturing defects, lowered production costs [3], and improvements in worker’s lives such as increased safety. As the more obvious “physical tasks” have become automated, automation attention has been turned towards “white-collar” or “knowledge-based” work processes. Examples of such work processes are employee onboarding for human resource management and account reconciliation for accounting. From experience, these processes are among some of the “easier” processes to automate due to a high degree of standardization. They are often among the first-wave of processes that are automated by firms.

Even for these “easier” business processes, there are still numerous difficulties that must be overcome. One key barrier is the presence of legacy IT systems that underpin much of the infrastructure that the world runs on today. When there are no modern Application Programming Interfaces to talk to these legacy systems, humans are needed to translate work between these old systems and the modern world. Robotic Process Automation technology is able to perform this translation on our behalf in an automated manner.

1.2 Robotic Process Automation (RPA)

RPA technology has gained prominence within the past 10 years as a real solution to enable computer-based automation even when legacy computer systems are involved [4]. RPA is able to work with any computer system due to how it interfaces with applications. When APIs are unavailable, RPA communicates with applications the same way that people do – through the Graphical User Interface (GUI), by recognizing what important GUI elements are on the screen, such as form fields, dropdown lists and buttons.

In addition to the visual-based interfacing technology, RPA vendors provide a software development platform that allow companies to create low-code “programs” that will perform steps at the computer the same way that an employee would, to complete a business process. These automation programs can include elements such as launching applications, moving the mouse, typing on the keyboard, identifying buttons on the screen, pasting data into forms and sending emails etc. [5]. In this way, we are able to replace an employee’s computer-based tasks with a software “robot” that can manipulate a computer exactly the same way we do. Current research suggests that one RPA “robot” can do the equivalent work of three to five humans on average [4], since they can work faster and during off-hours – as long as there is enough work.

The benefits of RPA technology have been well-studied and include improving accuracy, productivity, consistency and efficiency of work done on a computer [6], while reducing the number of errors present in completed work. RPA also allows humans to reduce the amount of time spent doing undesirable, rote computer work, giving them a chance to perform work that cannot yet be replaced by robots involving creativity, communication and compassion. As of 2020, it is considered to be the most rapidly growing segment of the global software market [7].

RPA has become an industry of its own, and has found significant success across almost all sectors and functions. This is especially true for highly-regulated settings with standardized procedures such as government, finance, insurance, telecommunications, utilities and healthcare [7], due to RPA’s consistency of executed process steps and ability to provide audit trails of exactly what steps were executed. Research suggests that RPA has the fastest return on investment of any enterprise technology that has been studied since 2017 [8]. Microsoft has recognized the importance of this technology to business and has acquired an RPA firm in November, 2019 [9].

RPA does have limitations, for example, the process steps that are to be automated should be well-defined, stable and highly rule-based. This limits the scope of which business processes can be automated, as much of the work that happens today still requires non-rule-based, human decision making. Around 10% to 40% of a firm’s total number of business processes are thought to be automatable by RPA technology alone, but with advances in Machine Learning (ML), some predict that eventually 100% of processes could be automated [10]. Despite the tremendous uptake in industry, RPA is a field with a notable lack of scientific research [7], [11].

Combining ML and RPA extends the range of what is automatable beyond standardized and rule-based processes. Knowledge-based tasks involving complex decision making can potentially be automated as well. This natural evolution of pairing RPA with AI technologies has already received attention the World Economic Forum [12] due to its predicted impact on society at large. In industry, combining RPA and ML is known under many different terms, such as Intelligent Process Automation, Hyperautomation, Cognitive Automation or Intelligent Automation (IA). I refer to these terms collectively as “Intelligent Automation”, meaning the combination of RPA technology with Machine Learning.

1.3 Intelligent Automation (IA) and Risks

“Physical labor was replaced by robots; mental labor is going to be replaced by AI and software.” – Andrew Yang, US politician [13].

The term IA can take on different meaning depending on the industry and context. To some, IA may mean injecting AI functionality into the RPA products themselves so that they are easier to develop automations in. To others, it may mean having an AI that can predict and recover from errors during the automated execution. In this thesis, IA specifically means the combined use of RPA technology with ML for the purpose of **automating cognitive work that is normally done by a person**. Under traditional RPA, only well-defined rule-based automation or processes using structured data as input is possible. When ML is used together with RPA, human-like decision making and analysis of semi-structured or unstructured data is enabled.

IA is a relatively new field both in industry and in academia [14] and is on the cutting edge of business transformation [15]. The IEEE Standards Association released its “Guide for Taxonomy for Intelligent Process Automation Product Features and Functionality” in July 2019 [16], outlining the importance of IA as a developing sector. Interest in IA is also being accelerated due to advances in machine learning research [17]. Because of this, businesses are increasingly combining RPA with AI to automate human-decision making in their processes. As a response, the three main technology vendors for RPA: Automation Anywhere, Blue Prism and UiPath have all launched ML-based capabilities to their product line-up within the past two years [18], [19], [20]. Industry research firm Gartner predicts that IA will be a \$600 billion market in 2022, and that IA is not a “nice to have”, but a condition for the survival of firms [21].

This push towards IA poses challenges that are absent under traditional RPA. Under traditional RPA, work output is deterministic. This allows employees to remain relatively hands-off, as intervention is rarely needed while the robot completes its work. A human worker only needs to know whether the work performed by the RPA robot has successfully completed or not. If it was completed, the work can be assumed to be done correctly. With IA, the possibility that completed work is done incorrectly must be considered and human intervention is often added back into the business process. A deterministic outcome under RPA becomes a probabilistic outcome under IA.

The world is turning towards IA as a solution to some of its largest challenges: business continuity due to labour shortages caused by COVID-19, an aging workforce and the Great Resignation [17], [22]. In turn, institutions will also face many challenges to meet society’s demands for IA. Some of these challenges include the lack of employees with necessary education and skillsets [14], research into explainable AI methods and the development of regulatory guidelines outlining the appropriate use of ML in daily business operations. But perhaps a more pressing issue that must be addressed is to first understand what risks IA poses to firms to begin with.

1.4 Research Questions

“[T]here are known knowns . . . there are also unknown unknowns—the ones we don't know we don't know . . . it is the latter category that tends to be the difficult ones.” – Donald Rumsfeld on the topic of risks [23].

I came to this thesis topic as a result of my own professional experience and intellectual curiosity, seizing on the opportunity afforded by the yearlong residential Sloan Fellows program at the Massachusetts Institute of Technology. Having spent four years providing detailed technical and strategic guidance and execution in the domain of automation for dozens of firms, I have yet to encounter a list of possible risks that IA implementation entails, let alone a structured methodology to evaluate the potential implications for business and society. The present thesis aims to fill the breach by taking on the challenge of turning some of IA's “unknown unknowns” into something that is better known. A refined understanding of risks that are currently under-appreciated and under-studied will contribute greatly to the adoption and success of IA in firms.

Since I did not find any industry-led guidance on the specific risks posed by IA, I turned to academia to see whether the topic of risks and IA has been addressed there. While some research has been published surrounding the risks of traditional RPA implementation [24], I have found few examples of what risks exist for IA. Others [25] have also noted the lack of research into **risk management in IA**.

Two key components of risks management are 1) identifying which risks exist and 2) understanding what risk mitigation measures are possible. I have designed four research questions to discover these two points within the context of IA. These four questions are shown in the table below. Answers to RQ1 and RQ2 provide the context needed to understand the current IA ecosystem and give us an idea of what risks and risk mitigation methods are possible. RQ3 and RQ4 map directly to identifying the risks and risk mitigation techniques.

Table 1: Research Questions

ID	Research Question	Objective
RQ1	What high-level machine learning use cases are being used in intelligent automation?	Understanding the high-level use cases will help determine what types of risks exist and what types of risk mitigation are possible.
RQ2	Which specific machine learning algorithms and technologies are being used in intelligent automation?	Understanding which algorithms are being used will help to determine what types of risks exist and what types of risk mitigation are possible.
RQ3	What are the risks of intelligent automation to the firm?	To create a list of risks that can be used as a risk register for firms looking to use IA in their organisations.
RQ4	What risk mitigation techniques have been put to use when implementing intelligent automation?	To create a list of risk mitigation techniques that can be implemented based on the characteristics of their business process and the machine learning algorithms used.

All four research questions are interrelated; RQ2 (algorithms and technologies) follows naturally from RQ1 (use cases), and RQ4 (risk mitigation techniques) follows naturally from RQ3 (risks). Specific IA use cases (RQ1) can also lead to the use of specific technologies which may lead to certain risks and constrain which risk mitigation techniques are available. The selected algorithms and technologies (RQ2) may also do the same. Note that RQ3 and RQ4 aim to surface the risks of implementing IA for firms generally speaking, as opposed to using IA for risk mitigation purposes, which is a commonly cited use case for ML [26], [27].

Answers to the four RQs are then used to design the final contribution of this thesis, which is to develop a practical risk register that firms can use to assess, track and control the risks of their IA projects.

1.5 Research Goal

In answering the four research questions, I provide companies looking to implement IA with a specific list of risks that they may face as well as specific guidance on how those risks can be mitigated. Understanding the risks are a key part of any project's planning process, and controlling the risks greatly increase the odds of achieving project goals. The mapping between the risks and risk mitigation techniques is presented in the form of a risk register which can be customized to suit the needs of the firm.

The resulting risk register provides firms with a practical starting point for understanding and addressing the risks posed by IA. To enable the next steps required to deliver better solutions, this thesis also aims to catalyse the adoption of this transformative technology throughout industry.

The thesis is separated into five additional chapters. **Chapter 2: Research Methodology** discusses the chosen research methodology and data collection process. **Chapter 3: Data Analysis** presents a high-level summary of the collected data along with answers to the four RQs. **Chapter 4: Data Discussion** digs deeper into the identified use cases (RQ1), technologies (RQ2), algorithms (RQ2), and clusters the risks (RQ3) and risk mitigation techniques (RQ4) for a clearer understanding and analysis. In **Chapter 5: Risk Register Development**, I develop the risk register that is meant for practical use. Finally, **Chapter 6: Conclusion and Final Discussion** provides a summary of the research, and closing thoughts on IA and its role in the future of work.

Chapter 2: Research Methodology

Systematic literature reviews (SLR) have a rich history of use in business and management research [28] and remains a popular research methodology today. A SLR outlines a transparent and reproducible method for collecting, analysing and synthesizing data across a wide body of existing literature [11]. The IA field is situated across a broad range of disciplines, including management, information systems, computing science, economics and social sciences. I selected

SLR as the best methodology to consolidate research across numerous different disciplines and find answers to the four research questions.

Naturally, the SLR methodology has already been selected by other authors to study the topic of IA [29], [30], [17], [11]. However, the research questions posed in those SLRs do not overlap with the questions posed in my thesis. The usual four-step methodology of conducting a SLR that is also used in this thesis [31] is 1) Data collection, 2) Data coding, 3) Data analysis and 4) Interpretation of coded content.

2.1 Literature Selection Process

I chose peer-reviewed journal articles and conference papers from the electronic databases of Web of Science and ScienceDirect as the main sources of data. These two databases were selected due to their cross-disciplinary nature and their focus on peer-reviewed research. As IA research is rather new with a limited number of articles, no constraints were imposed on the journal quality nor the impact of the articles themselves. Only English language articles with open full-text available were considered. The first academic paper published with the keyword “robotic process automation” appeared in 2016, so the publication date range was set from January 1, 2015 to December 31, 2021. “Grey literature” such as industry-led publications from technology consulting firms and software vendors are not examined in this SLR. The specific document inclusion and exclusion criteria used are shown in **Table 2: Systematic Literature Review Document Inclusion Criteria** and **Table 3: Systematic Literature Review Document Exclusion Criteria** below.

Table 2: Systematic Literature Review Document Inclusion Criteria

Inclusion Criteria	Description
Dates	From January 1, 2015 to December 31, 2021
Languages	English
Publication Types	Peer-reviewed journal papers, conference papers, book sections
Search Engines	Web of Science and Science Direct

Table 3: Systematic Literature Review Document Exclusion Criteria

Exclusion Criteria	Description
Publication Types	Industry white papers, non-peer-reviewed papers, webpage articles
Topics	Any topics unrelated to RPA technology, or when RPA and ML are treated separately

2.2 Search Term Keyword Selection

As an industry expert, I know the different terminology used in business referring to the combination of computer automation and AI. The exhaustive list of industry terminology that I have encountered in the field was included in the search terms of the databases. I was however unaware of which terms were being used in academia at the beginning of this research. To

address this, I performed two weeks of exploratory searching through the two databases and read papers to uncover search terms that capture the meaning of combining RPA with ML, that are specific to academia.

The search terms in the list below represent all of the terms that I have encountered both in industry and academia that refer to the combination of RPA and ML. This list of search terms was created to capture the concept of IA as opposed to just traditional RPA. Quotation marks around the terms indicates that the entire content inside the quotation marks should be present.

- “rpa” “machine learning”
 - The “machine learning” term is added to focus the search on combining RPA with machine learning, given that the terms IA, hyperautomation, etc. are industry terms that have only emerged within the past three years
- “robotic process automation” “machine learning”
- “hyperautomation”
- “intelligent automation” “rpa”
 - The reason why term “intelligent automation” is not used alone is because it appears in many contexts unrelated to robotic process automation which leads to false positives, for instance, in manufacturing
- “intelligent process automation”
 - This term is used primarily in academia rather than in industry
- “cognitive automation”
 - This term is used primarily in academia rather than in industry

Readers wanting to reproduce the search results exactly for their own analysis can make use of the exact search queries for each search database, found in: **Appendix: A. Exact Search Terms used in Document Search Databases.**

2.3 Search Results

The web searches were performed on January 2nd, 2022 in order to capture the complete set of published results for year 2021. Searching was performed in the title, abstract and full text of the papers. The initial search yielded 539 results, 72 of them being duplicates, leaving a total of 467 articles to be considered. The abstracts and titles of the remaining 467 articles were analysed, to either explicitly include or exclude articles unrelated to IA technology. 39 articles were initially accepted and 332 were rejected, leaving 96 papers as undecided to fully review. The most common reason for rejection was the presence of the term “intelligent automation” in manufacturing, and the term “RPA” used to mean “remotely piloted aircraft”, and numerous other medical abbreviations.

After fully reviewing the 96 papers, I accepted an additional 14 articles, giving a count of 53 papers. From these 53 papers, forward and backward searching based on the citations were performed. The forward search was conducted based on the references of the 53 papers, yielding

3 additional papers. The backwards search was conducted using the “cited by” functionality in Google Scholar, which yielded 21 additional papers. A grand total of 77 papers were used in the SLR. A summary of this search process can be found in: **Appendix: B. Summary of SLR Search Results**.

2.4 Qualitative Coding Scheme

Coding schemes should be taken from existing research, developed by researchers in conjunction with field experts, or developed by researchers who are field experts themselves [28]. Reusing an existing coding scheme was not possible since no previous SLR has tried to address the research questions raised in this thesis. As a field expert with applied experience and requisite knowledge in both RPA and ML, I drew on my own professional and academic expertise to develop a coding methodology without needing to consult external experts.

The guiding principle behind the choice of coding fields was to capture data that is relevant to discovering the risks and risk mitigation measures of IA. Following a systematic protocol, I read each of the 77 articles three times: an initial pass to gather high-level notes, then twice more to perform coding and to ensure coding reliability.

2.4.1 Relevance Scores

I developed a simple metric to enable a structured assessment of this wide-ranging literature. A qualitative measure, which I dub the **Relevance Score**, captures high-level metrics for my research questions for every research paper in the SLR (77 papers * 4 RQs = 308 Relevance Scores captured). Each Relevance Score item can take on an integer value of 0, 1 or 2. A 0-valued Relevance Score indicates that a particular research paper did not address a specific research question in a meaningful way. A Relevance Score of 1 means that the research paper only addressed the research question in passing or partially. A Relevance Score of 2 means that the research paper directly addressed the specific research question.

2.4.2 Text Highlighting Method for Free-Text Capture

Many of the coded fields, such as industry (captured across every RQ), ML algorithms (RQ2), technologies (RQ2), risks (RQ3) and risk mitigation techniques (RQ4) are captured in free-text through an open-coding approach [32]. While reading through each paper, I highlighted words and sentences pertaining to one of the free-text categories using Zotero’s PDF reading plugin. Once all of the highlighting was complete, all of the words and sentences were copied to a spreadsheet for a high-level analysis. Sentences were clustered and inductively categorized into headings. Determining the heading granularity for free-text coding was a challenge. When thinking about whether to combine sentences into a heading or to create a separate heading, the main criteria used was to avoid a loss of meaning from the authors’ original intent.

2.4.3 Research Question 1 (IA Use Cases) Coding

Coding RQ1 required drawing from a list of high-level ML use cases. As no comprehensive list of use cases exist, I started to develop one incrementally through a first reading of each of the research papers. A second reading was performed to do the scoring and a third reading was done to ensure consistency of the scoring method. Many papers discussed ML use cases independently of automation. Those use cases were explicitly not recorded to ensure that only IA use cases were captured. The full list of coded fields for RQ1 are listed in **Appendix: C. Research Question 1 (IA Use Cases) Coding Scheme**.

Some of the use cases are interrelated, for instance, “Named Entity Recognition” is considered a subset of “Natural Language Processing” and “Email Classification” is likely a specific form of “Multi-Class Classification”. When capturing the use case names from the articles, I retained the **most specific category** for each use case example, in order to **preserve the most amount of detail**. For example, if “Facial Recognition” is mentioned in the article, it will be recorded as a use case for “Facial Recognition” and not for “Object Recognition” or “Computer Vision”.

2.4.4 Research Question 2 (ML Algorithms and Technologies) Coding

RQ2 aims to capture both the ML algorithms used, and which specific technologies were chosen to implement those algorithms. For the coding of ML algorithms, I first generated a list of widely used algorithms by reviewing research articles and book chapters related to summarizing the ML algorithm landscape [33], [34], [35]. Next, I consolidated the lists from the three sources and removed the duplicate algorithms. This resulted in the list of possible values that will be used for coding the ML algorithms, shown in the table below. Since I have industry knowledge, an existing background in traditional ML and have very recently taken courses on analytics algorithms and deep learning as part of my coursework at MIT, I believe that the list is suitable for the purpose of answering the algorithms portion of RQ2 with enough detail.

Table 4: List of Machine Learning Algorithms for RQ2 Coding Scheme

Artificial Neural Networks	Bayesian Networks	Convolutional Neural Networks
Decision Trees	Deep Learning	Genetic Programming
Gradient Boosting	K-Means Clustering	K-Nearest Neighbours
Linear Discriminant Analysis	Linear Regression	Logistic Regression
Long Short Term Memory	Naïve Bayes	Perceptrons
Random Forests	Recurrent Neural Networks	Support Vector Machines

Similar to the use cases, many of the algorithms are interrelated, for instance “Long Short Term Memory” is a type of “Recurrent Neural Networks”, which is a type of “Artificial Neural

Networks”. When capturing the algorithm names from research articles, only most specific version of the algorithm is kept, in order to preserve the most amount of detail. For instance, if “Recurrent Neural Networks” is recorded, then “Artificial Neural Network” will not, unless it refers to a completely separate use of the algorithm.

I performed the coding of ML technologies using the text-highlighting method described in **Section 2.4 Qualitative Coding Scheme**. During this coding process, each potential technology was double-checked to actually be a programmatic library, or commercial ML tool through online searching. The full list of coded fields, including the algorithms and technologies for RQ2 can be found in **Appendix: D. Research Question 2 (ML Algorithms and Technologies) Coding Scheme**.

2.4.5 Research Question 3 (IA Risks) Coding

Coding of the risks for RQ3 was done using free-text as there is no comprehensive list of risks that IA poses to the firms. During the capture process, the exact text used to describe the risks was recorded. After all of the risk descriptions were collected, they were analysed for common meaning and grouped under headings for further discussion as described in **Section 2.4 Qualitative Coding Scheme**. A full list of coded fields can be found in **Appendix: E. Research Question 3 (IA Risks) Coding Scheme**.

2.4.6 Research Question 4 (IA Risk Mitigation Techniques) Coding

The risk mitigation methods for RQ4 were coded in free text. After all of the descriptions were collected, they were analysed for common meaning and grouped under headings for further discussion. The full list of coded fields can be found in **Appendix: F. Research Question 4 (IA Risk Mitigation Techniques) Coding Scheme**.

2.5 Chapter 2 Summary

In this chapter, I outlined the chosen research methodology (SLR), the sources of data (journal databases) and the specific, reproduceable criteria for searching. Then, I briefly described the search process and results, which yielded 77 papers. Finally, I described the coding methodology for each of the four RQs. The next chapter discusses the results of the data coding procedure.

Chapter 3: Data Analysis

Section 3.1 Basic Summary Statistics of this chapter describes the results of the data coding process that was captured across all RQs and provides evidence on why this research is important. Following that, **Sections 3.2 Research Question 1 (IA Use Cases)** to **3.5 Research**

Question 4 (IA Risk Mitigation Techniques) describe the RQ specific coding results. A full table of coded values extracted from the 77 research articles which was used as the base data for this analysis can be found in **Appendix: G. Coded Data**.

3.1 Basic Summary Statistics

This section describes the basic summary statistics that were collected across all four RQs. 50 out of the 77 of the papers (65%) came from academic journals compared to 19 (25%) from conference proceedings. The remaining eight publications (10%) were books, book sections or reports. The first year where IA topics appeared in academic literature was in 2018, roughly two years after the first appearance of the term “robotic process automation” in literature. Researchers looking to study IA in the future can narrow their beginning search range to start in 2018 to lower the number of false positive search results. The number of IA-related publications is steadily growing from year to year, indicating increased interest in this research area.

Table 5: Search Results by Year

Year	Number of Publications
2018	6
2019	9
2020	24
2021	38

Twelve industries were captured across the four RQs. The top three industries were Finance, followed by Insurance with Education and Accounting/Auditing tied for third place. The presence of Finance, Insurance and Accounting is intuitive, as they were also early adopters of traditional RPA technology. Education is a less common industry to hear discussed in terms of RPA adoption. One potential reason why Education has started to climb the list of overall industry references is due to a higher number of IA use cases involving chatbots deployed to interact with students and computer vision used to track student activity during online learning. Overall, a broad range of industries were represented in the SLR, reinforcing the view that IA technology is appealing to most industries.

Table 6: Industry Counts Across all Research Questions

Industry	Count
Accounting/Auditing	9
Education	9
Finance	13
Forestry	2
Government	1

Healthcare	8
HR	6
Insurance	11
IT	6
Logistics	4
Sports	1
Utilities	4

After completing the coding process, an additional derived column named **Overall Relevance** was created by summing the Relevance Scores for each RQ. This “Overall Relevance” value (from 0 to 8) represents the overall relevance that the current body of IA research has to the four RQs posed in this thesis. Papers with an “Overall Relevance” score of “0” are those that met all of the SLR search inclusion criteria but did not respond to any of the RQs meaningfully.

A histogram of the “Overall Relevance” Score is plotted below in **Figure 1: Histogram of the Overall Relevance Score**. 29% (22/77) of the SLR papers received a zero “Overall Relevance” score. The highest value for “Overall Relevance” was 6, indicating that the holistic evaluation of IA risks and risk mitigation methods that examines the use cases, algorithms and technologies does not exist.

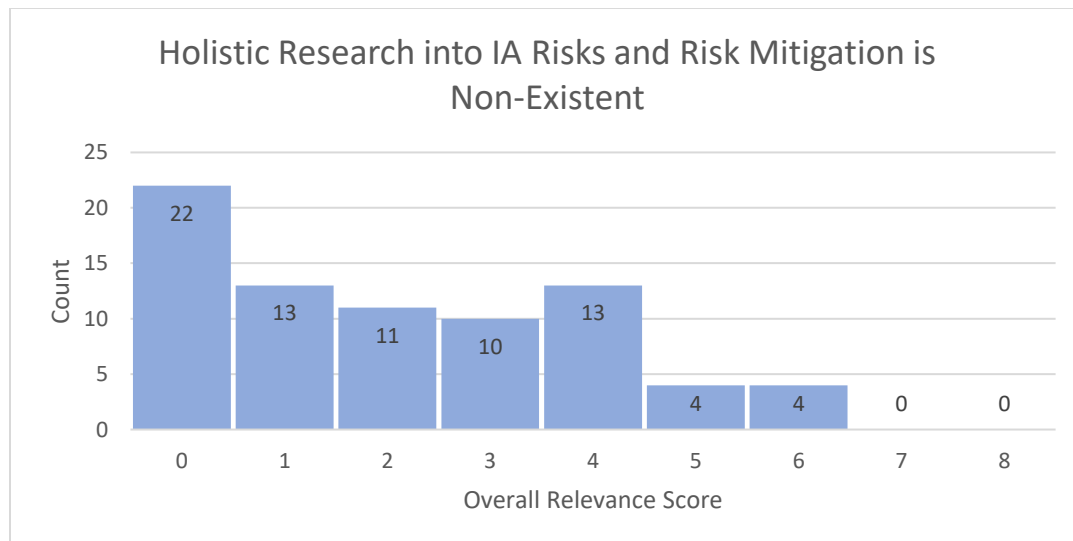


Figure 1: Histogram of the Overall Relevance Score (Source: Author)

Of the 55 papers that had an “Overall Relevance” score of at least 1, discussion of the use cases and technologies (RQ1 and RQ2) accounted for 58% of the total value. Risks and risk management have not been the focus of academic research compared to use cases and technologies. Fewer papers had relevance to RQ2 compared to RQ1, and similarly for RQ4 and RQ3. This means that more use cases are being discussed without diving into the actual

implementations, and more risks are discussed without mentioning the means in which they can be mitigated.

When looking at how the existing IA research has addressed specific combinations of research questions (see **Appendix: H. Number of Research Papers Addressing Specific Combinations of RQs**), I found that only 8% (6/77) addressed all four RQs to some degree. The conditional probability that research answered RQ2 (algorithms or technologies) given RQ1 (use cases) was 61% (25/41). When a research paper discusses risks (RQ3), an accompanying discussion of risk mitigation measures (RQ4) was found only 62% (16/26) of the time. Of the 41 papers that had answered either RQ1 or RQ2, 23 of them (56%) did not mention any risks or risk management at all.

The collection of use cases is a natural starting point for academic research into nascent fields, especially applied ones such as IA. This has been reflected in the derived “Overall Relevance” score. With the large number of collected use cases, it is appropriate to take the thinking further and use them to develop new theories for use in the field. The use cases need to be more systematically linked to the technologies, algorithms, risks and risk mitigation measures. The data collected for this section provides evidence towards the presence of missing links in the existing body of research. These missing links prevent firms and field experts from understanding how they can control the risks of implementing IA. This reinforces the need for the research that I am conducting in this thesis.

3.2 Research Question 1 (IA Use Cases)

This section presents data from the four coded fields, the Relevance, the Industry, the Use Cases and the Positioning. 53% of the examined papers contained use cases. As IA is still a relatively new field in both industry and in academia, the demand for research illustrating the actual use cases is still high. There was an equal proportion of research papers examining real IA use cases, as well as research referencing or suggesting theoretical use cases of IA. The definition of theoretical is that the use case was not implemented in reality.

Table 7: RQ1 – Relevance Scores

RQ1 Relevance	Count
0 – No relevance as there is no mention of any use cases in the research article	36
1 – When the article only discusses theoretical or use cases through references	21
2 – When the article directly implements or studies a use case	20

The “RQ1 Industry” score is only recorded when the “RQ1 Relevance” score is above 0. RPA is thought of as an industry-agnostic technology, and this was reflected in the list of industries captured in the SLR. A broad range of industries were represented, confirming the wide appeal that IA technology has today. Notably missing from the list of IA use case industries is

Government and Telecommunications, two industries that have been at the forefront of RPA adoption.

Table 8: RQ1 – Industries

RQ1 Industry	Count
Accounting/Auditing	4
Education	4
Finance	4
Healthcare	3
HR	3
Insurance	3
IT	2
Logistics	1
Utilities	2
Total	26

A total of 122 use cases were captured across the 77 papers. 42 (34%) of the described use cases were purely theoretical, meaning they were not directly studied, or referenced. Only 48 use cases (39%) were directly studied by the authors. Focusing on the real ML use cases used in IA, the majority (52%) involve handling written language and documents (Translation, NLP, NER, OCR, Document Classification). 15% of the use cases involve using Chatbots or Virtual Assistants as the input interface to trigger IA. These language, document and chatbot models are commonly available as generic, off-the-shelf machine learning services.

IA use cases that are less likely to be serviced by a generic pre-trained algorithm, such as custom developed models (Anomaly Detection, Forecasting, Risk Management, Binary Classification, Object Recognition, Multi-class Classification) represented only 25% of the use cases. This suggests current industry reliance on using pre-built models or using Machine Learning as a service.

Table 9: RQ1 – Use Case Scoring

Use Case Name	Theoretical	Referenced	Real	Total
Anomaly Detection	3	0	2	5
Binary Classification	1	0	6	7
Chatbot / Virtual Agents	6	5	7	18
Computer Vision	6	3	1	10
Document Classification	3	2	3	8
Email Classification	2	1	1	4
Facial Recognition	0	0	1	1
Forecasting	4	2	2	8
Multi-class Classification	3	0	2	5
NER	1	5	6	12

NLP	4	7	5	16
OCR	3	4	11	18
Object Recognition	1	2	0	3
Risk Management	3	0	0	3
Sentiment Analysis	1	1	1	3
Translation	1	0	0	1
Total	42	32	48	122

Under IA, a business process is codified for automatic processing, with a clear boundary of when automated work begins and ends. I tried to capture where in the automated process ML was used, either at the very start, someone in the middle or at the end as the very last step. The positioning of where ML is used, gives us an idea of “how automated a process is”. If ML is used somewhere in the middle of the business process without human validation of the ML prediction, the business process fits the mental image of being “truly automated”. However, if the IA process was designed to have ML as the very last step of automated processing, before a human picks up the work again, the use of ML might not be a true replacement of human knowledge or decision making.

Few papers described the use cases in enough depth to understand where ML was used, or positioned in the automated business process. Just one single use case described using ML in two different parts of the automated process – once at the beginning and again somewhere in the middle. Five cases described using ML at the very beginning of the process; four used document processing or chatbots as the ML entry point to the automation. Three use cases had ML being used somewhere between the start and end of the automated process. Anomaly detection and object detection was used in two of those cases, with the other case being NER. There were no instances of having ML at the very end of the process, which was a surprising finding as it would seem to be a natural point for human interaction to occur to verify a ML prediction.

Table 10: RQ1 – Positioning Scores

Positioning of ML in the Use Case	Count
Empty (no use case)	37
Unknown (use case present but unknown positioning)	32
Start	5
Middle	3
End	0
Multiple Entries	1 (Start and Middle)

The key points to retain from looking at the use case data are that 67% of the use cases involve converting text into structured data and that these use cases can readily be implemented as a solution. For the automated processes where the positioning was captured, most automation designs placed ML at the beginning of the process. This is intuitive given what I have

captured from the use cases – text digitization done at the beginning can enable the automation of many processes. This suggests that IA currently is mostly used to automate “simpler” intelligence tasks as opposed to replacing more complex human decision-making or knowledge tasks.

3.3 Research Question 2 (ML Algorithms and Technologies)

The Relevance Score, Industry, ML Algorithms and ML Technologies coded for RQ2 are presented below. Only 25 (32%) of the SLR research papers discussed the details of the ML technologies or algorithms deployed. Among the 25 papers, seven discussed only technologies without any algorithms, nine discussed only algorithms and nine discussed both.

Table 11: RQ2 – Relevance Scores

RQ2 Relevance	Count
0 – No relevance as no mention of any specific ML algorithms nor technologies are in the research article	52
1 – When the article only discusses ML algorithms or technologies theoretically or through references	17
2 – When the article directly implements an ML algorithm or uses an ML technology in an IA process	8

Insurance and healthcare industries lead the way in discussing the details of their IA implementations. The use of IA in healthcare was a major trend for research conducted in the past two years as automation was seen as a way to minimize the number of workers physically present in healthcare settings and help minimize Covid transmission.

Table 12: RQ2 – Industries

RQ2 Industry	Count
Accounting/Auditing	1
Education	2
Finance	2
Healthcare	3
HR	3
Insurance	4
IT	2
Utilities	2
Total	19

36 algorithms were counted among 18 publications which discussed specific ML algorithms being used in IA processes. Neural network-related algorithms (ANN, CNN, RNN,

LSTM, Deep Learning) represented 13 or 36% of all cases. Random Forests was the most common algorithm with seven papers referencing its usage, followed by Deep Learning networks.

In terms of the interpretability of the algorithms used [36], only Linear Regression, Logistic Regression, Decision Trees, K-Nearest Neighbours and Naïve Bayes are inherently explainable models. Together, interpretable algorithms account for only 22% of all algorithms found in during the SLR. The remaining 78% are black-box algorithms that cannot be readily understood.

Table 13: RQ2 – Machine Learning Algorithms

Machine Learning Algorithm	Count
Artificial Neural Networks (ANN)	3
Bayesian Networks	0
Convolutional Neural Networks (CNN)	2
Decision Trees (DT)	2
Deep Learning (DL)	5
Genetic Programming	1
Gradient Boosting	3
K-Means Clustering	0
K-Nearest Neighbours (KNN)	1
Linear Discriminant Analysis (LDA)	1
Linear Regression	2
Logistic Regression	2
Long Short Term Memory (LSTM)	2
Naïve Bayes (NB)	1
Perceptrons	0
Random Forests (RF)	7
Recurrent Neural Networks (RNN)	1
Support Vector Machines (SVM)	3
Total	36

A complete list of 28 captured technologies, of which 21 are unique is listed in the table below. Technologies mentioned more than once across different research papers are listed with parenthesis, and the number of papers it was mentioned in. The grouping of the technologies into categories, “Programming Languages”, “Commercial or Cloud Services” and “Machine Learning Libraries” was not captured during the coding phase, but added in the table to simplify presentation. The main distinctions between commercial vs. ML libraries (besides the cost) is the amount of control that firm has over the IA solution and the amount of expertise required to implement it. If a commercial solution is used, the amount of control over the ML algorithm is likely less, with fewer expertise needed to deploy the solution.

Almost all of the Machine Learning library technologies are Python based and Python is the only programming language mentioned to be in use. Eight of the technologies (29%) were cloud-hosted solutions provided by large Machine Learning vendors, whereas 13 (46%) were solutions developed using ML programmatic libraries. Out of the “Commercial or Cloud Services”

category, two of the technologies (“Azure Databricks Machine Learning” and “Azure Machine Learning”) are similar to “Machine Learning Libraries” category in that they are web-hosted platforms that allow for the development of custom ML models.

Table 14: RQ2 – Machine Learning Technologies

Machine Learning Technologies	Machine Learning Technologies
Commercial or Cloud Services	ABBY, Azure Face API, Azure Databricks Machine Learning, Azure Machine Learning, Google Dialogflow, Google Vision, Kommunicate, IBM Watson Assistant (2), IBM Watson, Orbograph OrboAnywhere
Machine Learning Libraries	Jieba, Keras, LibSVM, Microsoft Bot Framework, Pytorch, Scikit-learn (2), Rasa, Spacy, Tensorflow (2), XGBoost (2)
Programming Languages	Python (4)

The 28 technologies from the previous table have been reorganized into the Use Case categories from **Table 9: RQ1 – Use Case Scoring**. Many of the technologies span across multiple use cases. The most common use case addressed by the technologies are Chatbots and Virtual Agents. The use case named “Generic” refers to the technologies that can be broadly used across many use cases. “IBM Watson” was extracted from one paper, but as a technology, it can potentially refer to a whole suite of ML services. It was not specified which part of IBM Watson was in use, so it stayed categorized as a cloud service for **Table 14: RQ2 – Machine Learning Technologies**, and as a generic use case in **Table 15: RQ2 – Machine Learning Technologies by Use Case**.

Table 15: RQ2 – Machine Learning Technologies by Use Case

Use Cases	Technologies
Chatbots / Virtual Agents	Google Dialogflow, Kommunicate, IBM Watson Assistant (2), Microsoft Bot Framework, Rasa
Computer Vision Facial Recognition / Sentiment Analysis	Azure Face API, Google Vision
Generic	Azure Databricks Machine Learning, Azure Machine Learning, IBM Watson, Keras, LibSVM, Python (4), Pytorch, Scikit-learn (2), Tensorflow (2), XGBoost (2)
NLP	Spacy, Jieba
OCR / NER	ABBY, Orbograph OrboAnywhere

Data regarding the eight actual (as opposed to theoretical or referenced) algorithms and technologies in use are shown in the table below. There do not seem to be any patterns as there are a wide variety of industries, algorithms and technologies being used.

Table 16: RQ2 – Data from the Eight Papers with Relevance Score 2

Industry	Algorithms	Technologies
Education		Watson Assistant
Healthcare	ANN, Gradient Boosting	Google Vision, Spacy, Keras, Python, XGBoost
Healthcare		MS Bot Framework
HR	Random Forests	
HR	LSTM, Naïve Bayes	
Finance	Deep Learning, Random Forests, Linear Regression, RNN, SVM	LibSVM
Utilities	Logistic Regression, Decision Trees, Random Forests, Gradient Boosting	Azure Databricks Machine Learning
Utilities	Linear Regression, Random Forests	Python

Beyond the document processing and chatbot use cases identified in RQ1 which are readily deployed as pre-built commercial software, the remaining use cases are largely implemented using black box ML algorithms. Model predictions for black-box algorithms cannot readily be explained to management or to users affected by the prediction. Python is clearly the dominant language being used in IA, with almost all of the surrounding technology pieces supporting Python in some way.

3.4 Research Question 3 (IA Risks)

The Relevance Score, Industry and Risks for RQ3 are presented below. Only 26 (34%) of the 77 research papers mentioned any risk associated with IA. There were more theoretical risks, meaning that the risk was not directly elicited or observed by the researchers, rather than risks actually encountered and described by authors as part of a real implementation.

Table 17: RQ3 – Relevance Scores

RQ3 Relevance	Count
0 – No relevance as there is no mention risks in the research article	51
1 – When the article only discusses risks theoretically or through references	15
2 – When the article directly calls out risks due to implementation of IA	11

Finance was the most commonly cited industry for papers that discussed risks. Highly regulated industries (Finance, Healthcare, Government, Insurance, Accounting) accounted for 67% of the industries counted. This makes intuitive sense as they are more likely to care about risk management to meet regulatory demands.

Table 18: RQ3 – Industries

RQ3 Industry	Count
Accounting/Auditing	2
Education	2
Finance	4
Forestry	1
Government	1
Healthcare	1
Insurance	2
Logistics	1
Sports	1
Total	15

A total of 70 risks were captured across 26 research papers. These have been coded into 36 risks headings. The most commonly cited risks involve employee deskilling, loss of job security, data bias, data quality, data privacy, financial loss and reputation loss to the firm. It is important to note that this is not an exhaustive list of risks that IA poses to the firm, simply a list that has been extracted from the research papers during the SLR process.

The 36 risk headings and a brief description are provided in the table below. A more in-depth discussion on the 36 risks is provided in **Section 4.4 Risks**.

Table 19: RQ3 – Risks

Risks	Description	Count
Adversarial Attacks	ML algorithms used in IA are subject to adversarial attacks, which would propagate in unwanted automated work being processed	1
Assignment of Liability	When multiple companies are involved in the development and operations of ML predictions in IA, liability becomes unclear between the firms if something goes wrong	3
Attract Competitive Response	Publicly investing in IA can trigger responses from competition, either encouraging them to pursue IA themselves or to decry your use of IA	1
Cognitive Work Overload	Removing simple cognitive work may leave only difficult cognitive work, leading to cognitive overload and increased job stress	1
Compliance	IA may complicate the compliance terms of existing regulations. Third-party ML vendors must also be compliant with necessary regulations	2
Conflicts of Interest	AI Vendors may be incentivized towards rent seeking behaviour	1
Control Flow Drifts	Changing business process logic and pathing in the control flow may necessitate rebuilding ML models	1
Data Bias	The data may have biases and perform poorly on real life data	5
Data Drift	The underlying nature or distributions of the data may change over time	2
Data Privacy	Sending sensitive data to third parties for use or model development may lead to data leaks	4
Data Quality	The data quality may lead to poorly performing models	4
Departmental Resistance	Managers may worry that their headcounts or budgets will get frozen or reduced due to IA, leading to non-cooperation or sabotage	1

Difficult Error Detection	The use of automated ML decisions may make detection of errors in the business process more complicated	3
Ethics	Biases in the automated processing may lead to ethical concerns	2
Employee Turnover	Employees may quit the organization in large amounts because of IA	1
Financial Loss	Incorrect IA work may lead to financial loss to the firm	4
Information Asymmetry	Power balance between different units inside the business may shift, for example the data scientists who control valuable decision making may gain leverage over the business unit they are servicing	1
Loss of Control	If the firm relies on vendors to develop or host the ML predictions, there is a risk that their service will go offline or cease operating	1
Loss of Job Meaning	IA may be automating work that is meaningful to employees, leaving them less satisfied with their jobs	2
Loss of Job Security	Knowledge-based automation can lead to a much larger group of employees to worry about their jobs, increasing their stress levels and impairing their health	4
Low Predictive Performance	IA predictions may be less accurate or reliable than human predictions, resulting in worse outcomes after automation	1
Low ROI	The ROI of IA may be unattractive compared to RPA due to the additional needs of constant monitoring, retraining of models, and the costs of data scientists	1
Missed Servicing Opportunities	When narrow AI is used to interact with customers, there is no flexibility to discover additional ways in which the customer can be serviced	2
Mistrust in Management	A push towards IA may lead to mistrust being formed between workers and management	1
Mistrust in Model Predictions	A lack of trust in model predictions may lead employees to actively resist or sabotage IA efforts	2
Performance Agreement Breaches	Existing performance agreements or SLAs may need to be renegotiated after implementing IA	1
Performance Degradation	Model predictive performance are known to reduce over time unless actively managed	1
Prediction Accountability	Which employee(s) should be held responsible if a prediction or business outcome is incorrect?	2
Reduced Understanding of Business Logic	The overall knowledge about the business process may reduce over time if business decision making is automated	2
Reduced Work Preparedness	When IA is in place, employees spend less time looking at work cases meaning that fewer details are known about a case if it needs to be manually worked on	1
Regulatory	Government regulations may appear in the future, governing how and when algorithmic decisions can be used	1
Reputation Loss	Incorrect IA work may lead to reputational loss to the firm	4
Time Lag Effects	There is a time gap between when an error is detected and when the error actually occurred. Automatically processed work during this time gap needs to be reviewed to see if it needs correction or re-doing	1
Transfer Learning Bias	If transfer learning is used to develop the model, the base model may have hidden biases, with no way to fix it	1
Unmeasurable ROI	The ROI of IA may not be measurable due to being unable to quantify the value of knowledge or decision work	1
Worker Deskilling	Worker's skill in performing their knowledge tasks is reduced due to automation	4

3.5 Research Question 4 (IA Risk Mitigation Techniques)

The relevance scores for RQ4 were the lowest amongst all of the research questions, indicating that research on the topic of risk mitigation for IA still needed. Only 29% of the research papers discussed risk mitigation to any degree.

Table 20: RQ4 – Relevance Scores

RQ4 Relevance	Count
0 – No relevance as there is no mention risk mitigation for IA in the research article	55
1 – The article discusses risk management only theoretically, through references, or only in passing without details	14
2 – The article directly discusses at least one risk mitigation technique for IA in detail for at least two sentences	8

The discussion of risk mitigation was most common for papers discussing the Finance industry. There was surprisingly only one mention of risk mitigation methods for the healthcare industry.

Table 21: RQ4 – Industries

RQ4 Industry	Count
Accounting/Auditing	2
Education	1
Finance	3
Forestry	1
Healthcare	1
Insurance	2
IT	2
Logistics	2
Total	14

A total of 29 risk mitigation techniques were extracted, grouped under 15 headings. “Human in the Loop” (HITL) is by far the most popular method with nine mentions overall. HITL is when human input is deliberately introduced into the automated process to monitor, audit and change predictions that have been conducted by ML. HITL was also the most frequently cited method used in practice, with four mentions having a relevance score of two.

The next most frequently cited risk mitigation technique at three mentions is “Thresholding”, where the probabilistic confidence measure of each prediction is taken and compared to a chosen threshold level. If the confidence of the prediction is above the threshold, then automated processing continues and if not, the automated processing is paused until a

human can verify the prediction, or it is simply routed to a human for manual processing from that point forward. All three mentions were theoretical, with a relevance score of one.

Two of the risk mitigation techniques are contradictory: “Self-learning” and “Avoid Self-learning”. Both have only one mention. Note that “Self-learning” does not refer to unsupervised learning, which is when manually labelled data is not needed in order for the algorithm to learn. Here, “Self-Learning” refers to a configuration of the IA solution such that it will automatically take in new training data, re-train and deploy models into production without the need for human input. The idea is that prediction accuracy can automatically improve over time, reducing the risks of incorrect predictions, financial loss etc. The other opinion argues for the opposite, which is to maintain human control and oversight for any improvements to the ML model. The proposal to use “Self-learning” [29], to reduce risk was conceptual and not implemented in reality.

A summary of the 15 coded risk mitigation techniques can be found in the table below. An in-depth discussion of the risk mitigation methods is provided in **Section 4.5 Risk Mitigation Techniques**.

Table 22: RQ4 – Risk Mitigation Techniques

Risk Mitigation Technique	Description	Count
AI Liability Terms in Contracts	Liability in case of incorrect work or predictions in IA should be codified into formal contracts	1
Avoid Self-learning	Avoid using any techniques that involve self-learning, preferring to approve any changes or improvements to the underlying models before use in production	1
Contract Renegotiation	The use of IA may fundamentally change the premise(s) on which previous contracts with other firms were based on	1
Explainable AI	Choose algorithms that produce inherently interpretable models or use methods that can explain predictions after they have been made	3
Governance	Put governance and documentation into place to manage aspects of the machine learning lifecycle and to prevent the loss of process knowledge	2
HITL	Having a human monitor, review or audit the work performed by the ML algorithms. This may involve redesigning the process to have some cases routed to humans for manual processing	9
Minimize False Positives	Design or modify the ML algorithm to explicitly minimize false positives as opposed to another measure of accuracy	1
Monitor Data	Monitor and update the training data regularly and rebuild the related ML models	1
Monitor Models	Monitor and update the ML models on a regular basis	2
Process Runtime Controls	Provide controls that allow automated processes to change between human and ML prediction, validation or no validation during process execution	1
Random Sampling	Choose a fixed percentage of work cases that will always be sent to a human for processing instead of automatic processing	1
Self-learning	Improve existing ML predictions automatically through self-learning and automatically deploy them for use in production	1
Staged Deployments	Use deployment techniques such as canary testing, A/B testing that allow for fast deployment of model changes and rollbacks if a problem is encountered	1

Thresholding	Define thresholds for the ML algorithm such that any prediction that results in a confidence score above the threshold is processed automatically, but anything lower is sent to a human for processing	3
Understanding Employee Sentiment	Understand what employee sentiment is regarding IA and plan IA projects with these sentiment segmentations in mind	1

3.6 Chapter 3 Summary

Chapter 3 has presented a wide-angle view of the literature on uses of IA and their risks, distilled into metrics that codify the content of 77 papers selected for their relevance. The Relevance Scores reveal that the landscape is currently patchy and underscore that this thesis does indeed represent a novel research topic. Although across the study set many use cases were mentioned, two-thirds of them used “off-the-shelf” AI, focused on processing documents and performing human interactions with chatbots. Such a focus has implications for the specific firm-level risks and risk mitigation measures that are considered. As the extent of intelligent automation extends and additional parties are introduced into the automation service chain, such a limited consideration of use cases may in turn limit the scope of real-world risks that decision makers attend to. I also found that for the ML models that were developed in-house, the vast majority of the models were not interpretable. The most commonly used technologies used to develop the AI portion of IA is Python and its related libraries.

The risk coding in RQ3 resulted in 36 unique risks that potentially need to be addressed during the implementation of IA. The coding of RQ4 identified 15 risk mitigation techniques to some of those risks. In the next chapter, the important findings from this chapter are examined in greater detail.

Chapter 4: Data Discussion

Many of the use cases, technologies, risks and risk mitigations coded from the previous Chapter require further clarification as they were not discussed in enough depth to be used practically. This chapter will take a deeper dive into the findings from **Chapter 3: Data Analysis** to come up with useable insights.

4.1 IA Use Case Riskiness

In order to develop a more nuanced sense of the risk levels present in the IA use cases, a second-level coding was performed on the captured use case data. My goal was to separate the IA use case risk levels into two categories, low (L) and high (H). A low-risk level means that there is no, or very little chance of financial or reputational loss due to an incorrect ML prediction in the IA process. If human validation of **every** ML prediction is maintained in the process, the risk is rated as low. Any other situation, for instance, where the use case is not clear, where only some of the predictions are human-validated, or otherwise, is categorized as high-risk. This deliberately

skews the coding to a high-risk level, to present an optimistic (but probably unrealistic) rating of how many firms are doing high-stakes replacement of knowledge work through IA.

Of the 17 use cases that were actually implemented (had a Relevance Score of 2), five of them (29%) were rated high-risk, and 12 (71%) were rated low-risk. It was not clear in many of the research papers whether humans were kept in the process to validate predictions and if so, whether all predictions were validated. Of the 12 low-risk use cases, seven are business processes that are inherently low-risk to begin with, unlikely to cause any financial loss or harm to the company regardless of whether there is HITL. This includes cases where chatbots are used to retrieve and consolidate informational data and triaging work for internal employees.

The remaining five low-risk use cases would have been categorized as high-risk, if it were not for the risk controls put into place. Optimistically, only 29% to 59% of IA use cases are high-risk, despite all the media chatter about robots replacing the high-level cognitive skills of people. Although low-risk does not imply low-impact, firms are in some way intuitively assessing the risks of IA and are settling at implementing low-risk use cases. The coding of the risk level, justification and SLR article reference can be found in **Appendix: I. Risk Levels of Real IA Use Cases (Relevance Score of 2)**.

4.2 Machine Learning Algorithms

Coding of the ML algorithms used in IA was performed to understand if there is any link between the algorithm and risks and to understand whether algorithm choice influences the risk mitigation measures that are available.

ML problems can be broadly categorized by whether they deal with classification or regression [37]. Regression problems involve predicting numerical values, such as forecasting sales figures or predicting a credit score. Classification involves the prediction of labels, such as “spam email” vs. “non spam email”. Three commonly captured user cases, OCR, NER and Chatbots that predict user intent all deal with classification problems. In fact, the vast majority of use cases that have been coded are classification problems; only forecasting and risk management, or 11/122 (9%) of the use cases dealt with regression.

Many of the risks and risk mitigation techniques, for example, “Mistrust in Model Predictions” and “Performance Degradation” require an understanding of the underlying performance measure of the algorithms. For regression algorithms, measuring the performance or accuracy is straightforward, since we can compare the predicted numerical values to the actual values during the model training period. However, measuring the performance of a classifier, representing the majority of IA use cases, is a more complicated topic which requires understanding the concept of “confusion matrices”.

4.2.1 Measuring Performance of Classifiers – Confusion Matrices

The “performance” of a classifier can be measured in many ways. One popular way is to look at the accuracy, which is the ratio of correct predictions to the total number of predictions made. Most performance measures are based on a “confusion matrix”. The structure of a 2x2 confusion matrix is shown in the table below. In a 2x2 matrix, the classification problem predicts between one of two labels, for instance whether someone “will purchase a product” vs. “will not purchase a product”. “Actual Positive” are the people who have actually purchased the product and “Actual Negative” are those who have not. “Predicted Positive” are the people that the model believes will purchase the product, and “Predicted Negative” are the people that the model predicts will not. Note that the size of the matrix is dependent on the number of labels that we are trying to predict. A problem where we want to predict between three labels will result in a 3x3 confusion matrix.

Table 23: Confusion Matrix Structure

	Predicted Positive	Predicted Negative
Actual Positive	True Positive Count	False Negative Count
Actual Negative	False Positive Count	True Negative Count

Continuing the 2x2 example in the table below, assume that we have data on 100 customers, which will be used to train a ML model. In the dataset, 30 people actually purchased the product and 70 did not. A model is built from the data and the following confusion matrix is obtained. 26 people actually purchased the product, and were predicted to purchase the product, which correct. 62 people did not purchase and the model predicted that they would not, which is also correct. Eight people did not purchase, but the model predicted that they would, which is incorrect. And finally, four people actually purchased the product but the model predicted that they would not, which is also incorrect.

Table 24: Confusion Matrix Example – Will Someone Purchase the Product?

	Predicted Positive	Predicted Negative
Actual Positive	26	4
Actual Negative	8	62

There are many ways to measure the performance of a classifier, with the measure chosen being highly use case dependent. Suppose that we want to build an algorithm that classifies videos as “safe for kids” vs. “not” for parental controls. “False Positives”, when videos are predicted as being child-friendly when they are not are unacceptable. In this case, a performance measure that takes into account “False Positives” is desired. An example of this would be to use $\text{True Negatives} / (\text{True Negatives} + \text{False Positives})$.

Another example is predicting whether a bank transaction is fraudulent or not. Unlike the previous example, “False Positives”, or predicting that a transaction is fraudulent when it really is not is acceptable to people, as long as every actual fraudulent activity is correctly predicted as such. What is undesirable is if an actual fraudulent activity is predicted as non-fraudulent, or “False Negatives”. An error measure that could be used here is True Positives/(True Positives + False Negatives).

Other performance measures exist that only care about the pure accuracy of correct predictions, while not caring about incorrect predictions, or aim for some balance between the two. Regardless of which performance measure is chosen, the measure is typically based on some combination of the number of True Positives, True Negatives, False Positives and False Negatives. The Confusion Matrix is readily available when an algorithm is developed in house or custom developed solely for your own use by third parties. However, if the classification algorithm is a consumed as a service across many customers, it is likely a black box. Performance measures based on the confusion matrix will likely not be provided. Instead, most AI vendors will only provide a “Confidence Interval” attached to each prediction, to indicate how accurate that prediction is [38], [39], [40].

4.2.2 Confidence Intervals

The data coding of RQ2 revealed that 29% of the technologies used fell under the category of black-box ML services. These services do not provide visibility into the confusion matrices. Instead, they provide a number alongside each prediction, typically between 0 and ,1 (or 0 and 100) representing how confident the service is of the prediction. The purpose of this confidence value is to explicitly let users of the ML algorithm “calibrate custom thresholds for their content and scenarios to route the content for straight-through processing or forwarding to the human-in-the-loop process” [41]. An example of this would be submitting a photo of an animal into Google Vision. Google Vision may reply that it predicts that the image contains a “dog” with “0.97” confidence. Whether 0.97 confidence is “good enough” is completely dependent on your application, and the correct value for the score must be figured out through experimentation using your own datasets.

Third-party services will only provide you with confidence scores for every prediction. Custom-built models can provide confidence scores for every prediction in addition to the confusion matrix. The three most widely used technologies identified, Scikit-learn, Tensorflow and XGBoost all have ways to provide confidence values [42], [43], [44] to accompany individual predictions. It is reasonable to assume that almost every ML model, whether it is custom developed or provided as a black-box service, will provide confidence scores along with the actual predicted labels for classification problems, which represent the majority of IA use cases.

The presence of confidence scores with every prediction enables the use of “Thresholding”, which has been identified as a risk mitigation technique. The way that thresholding is used in IA decision making is as follows. First, the user defines one or more threshold values. Different actions for the different ranges of confidence intervals created can

then be prescribed. For example, if a single threshold is set to 0.95, we can decide to allow for fully automated processing if the prediction confidence is above 0.95, or manual human processing if below. Thresholding is further discussed in **Section 4.5.3 Human Interaction Design**.

4.2.3 Interpretability

Having an interpretable ML model helps address many risks including, “Mistrust in Model Predictions”, “Difficult Error Detection” and “Reduced Understanding of Business Logic”. There is currently no mathematical definition, or even consensus among academics on the definition of what “interpretable” means in ML. The differences between a model being interpretable and being explainable has also been discussed at length in literature [45]. For this discussion, interpretability and explainability are used interchangeably under the idea of being able to present a model in understandable terms to a human [46]. In terms of risk, the main groups that we would need to explain a model to are management, end users affected by the automated decision and courts of law in case liability arises.

There are two broad paths towards ML interpretability [45]. The first is to choose an inherently interpretable algorithm. The downsides to this are that there are very few inherently interpretable models and there is a misconception that these types of algorithms are simpler, thus lacking in predictive power compared to black-box models [47]. Among the limited list of intrinsically interpretable models are: Linear Regression, Logistic Regression, Decision Trees and Naïve Bayes [45]. Despite their inherent interpretability, when the number of features is high, even these models become difficult to understand. Limiting the choice of ML algorithms to purely interpretable ones is impractical. The five most common algorithms used in IA were Random Forests, Deep Learning, Gradient Boosting, Support Vector Machines and Artificial Neural Networks. ANNs is a generic catch-all term for any Neural Network. None of these algorithms are inherently interpretable, which is surprising given multiple citations of using “Explainable AI” as a risk mitigation technique. Only 22% of the coded algorithms are inherently interpretable.

The second path towards interpretability lies in creating an explanation method to explain the model or the predictions, after the model has been trained. There are currently many more black-box models in use for IA use cases, so this is a more realistic path to pursue. Explainable methods are used after the model has been built, and can also be used to derive understanding from inherently interpretable models as well, as long as the explanation method is model-agnostic [45]. Still, there are many different ways in which a model can be understood. We can try to understand a model in aggregate, without looking at any specific prediction. For instance, we can try to understand the average behaviour of the model, or how slight differences in feature values will affect a theoretical prediction. Algorithms that provide for this type of overall model understanding are termed “Global Model-Agnostic Methods” [48]. Global Methods typically look at the interactions between the features presented in a graphical plot and often require access to the underlying data, making it impractical to use with third-party developed algorithms.

While the overall behaviour of the model is important, for the purposes of IA, understanding why a model made a certain prediction for a **specific case** is extremely important

for audit purposes. This is referred to as the “Local Interpretability” of a single prediction [48]. The two most mature and widely researched local interpretability methods are called LIME and SHAP [46]. The technique underpinning LIME is to provide slight variations of the data that you want to predict against a black box model, get the predictions and use that to train a new interpretable model. LIME works on any type of data, whether it is tabular, image or text. LIME can be readily implemented in Python, and is commonly used together with Lasso or Decision Trees with limited depth as the interpretable model. This creates human-friendly explanations that can easily be understood by a layperson. However, the method is still under active research and great care needs to be taken to safely apply it [48]. SHAP is another popular method used to explain individual predictions based on the concept of Shapley values from game theory [48]. A major disadvantage of SHAP is that it requires knowledge of the average prediction which means that the data must be known, ruling out its use for non-custom-built models.

There has been notable critique for using explainable methods instead of inherently interpretable model for high stakes decision making [47]. Any explanation will lack fidelity compared to the original model and can very well be an inaccurate representation of it. An explanation can be right or wrong, partially right or partially wrong and it is not possible for us to know when. [47] argues that with careful design of the features, an inherently interpretable model can likely reach comparable accuracies to black-box models. However, there are still valid reasons for wanting to create a black-box model, for instance, to protect intellectual property and trade secrets.

ML model interpretability is likely to gain importance over time, as there is strong belief that the demand for explainable AI will be high from judges, as more and more court cases demand explanations from algorithmic decisions [49]. It is further suggested that local explanations about a specific decision instead of a global explanation of the model is expected by the US legal system when an explanation is needed [50]. For firms looking to seriously reduce their liability footprint from the use of IA, a modelling approach that tries inherently interpretable methods first would be wise, even if it comes at the cost of predictive power.

4.3 Machine Learning Technologies

Python was the only programming language discovered in the data coding, which is intuitive as Python is the most popular programming language as of February 2022 according to TIOBE, an index which measures the popularity of programming languages [51]. Other popular languages used for ML are R, which is ranked 13th, and Julia which is ranked in 30th place. Every one of the 10 ML libraries that were extracted from RQ2 support Python, with many of them only supporting Python. Only XGBoost was reported to support R and Julia in addition to Python. Even though Python was only mentioned four times, it is probably safe to infer that Python is overwhelmingly being used in IA compared to other languages.

Table 25: Supported Programming Languages for Machine Learning Libraries

ML Library	Supported Language(s)
Jieba	Python [52]
Keras	Python [53]
LibSVM	C, C++, Java, Matlab/Octave, Python [54]
Microsoft Bot Framework	C#, Java, Javascript, Python [55]
Pytorch	C, C++, Python [56]
Scikit-learn	Python [57]
Rasa	Python [58]
Spacy	Python [59]
Tensorflow	C++, Java, Javascript, Python [60]
XGBoost	C, C++, Java, Julia, Python, R, Ruby, Scala, Swift [61]

The most commonly used ML algorithms uncovered by the SLR were Random Forests, Deep Learning (which could be in the form of any neural network), Gradient Boosting and Support Vector Machines. None of these models are inherently interpretable. It is therefore important to understand whether the popular explainable methods for local models, such as LIME and SHAP can be readily applied to those algorithms in practice instead of just in theory. The table below shows a mapping between the common IA algorithms and the technology libraries that can be used to implement them. We previously established that the technology libraries can all be used with Python.

Table 26: Algorithm to Technology Library Mapping

Algorithm	Technology Library
Deep Learning	Keras, Pytorch, Tensorflow
Gradient Boosting	Scikit-learn, XGBoost
Random Forests	Scikit-learn
Support Vector Machines	LibSVM, Scikit-learn

Both LIME and SHAP have readily usable implementations in Python [62], [63]. The SHAP library has specific support for tree-based models in Scikit-learn and XGBoost. It also has modules to work with deep learning models in Tensorflow, Keras and Pytorch. Finally, SHAP can work with Support Vector Machines produced in Scikit-learn, although it is unclear whether it works with LibSVM. SHAP is therefore technically compatible with the vast majority of IA use cases that have been examined, except potentially Support Vector Machine models created by LibSVM.

LIME's documentation says that it can work with any classifier that implements the "predict_proba()" method which is found in Scikit-learn [64]. From examining Scikit-learn's API [65], this would include SVMs, Random Forests and Gradient Boosting. The XGBoost library also contains the predict_proba() function, making it compatible with LIME [44]. There are also examples of using LIME with Keras and Pytorch [66], [67]. Given that Keras is actually a high-level API that runs on top of Tensorflow itself [53], this implies that LIME is compatible with

Tensorflow. Once again, all of the common ML algorithms are supported by LIME, except for LibSVM.

From this, we start to see a consistent story. Most of the models developed are not interpretable, are all implemented using Python-related libraries. This means that they have access to confidence thresholds, and the most common explainable AI libraries, SHAP and LIME. The practical conclusion is that both the LIME and SHAP explainable methods can work with almost all popularly used algorithms and technologies encountered in IA, except for LibSVM. If a SVM model needs to be developed, it is recommended to develop it in Scikit-learn instead, for compatibility with the LIME and SHAP libraries. However, a large gap between the theory of explainable methods and implementation arises here. It is already known that SHAP must have access to the underlying data in order to be used, but LIME can theoretically be used against any black-box model including third-party APIs, as we only need access to the predictions. However, in practice, the implementation of LIME also requires access to the underlying model. Neither LIME nor SHAP can currently be used to explain pure black-box models that are not controlled by the firm.

4.4 Risks

36 risks were uncovered from RQ3 of the SLR. Since most of the research papers did not focus on explaining risks, they were often only discussed in passing, without much depth. This section aims to add more description and bring clarity to each of the 36 risks. To facilitate discussion, the risks are organized into two categories: 22 Socio-Organizational risks and 14 Operational risks. Risks belonging to the socio-organizational category arise from the relationships between different social groups, such as the employee and the firm, or between firms. Operational risks are those that come from the day-to-day IA operations, and are often project-based or technical in nature.

4.4.1 Socio-Organizational Risks

The socio-organizational risks have been further sub-divided into four sections: Environmental, Enterprise, Employee and Third-Party risks. Environmental risks come from the interaction between the firm and the wider industry, society and legal structure. Enterprise risks sits between all other sections, involving risks that are posed to the firm. Employee risks are those experienced by individual employees affected by IA. Third-Party risks are the risks that arise between the firm, their customers, partners and vendors.

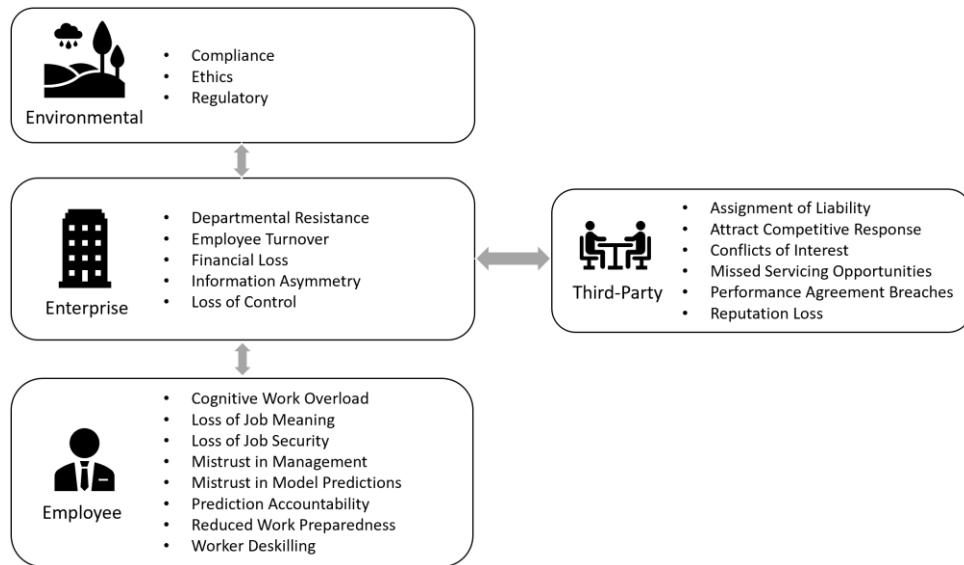


Figure 2: Four Categories of Socio-Organizational Risks (Source: Author)

4.4.1.1 Environmental Risks

“**Compliance**” was only cited once in a pharmaceutical setting [68], where the business process for drug prescriptions is automated, but human verification of the drugs prescribed is still maintained. Compliance may be industry specific or a part of national or international law. For example, articles 13-15 of GDPR, which came into enforcement on May 2018 provides Europeans with the right to “meaningful information about the logic involved in automated decisions” [69], which implies some sort of explainable AI or interpretable AI method being used. Article 22 of GDPR states that individuals “have the right not to be subject to a decision based solely on automated processing” [70]. It also states that people have “the right to obtain human intervention”, which implies that there should be a mechanism within IA to completely disable automated decision making for specific cases, and a mechanism to allow people to inform the firm and invoke their rights. The current penalties for GDPR violations are fines of up to 20 million euros, or 4% of global revenue, whichever is greater [71].

“**Regulatory**” risks [72] refer to the threat of being non-compliant with future enacted laws. We might imagine an adapted version of GDPR’s right to explanation being adopted outside of the EU. In the extreme case, new regulations may require completely disabling algorithmic decision making. To reduce the impact of regulatory risks, IA teams should plan for explainable AI and having ways to disable algorithmic decision making before new regulation is passed.

Two different interpretations of “**Ethics**” were surfaced in the SLR. One ethical concern is that the automated processing may lead to unfair outcomes due to bias in the algorithms or data [73]. The concept of bias is discussed later on in **Section 4.4.2.4 Data Risks**. The right to non-discrimination is fundamental to many societies, but the use of big-data and ML are easily able to produce biased predictions that harm vulnerable groups, leading to biased decisions [71]. Firms must take care to not build models that use certain sensitive data, for example, ethnic

origin, political affiliation, religious beliefs, sexual orientation etc. and to take care in using features that are correlated to sensitive data, such as geographic region and ethnic group, unless explicit consent is given. Another interpretation of ethics comes from the responsibility of the firm with respect to the employees, to properly educate and prepare employees in case of job nature changes or displacement by IA [74]. Strong collaboration with HR should be maintained during IA implementation to measure the impact of IA on employee well-being.

4.4.1.2 Enterprise Risks

“Departmental Resistance” is expressed by managers affected by IA in [75]. IA is often pitched as way to save costs through headcount reduction, leading managers to worry that their own headcounts and thus budgets will be reduced. This in turn will reduce their sphere of influence within the firm and their ability to meet KPIs. This can lead to significant resistance to IA projects, result in IA non-adoption or even active sabotage of efforts. [76] has extensively covered the topic of whether IA results in job loss on a macro-level. On a macro-level, the most likely scenario is not permanent job loss, but the replacement of certain tasks out of an individual’s job. However, on a micro-level, there is evidence of headcount reduction. One study at a bank saw a reduction in demand for low skilled jobs, although demand for higher skilled jobs increased as a result [77]. This likelihood of this risk is therefore dependent on whether the proportion of tasks being automated away warrant a reduction of headcount. Managers can counteract this risk by proposing new tasks or training to maintain their headcount.

“Employee Turnover” refers to large-scale voluntary turnover due to IA adoption. Existing research on the sources of employee turnover show that the reasons for quitting are wide and varied. Different studies emphasize different factors, such as stress levels, a sense of power loss, organizational instability and economic reasons [78]. However, in a study related to AI implementation [75], employees who held less positive emotional attitudes towards AI had a higher intention to leave the company due to its adoption.

Li et al. conducted a study in the hospitality field on the impact of AI and robotics on employee turnover intention, revealing three key points [79]. First, there is a positive relationship between AI awareness and employee turnover intention. Next, turnover intention is weakened when employee support from the organization is perceived to be high. Support can include many things, including employee development, team building exercises etc. This is a key lever that firms can use to reduce the risk of employee turnover. Finally, the positive relationship between turnover intention and AI awareness is stronger when the work environment is highly competitive. This suggests that lowering workplace competitiveness can weaken the relationship between AI awareness and turnover intention, however, this is not something that can be easily changed in practice. The study also adds evidence to the “Information Asymmetry” risk, as the authors suggest that the introduction of AI and automation can divide employees and lead them to hoard and not share information, leading to a more toxic workplace.

The risk of **“Financial Loss”** was cited four times with two distinct definitions. Two papers [25], [74] describe financial loss due to potential biases in the model, leading to financial loss

through litigation. Two other papers [80], [17] describe the losses arising from triggering erroneous actions due to incorrect predictions, which are losses specific to the business process being automated. Broadly speaking, most of the risks described in this discussion will eventually lead to some sort of financial loss.

IA can introduce a new type of “**Information Asymmetry**” into firms [81], which allows for far more processing power and information consumption by those who are in control of IA versus those who are not. Another form of asymmetry arises between those who understand the models and those who do not. [81] also argues that AI can lead to a “**Loss of Control**”, if an outsourcing model or ML as a service is used. An example of this is if the prediction service goes offline for planned or unplanned reasons, IA processing will stop unless there is an alternate ML system available to complete the predictions. This may seem farfetched, but in December 2021 alone, Amazon AWS had three large-scale unplanned outages, halting the services of many companies for multiple hours during the workday [82].

4.4.1.3 Third-Party Risks

Third-party risks exist due to the interactions between the firm and other firms or customers. The topic of “**Assignment of Liability**” [83], [17], [84] questions whom should be liable when automated processing causes losses of some sort and liability must be assigned. This is discussed in more detail in **Section 4.5 Risk Mitigation Techniques** of this chapter, but the short answer is that liability will almost certainly fall on the firm using the prediction for automated processing. It is therefore a significant risk for firms who plan on using IA for high-stakes use cases.

IA is being rapidly adopted, especially among highly regulated industries. For those industries where IA adoption or AI use is uncommon, publicly adopting IA can “**Attract a Competitive Response**” [85]. This can manifest through competitive advertising. For example, in industries that are very traditional, or where human-based service is the norm (hospitality and healthcare), competitive advertising campaigns to condemn the use of IA can be done to try to damage the firm’s reputation or steal market share. Adopting IA publicly can also spur competitors to adopt IA as well, narrowing the gap between advantages gained from the IA program.

“**Conflicts of Interest**” can appear between firms [81], when IA or AI are provided as a service. In a worst-case scenario, the service provider of the prediction may try to hold the algorithm hostage, in an attempt to extract more value from the firm through rent-seeking behaviour. They may also shop around the IA service or developed ML model to competing firms.

When IA replaces human communication with end users through chatbots or virtual assistants there is a risk of “**Missed Servicing Opportunities**” [86], [87]. Customer-facing employees that adopted IA indicated that the use of automation rigidly defined the boundaries of a service interaction. This can leave customers not fully satisfied if they have other issues that need addressing. It also prevents human customer service agents from surfacing additional ways

in which the customer can be serviced to boost satisfaction. Chatbots was one of the most popular use cases of IA, so this risk is especially important for industries where customer service quality is paramount.

For companies that have existing, long lived agreements with service level requirements, there are “**Performance Agreement Breaches**” that can arise due to automation [83]. Most service level agreements have measurable numerical targets to achieve, for instance minimum response time, minimum completion time, maximum downtime etc. and can incur financial penalties when targets are missed. IA will likely increase the throughput of work cases completed, lowering the risk of missing some types of service level agreements on average. However, it also introduces additional points of failure in the process (the infrastructure to manage and deploy ML) and sources of potential downtime. For example, if the ML prediction is cloud hosted, and the cloud platform goes offline, it could take some time to discover and recover from that situation, leading to performance breaches. A review of such contractual obligations is recommended to assess the impact of IA on these types of clauses.

“**Reputation Loss**” to the firm [25], [17], [88] can happen for numerous reasons. [25] describes reputation loss from automated processing caused by biases in the data, for instance race, gender or ideology. Despite not being an IA specific example, the ML system “COMPAS” that was developed to predict criminal recidivism led to a ProPublica criticism of the system, alleging that the algorithm was racially biased [89]. This gained enormous attention from the press and academic study at the time, and illustrates how reputation loss due to IA can be a very real concern. Reputational damage can also simply be from the loss of customer confidence caused by incorrect automated decisions and processing [17]. The use of IA can also cause reputation loss negative perception of the technology from end-users, or mismatch between brand image and the use of new technologies [88].

4.4.1.4 Employee Risks

“In the long term, artificial intelligence and automation are going to be taking over so much of what gives humans a feeling of purpose.” – Matt Bellamy, musician, on the impact of IA on human satisfaction in the workplace [90].

We have established that there are far more instances of low-risk cognitive tasks being automated compared to high-stakes ones. Imagine a scenario where someone’s manual tasks and low-risk cognitive tasks have been automated away, and that their time is filled with more high-risk, high-stress ones. This may lead to “**Cognitive Work Overload**” [91]. The authors of [91] hypothesized that this could be an outcome of automation and AI on employee well-being. However, their results lacked evidence to directly support this claim, so further study is needed.

As cognitive or customer facing work is replaced with IA, some employees report decreased job satisfaction due to a “**Loss of Job Meaning**” [86], [91]. Three broad themes describe the components that make up the concept of “meaningful work” [92]. The first is a “sense of self”, that is recognizing and developing one’s potential through work and being able

to bring one's whole self, including our minds, body and emotions to work safely. This "sense of self" concept is similar to Maslow's concept of self-actualization. The second component is "the work itself", or doing something worthwhile. The third component is a "sense of balance", or the alignment between the work life and personal life, so that no one side dominates the other. "Meaningful work" interplays between each of the three concepts and is highly individual.

The study in [86] examined the replacement of social workers who help the elderly apply for government benefits with RPA and a web interface, leading to a loss of professional discretion. This corresponds to taking away from the concept of "the work itself". Lowered personal agency (external locus of control) is negatively related to job satisfaction, and lowered job satisfaction reduces commitment to the firm and increases the intention to quit [93]. While [91] did not have evidence to support the "Cognitive Work Overload" hypothesis, it did have evidence to support the claim that IA can replace work considered core to an employee's sense of identity, reducing job satisfaction. Careful evaluation should be done to avoid automating the parts of someone's work that are core to their job satisfaction.

Job security refers to employee expectations about the stability and longevity of their job [94]. Numerous research papers state that employees risk feeling a "**Loss of Job Security**" [14], [95], [85] due to IA. Contrary to intuition, the research examined in the SLR [91] disagrees with this notion, as their study does not find a significant link between increased feelings of job insecurity specifically due to automation and AI. Other studies however do find this link, especially in industries such as retail and insurance [96]. One possible reason for this discrepancy is because workers lack awareness and therefore do not feel job insecurity due to automation. Another possibility is that people do not attribute the risk of job loss due to automation to themselves, but to others instead. It is however known that a perceived increase in job insecurity is positively associated with an employee's intention to quit [97] and resistance to change [96].

The three previous risks deal with employee's feeling of well-being. While there was only evidence to directly support the "Loss of Job Meaning" risk, further research is needed to determine to what degree "Cognitive Work Overload" and "Loss of Job Security" are risks.

"Mistrust in Management" can occur when a top-down approach is taken to implement IA, instead of implementing based on demand generated from the departments themselves [14]. While having an executive sponsor and treating automation as a strategic initiative is IA best practice [24], IA implementation should be carefully considered, rather than forced on everyone. There is no universally accepted definition of "trust" in academia [98]. The closest concept to a universal definition of trust is the "willingness to assume risk", meaning that if we trust someone, we are willing to become vulnerable and accept the negative consequences if something bad occurs. Viewed from this lens, a lack of trust in management can manifest into resistance to organizational change and unwillingness to accept blame for problems that arise from automated processing. "Trust" represents the employee's psychological state towards their employer. "Trustworthiness" on the other hand is what leads to trust.

More specific to IA, the concept of "organizational trustworthiness" [99] is used to evaluate management behaviour during workplace transformations. The idea of trustworthiness

in this context is defined as the expectation that the organization and managers will act in the interests of its employees. The behavioural norms that define organizational trustworthiness in the workplace include two parts. The first norm is management competence, or that the firm must maintain an effective management system that allows employees to meet the demands of their jobs. Second, is that employees should be treated with dignity and respect through supportive employment practices. A top-down approach to IA touches both norms; employees may feel that IA will hinder their ability to complete work, and the thought that IA can replace their tasks can be taken as a sign of disrespect. When organizational trust is low, there is increased conflict between management and employees.

The unwillingness to shoulder blame is closely related to the next two risks, which are **“Mistrust in Model Predictions”** [25], [95] and **“Prediction Accountability”** [84], [74]. Under IA, there is a need to assign accountability of incorrect predictions and incorrect processing to individuals in the organization. For incorrect predictions, the natural assignment of accountability would go to whomever maintains the ML model, but this may be an external firm to the company. If the model is not developed in-house, the next natural assignment of accountability is to either the business users who make use of the prediction to complete their work, or members of the IA team. A mistrust in model predictions could cause employees to reject accountability and finger point between the business users and IA team.

One study examined a financial-services use case where RPA and ML were used to handle customer investment tasks. While IA resulted in considerable time savings for each affected employee, removing their manual data entry tasks meant that they spent less time understanding their customer’s data, resulting in **“Reduced Work Preparedness”** [95] when they actually needed to interact with the customer. Reduced work preparedness is a less severe form of the next risk, which is **“Worker Deskilling”** [17], [72], [87], [100]. This was the most common socio-organizational risk extracted in this SLR. Decision making is a cognitive skill, which is known to degrade when not used. If IA is increasingly used in organizations, there is a risk that human knowledge is shifted away from people into black-box ML models, leading to a permanent loss of organizational decision-making skill, reduced process knowledge and lowered ability to perform manual work in case it is actually required.

4.4.2 Operational Risks

14 Operational risks can appear during the planning, implementation or operation of an IA project. These risks are divided into four categories: Project, Process, ML Model and Data.

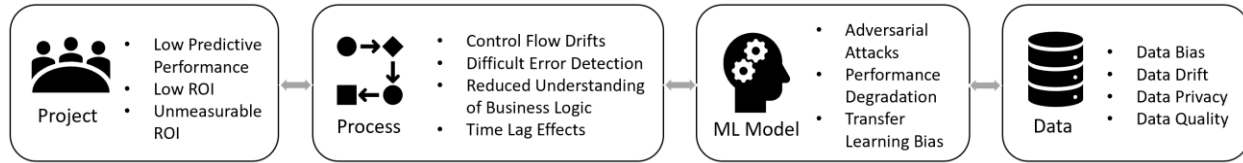


Figure 3: Four Categories of Operational Risk (Source: Author)

4.4.2.1 Project Risks

Project-based risks can prevent an IA project from being green-lit or lead to the project being shut down prematurely. “**Low Predictive Performance**” [101] refers to how well the model performs relative to existing alternatives, usually human prediction. Although ML can already outperform humans at Chess and Go, AI experts believe that the time when ML can outperform humans at every task is at the earliest 25 years away [102]. When moving toward IA today, this would lead to a trade-off between faster process throughput and the need to fix incorrect work caused by poor predictions. Poor prediction performance is also frequently linked to Data Bias and Data Quality, which are risks discussed further below.

Two closely linked project risks were coded: “**Low ROI**” and “**Unmeasurable ROI**”. Low ROI [25] refers to the higher cost to build and maintain an IA solution compared to a traditional RPA solution. There are presently three distinct ways that RPA vendors enable the use of ML with their products. All major RPA vendors companies provide add-on document processing solutions [103], [104], [105] that can perform OCR, entity extraction, document classification etc., enabling document-based ML. Next, they can provide a full ML development environment to build and deploy ML models using tools like Python, Keras, Scikit-learn etc. into an automated process [106]. Finally, vendors make it simple to connect an RPA process to an online ML model for web-based prediction. Regardless of which method is chosen to integrate ML and RPA together, the ML portions of IA can typically be priced independently from the RPA costs of IA.

Cost-benefit analysis or total cost of ownership are key components of the ROI calculation and are the major criteria for getting an IA project approved [107]. While [107] provides a list of direct and indirect costs to consider for a traditional RPA project, there has not been any mention of cost drivers for IA projects. Estimates for an initial deployment of a ML solution are around \$100,000 [108], [109] on top of existing RPA costs. Like any long-lived technology product, maintenance is required for ML. The costs of technology maintenance typically accounts for 40% of the amount invested on IT projects [107]. High start-up costs are a legitimate concern for first time IA projects, but the marginal costs go down as more AI-enabled processes are automated, making investment more attractive as long as there is continued demand. The costs of a traditional RPA project can be well estimated, and the incremental costs of adding ML can be realistically estimated as well.

The second half of the ROI equation are the benefits of the IA solution. The benefits of RPA are process specific; the most straightforward measurement are the estimated hours saved multiplied by the salaries of the people who do the work manually. However, if ML is used to

replace someone's decision making which only takes a few seconds, there are no hours saved to measure and add to the benefits side, leading to "Unmeasurable ROI" [85], or not knowing how to quantify the benefits IA bring over RPA. In order to capture the benefits of replacing decision making, a shift from a time or dollar savings mindset to a business value mindset should be made [107]. A list of potential criteria for a value-based assessment of IA can be found in [107].

4.4.2.2 Process Risks

Process Risks occur during the day-to-day operation of the automated work platform. "**Control Flow Drifts**" [14], are described as the need to rebuild ML models as a result of changes in business requirements. For example, imagine that an IA process uses ML to classify between sending a customer service request to the fraud department or generic customer service. If the process is redesigned to include redirecting to the credit card department, a new model needs to be designed, trained and deployed. Changes to the business execution flow may be a high-effort, high-cost exercise if new data needs to be collected and the ML model rebuilt.

One of the job roles that has emerged to monitor the progress of RPA work is called the "Process Controller" [110]. Part of the Process Controller's job is to ensure that automated work is completed according to service level agreements and to investigate any errors that occur during robot processing. Errors during processing can be caused by many factors, for example, a business application can hang, or a website may have changed its layout, making the robot unable to find the right field to enter data into. These sorts of errors will halt automated processing and are readily reported to Process Controllers as the robot will inform the RPA control dashboard that it was unable to execute the next processing step. However, errors encountered during the automation of knowledge work will still allow the robot to continue processing and will remain undetected by Process Controllers. This leads to a risk of "**Difficult Error Detection**" [95], [74], [75]. A new type of processing error is created by mispredictions, which is not detectable unless someone actively dives into the processing details for an audit, or if a business user further downstream catches the error when they use the result of IA processing as an input to their own work.

Closely linked to the employee risk "Worker Deskilling" is the idea of an overall "**Reduced Understanding of Business Logic**" for a business process [95], [111]. Worker Deskilling has to do with the "how", as in how should an employee perform the manual steps or make a decision to achieve a business task. This "how" is actually codified into the RPA process as discrete steps and can be extracted as documentation retroactively and taught back to employees in case the steps have been forgotten. "Reduced Understanding of Business Logic" concerns both the "how" and the "why". "How" to perform the business steps is impacted because previous human-based decisions, or large chains of formal logical steps can be replaced by black-box ML models. The "why" or the purpose of performing the process steps is reduced as people in the business unit are increasingly removed from the daily operations of completing the work, as the actual monitoring is usually performed by a Process Controller. In an extreme case, all of the staff members who worked a process manually might transfer elsewhere or leave the firm after implementing IA, leaving only the IA steps and a black-box model as the authoritative source of

how and why the process is performed. The risk mitigation measures needed to address this would need to either codify or maintain the decision-making skills of staff, and carefully document manual steps and the reasons why they are performed as they are.

If incorrect predictions are made, the automated process will continue and perform processing steps that it should not have. Imagine a business process at a bank which processes hundreds of personal bank account opening applications per day, and that a ML model is used to perform background checks and determine whether an application represents a real person or not. If the model is not tuned properly, or if the nature of the data is changed, many fraudulent bank account opening applications will be processed, with their data being erroneously entered into many other systems. Hundreds or thousands of bank opening applications might need to be undone by the time the issue is discovered. This delay between when a misprediction is made, the amount of incorrect work that is done as a result, and when it is finally discovered is known as “**Time Lag Effects**” [112]. At the very least, incorrect work must be undone or deleted, leading to lost time, but there could be financial and reputational repercussions as well if the automated work directly affects customers or other firms.

4.4.2.3 Machine Learning Model Risks

ML model risks involve the specific ML model used in an IA solution. The use of ML inside of a business process can result in new security risks, opening up the firm to new attack vectors specific to the ML model. These are known as “**Adversarial Attacks**” [14]. The field of adversarial machine learning appeared in 2004 has gained prominence over the past 10 years [113]. Even if an attacker has no knowledge on the inner workings of the model, no access to the training data or no understanding on how it was trained, they can still make an attempt to query the model and receive some sort of feedback. If the outcome of a black-box model is observable, an attempt can be made to attack it. Developers or data scientists whom have access to the models and data can also leave backdoors to manipulate the prediction results after a model has been deployed into production. The ways to mitigate the threat of this risk are to take proactive measures in modelling the threats to a developed ML model and through IT security controls.

[114] reports that ML models in production suffer from progressive “**Performance Degradation**” over time. The primary reason is due to data drift (discussed in the next section), but other reasons can include changes in business requirements, technology and the environment, leading to the deployed ML model or technique no longer being fit for use. Models and technologies can be rendered obsolete due to newer, more predictive ML algorithms or the deprecation of libraries used for implementation. Addressing this risk requires a plan to regularly monitor data, monitor model performance, retrain models and update technologies to ensure that the accuracy and performance for an IA solution are maintained. Addressing this risk necessitates a reduction in the ROI due to increased maintenance costs.

Transfer learning refers to using an existing ML model as a base for the development of an application specific model. The use of transfer learning represents a breakthrough in ML, allowing models to be built more quickly, and less expensively, especially for image and text

processing [115]. An example of this would be using a model that was trained to identify animals generally from images to identify between different dog breeds. “**Transfer Learning Bias**” [14] occurs when we develop a model on top of a base model that is biased to begin with. When new connections are built on top of an already biased network, the bias remains and will influence the predictions made on the new model in unpredictable ways.

4.4.2.4 Data Risks

The root cause of many of the socio-organizational and operational risks can be traced back to issues with the data. “**Data Bias**” was the most common risk uncovered in this SLR with five mentions [87], [14], [84], [112], [75]. The meaning of “biased data” is unclear from reading the SLR literature as it was used in a very broad sense. The only commonality between how biased data was used in the literature, was that the data has some sort of “undesirable properties”. Practitioners will need to understand what actually comprises “biased data” in order to fully understand what this risk entails. [116] proposes five sources of biases that occur at different points of the machine learning pipeline. A representation of the five types of biases and when they can occur in a standard four stage ML lifecycle [117] is shown in the figure below.

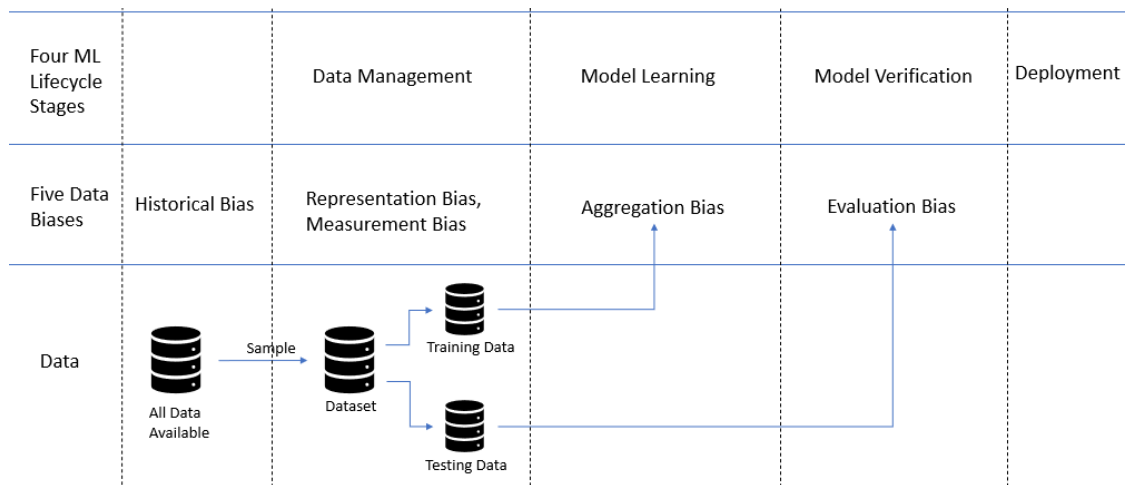


Figure 4: Five Data Biases [116] and Where They Occur in the ML Lifecycle [117] (Source: Author)

First, there is **historical bias**, which exists in the world as it is today; even with perfect sampling and feature selection, bias would still be present in the data. Historical bias tends to reinforce stereotypes of a particular group. For example, the prison population in the US has a disproportionately high number of African Americans (38%) despite representing only 13% of the overall US population [118], [119] caused in large part by social and economic disadvantage [120]. Any system that uses this data, even though accurately captured, could cause harm to that group of people through the reinforcing of stereotypes.

Representation bias is when a certain population is underrepresented in the data when it should not have been. This bias usually occurs during the data collection phase of ML. Potential reasons for representation bias include the method used for data collection (online data collection will exclude those without Internet access and the elderly), insufficient data collection on certain populations, geographies, or older data being used for newer model building. A result of this kind of bias is poor predictive power on the groups that have been excluded during the data collection process.

Data that can be captured and used in a ML model is often just a proxy for what we actually want to measure. An example of this is using the number of arrests as a proxy for the crime rate, which is much higher. **Measurement bias** occurs when the proxy data that is captured is different across different groups. Measurement bias can arise in multiple ways. First, the act of recording the measurement can change the behaviour of what is being observed. Monitoring a group of factory workers will likely induce different patterns of behaviour than monitoring office staff. Next, the quality of data can differ for different groups, for example there is a large difference in the reliability of COVID data from different countries [121]. Finally, if the ML model is supposed to predict a proxy label, that label can have different meanings for different groups. An example is using GPA as a proxy for university success. While GPA might be the appropriate proxy measure of university success for some students, other students may weigh other forms of success factors more heavily such as the ability to find jobs or high salaries.

Aggregation bias occurs when a single machine model is used to predict against a population when multiple models should have been used instead. For example, it is known that the clinical effectiveness of certain drug treatments differs between genotypes and ethnicities [122]. A “one size fits all” ML model, even if the training data has balanced numbers of ethnicities, would not produce optimal predictions as there are differences in the genetic makeup of different races and how they respond to different drugs.

ML models are developed using training data and evaluated against a testing data set which is held out during training. The prediction performance on the testing data is used to compare different models against each other, pending final selection of one for production use. **Evaluation bias** occurs when the testing data is not representative of the target population. The result of this is that the final model will consistently underperform on certain segments of the population despite the overall prediction performance being high. The main purpose of using testing data is to discover and penalize problems in the training data and model. Evaluation bias has a high chance of occurring if there is already representation bias in the complete dataset, and random sampling is used to split it into training and testing data. The way to avoid evaluation bias is to ensure that the evaluation data is balanced.

Some of the most commonly protected attributes include race, age, gender, religion, marital status and socio-economic status, but these can be relaxed depending on context. The conventions and laws regarding which attributes are protected differ by country and function and must be understood before being used as a basis for making decisions.

“Data Drift” [25], [14], also known as concept drift, refers to changes in underlying data trends or distributions that make them no longer representative of the original training data. Under IA, data drifts can lead to large volumes of incorrect predictions and incorrect actions performed based on those predictions. While there are some automated methods for detecting data drift, the vast majority of data drift detection is done manually [123]. This typically involves costly continuous monitoring and evaluation of the data, to identify changes in the trends or in the population distributions. The risk of data drift is closely related to the risks of mistrust in model predictions and performance degradation.

The need to consider the **“Data Privacy”** of customers is an issue in IA [30], [14], [74], [81]. The concept of data privacy can be separated into two parts. The “privacy” component concerns how personal information is collected and used, and the “security” component concerns how data is prevented from unauthorized transmission and access [124]. All four SLR papers which highlighted data privacy as a risk, did so from a security perspective. Perhaps consent was already obtained to collect and use the data in those instances. Both privacy and security are closely related to compliance and regulatory risks – meaning a risk of fines if improper measures taken to secure the use and access of data.

A key determinant in the success of any IA project is the **“Data Quality”** [125], [85], [30], [126]. A lack of data quality typically results in lower predictive power and possibly results in the IA project getting shuttered. Data quality is also composed of a number of factors [123]. First, the **number of data samples** must be sufficient. The question of what minimum number of data samples is needed is an open problem [127], but the general viewpoint is the more data the better. Next is the **proper data structure**, which affects the amount of processing needed to bring the data into useable format. Next is the data **cleanliness**, or lack of errors or noise in the data. Data **completeness** refers to the lack of missing columns in the data, which may make data samples unusable. A quality dataset implies the presence of **highly relevant features** to the prediction problem at hand. This is completely context specific. Quality is also linked to **bias** as an imbalanced dataset is thought to be of low quality. [126] and [125] takes a high-level view of data quality, noting that poor data quality will result in poor predictions. [30] describes the low quality of scanned documents, skewed images and old datasets being responsible for poor OCR results. [85] describes the lack of a reliable ground truth and the need to perform lengthy data collection and to filter out good from bad data to build a model with. Data quality is closely tied to the risk of low predictive performance.

4.5 Risk Mitigation Techniques

The 15 extracted risk mitigation techniques can be organized by where they can be applied during a typical IA project lifecycle. In the figure below, risk mitigation techniques have been organized into four categories. The techniques under “Planning and Due Diligence”, “Algorithm Selection” and “Human Interaction Design” can be used during the planning and design phases before an IA solution is deployed for use. The techniques listed under the

“Operations” category are to be used post-implementation on an ongoing basis, for as long as the IA solution is in place.

The bidirectional arrows in the figure below indicate that the choices of risk mitigation interact with other potential choices. For example, choosing an “Explainable AI” algorithm may reduce the need for “Human-in-the Loop”, and selecting “Random Sampling” may affect the “AI Liability Terms in Contracts”.

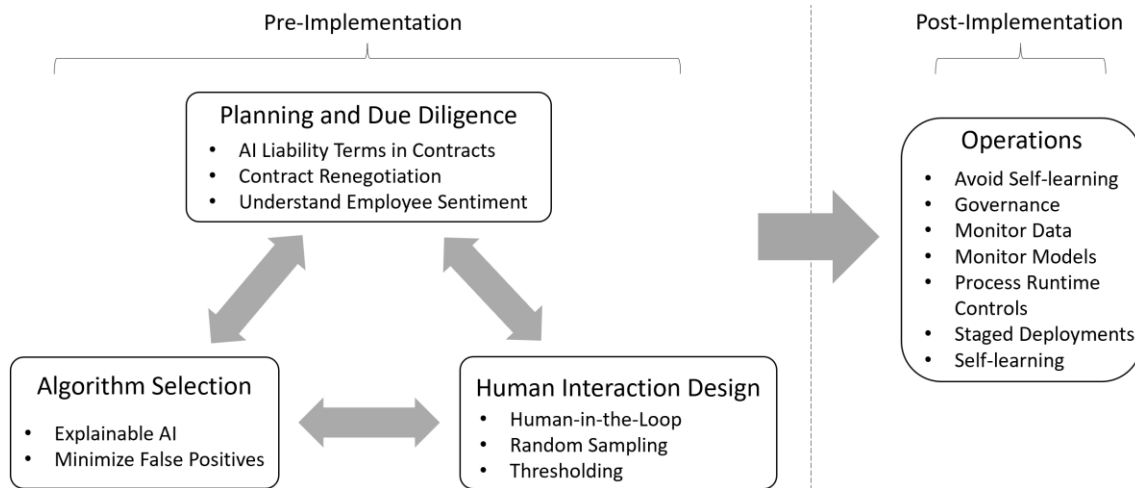


Figure 5: Four Categories of Risk Mitigation Techniques, Pre and Post-Implementation (Source: Author)

4.5.1 Planning and Due Diligence

Prior to undergoing an IA project, an impact assessment of the proposed automation on existing contracts with third parties should be conducted, especially if they are ongoing and of long duration [83]. “**Contract Renegotiation**” can be seen as both a risk mitigation and a value capture technique. On the one hand, the addition of AI decision making may introduce errors in the execution of day-to-day work, leading to intermittent SLA failures. On the other hand, automation may significantly reduce the amount of time needed to turnaround work to the customer. Contracts affected by IA should therefore look into renegotiating SLA breach conditions and look towards capturing value from general improvements in service levels if necessary.

Also discussed in [83] is the idea of using “**AI Liability Terms in Contracts**”. If a third party is responsible for the development or maintenance of an ML solution used in IA, and this causes losses to either the firm or the end users, who should be held responsible? ML liability is an active field of discussion in law, as the rate of technology development has far outpaced legal discourse [128]. It may be impossible to determine which party should be held accountable after some loss has occurred. None of the reviewed literature during the SLR spoke in depth about the legal aspects of using IA, so additional papers in the legal field were examined.

Čerka et al. indicate that in almost every common and civil legal systems, the “Vicarious Liability Doctrine” would apply to AI [129]. This means that the supervising party (the firm) can be held liable for actions or omissions taken by subordinates, such as employees or ML algorithms. If the ML algorithm was created by a third-party, the firm could seek legal damages against the third-party. However, the firm would have to argue that the AI was defective during the time when some damages occurred and that the AI defect was the cause of that loss, which may be difficult to prove. Under IA, it would be easy to show than an incorrect prediction was made, as there as audit logging systems that tracks every step performed in the automated process. However, this does not prove that the algorithm was defective. Ultimately, the firm that is executing the IA process is likely to be held liable, unless a third-party provider of ML is willing to take on additional liability (that they normally would not have to bear), which seems unlikely. Although trying to add “AI Liability Terms in Contracts” sounds like a good idea, it would be difficult to achieve in reality.

Instead, when dealing with a third-party AI provider, a firm may wish to reference Truong and Nguyen who have provided a list of possible ML attributes, contract constraints and monitoring methods that can be used to define ML service level performance clauses in contracts with third parties, such as data quality and prediction accuracy [130]. It would be much easier to prove a breach of a measurable ML service attribute, rather than trying to prove that there was some defect in the algorithm. However, including ML specific attributes inside of a contract is very forward-thinking as even the incumbent AI vendors do not include them in their standard service contracts [130] . Since the addition of “AI Liability Terms in Contracts” risk mitigation technique is not very practical, it should be transformed into including “Measurable ML Attributes in Contracts” instead.

Table 27: Revising Risk Mitigation Techniques

Original Risk Mitigation Technique	New Risk Mitigation Technique
AI Liability Terms in Contracts	Measurable ML Attributes in Contracts

Zhu et al. [75] present the risk mitigation “**Understand Employee Sentiment**”, under the context that there are four profiles of AI adopters, some of whom are more willing to adopt or work with AI technologies, and others who do not want to work with AI. They argue that it is crucial to understand and plan around the profiles of the people in the organization to better ensure the success of AI endeavors. Otherwise, non-adoption of the solution or even sabotage of the project may occur. The four profiles from [75] are reproduced below in a 2x2 matrix. The axes of the matrix are separated by emotional attitude, and rational attitude towards AI. AI Reticents are rational in acknowledging the that technologies have value, however they are reluctant to embrace it due to their own negative feelings. AI Intrepids are accepting of AI technologies and need not be worried about from a risk management perspective. AI Dissenters have the least favorable views of AI, from both a rational and emotional perspective. They believe that there is little value in implementing AI and may be more vocal in opposing its use in the workplace. Finally, AI Skeptics have interest and optimism about AI, but do not feel that the

commercial value or maturity of the technology is there yet. Age, gender and position (for example front-line vs. managerial) was not found to vary significantly between the four categories.

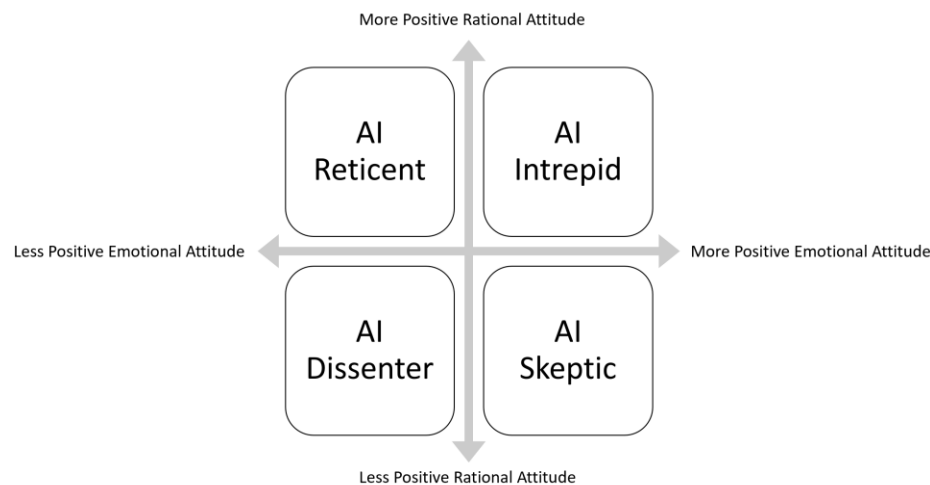


Figure 6: Understanding the Four Profiles of AI Adopters (Reproduced from [75])

From a risk management perspective, staffing an IA project with AI Intrepids and AI Skeptics will help improve the chances of implementation success. Both groups have a positive emotional attitude towards AI, but AI Skeptics remain unconvinced of the commercial value of implementing AI technology works. Skeptics however, could be convinced of the value of AI through participation in a project and experiencing positive results firsthand.

In terms of the target employee group whose tasks are to be automated, it may make sense to avoid automating tasks that belong primarily to AI Dissenters. They are the group that worries most about job loss and dislike changes to their current ways of working. Automating the work of Dissenters may lead them to develop increasing levels of job stress, unhappiness or even lead them to quit. If Dissenters are the target of an IA project, their concerns must be actively managed by the project team. The authors in [75] did not provide a methodology for eliciting which employees belong to which category, however, they do suggest holding group discussions, encouraging feedback and careful listening. Questions can be designed across both the rational and emotional side, with rational questions asking more about their perceptions on the commercial value of AI to the firm, and the emotional side discussing the impacts on people's jobs and the future of work.

An extended description of the "Understand Employee Sentiment" risk mitigation technique is to identify those with positive emotional attitudes towards AI, and to use them as project team members or the primary targets for IA.

4.5.2 Algorithm Selection

“**Explainable AI**” [126], [14] is proposed as a way to add transparency so that the business and management can understand why a ML prediction was made. This topic was treated in depth in **Section 4.2.3 Interpretability**. Current explainable techniques are not perfect, and require expertise to use. They also require access to the underlying data used to train the model, making them unsuitable for outsourced model development. The computational requirements are also very high. Finally, there do not seem to be any public libraries available for explaining third-party developed models. Interpretable methods require statistical knowledge to understand, so training to business users and management is needed if they need to understand how a prediction was made.

A study [131] conducted roughly 20 interviews of companies looking to implement explainable methods, and 30 interviews of firms that have actually implemented explainable AI techniques. For those firms that have not yet adopted explainable techniques, explainable techniques are perceived to help understand why performance is poor, to monitor model drift in feature and prediction distributions, increase transparency, increase compliance with current and future regulations and enable internal auditing of models. Almost all of these are risks that have been uncovered by this SLR. Among those firms that have implemented explainable methods, none used any global techniques. The focus was on local explainable methods and the major use case was to understand the importance of specific columns of data to the model. Explainable method output was primarily consumed by engineers and data scientists as sanity checks and to improve their understanding of the model. End users face large barriers in understanding explanation output until methods are improved. This is at odds with the stated reason for using explainable methods, which was to inform management, end users and potentially courts of law on why a decision was made. The conclusion is that the current state explainable methods are not fully ready to be used for risk mitigation purposes, except as an auditing step conducted by data scientists before a model is released into production.

In order of priority for risk mitigation purposes, preference should be given to develop inherently interpretable models first. Next is to try developing non-interpretable models with an explainable method alongside it. Current explainable methods are well-suited for the auditing of models before they put into use. The results of the explainable method should be accepted by management before deployment of the model for use in production. The riskiest way to deploy IA would be to use models developed by third-parties.

Chalmers [80] proposes the use of ML algorithms tuned to “**Minimize False Positives**” as a way to reduce risk in IA. This algorithm is explicitly designed to reduce the number of false positives for classification problems. This is a desirable property when the cost of triggering erroneously automated work is high. The concept behind the custom algorithm is to create a tight classification boundary around positive classes, such that false positives are minimized. A visualization of this is shown in **Figure 7** below. The positive class samples are shown in dark dots and the negative class samples in white dots.

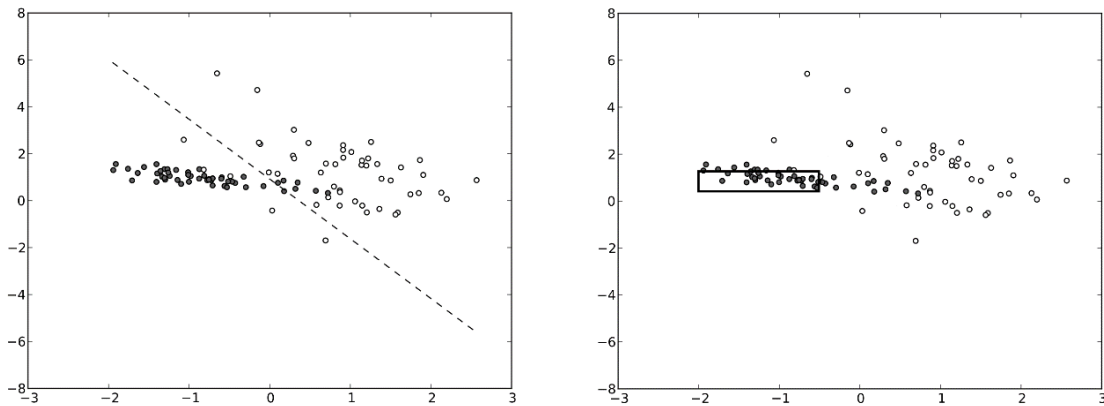


Figure 7: (Left) A Traditional Boundary vs. (Right) A Boundary that Minimizes False Positives (Adapted from [80])

The dotted line in the left diagram shows the boundary for a traditional binary classifier. White dots appear to the left of the line, meaning that there are false positives. The diagram on the right shows a black box that is tightly fit around the positive class. There are no false positives inside, but everything outside of the box is classified as negative, leading to many false negatives. This type of algorithm would be useful in high-stakes use cases where false positives must absolutely be avoided. We can imagine a different boundary for the picture on the left, where the slope of the dotted line is kept the same, but shifted towards leftwards such that there are no more white dots to the left of the line. This would give once again a classification boundary that has no false positives. However, this boundary still has many areas with which there are no training examples. For instance, the area between (-1, -2) has no samples and unknown behavior. It would be risky to classify samples in that area as positive, which is why Chalmers proposes using a tight boundary around known samples.

While the algorithm was implemented and tested by the author in the form of decision trees, the implementation has not been made available to the public for use. It is probably outside of the skillset of most data scientists to re-implement the algorithm for proprietary use, so this risk mitigation technique is out of reach for most firms, until a public version is available.

4.5.3 Human Interaction Design

“HITL”, “Random Sampling” and “Thresholding” all relate to a process design choice that determines when and how a human is supposed to intervene in an IA process following a ML prediction. HITL is a broad term meaning that humans should interact with the automated process in some way. In IA, this human interaction requires determining:

1. The criteria for triggering a human to intervene in the automated process

2. The actions available to the human during intervention. There are two sets of actions available. The first is deciding whether to accept, or modify the prediction. Next, the status of the automated processing should be decided, whether to cancel or to continue
3. Where in the business process should intervention occur

For 1., the criteria for triggering human intervention can be either Random Sampling or Thresholding. Under Random Sampling, a percentage is chosen (for example 90%), and a random number between 1 and 100 is generated whenever a ML prediction is made. If the random number is under the chosen percentage (90 in this case), that IA prediction will be recorded and automated processing will stop, pending input from a human. After human validation of the prediction, the person can choose to continue automated processing of that work case. In [68], where Random Sampling was proposed, a split of 75% automated to 25% manual processing was adopted, although the purpose was not to reduce risk, but to measure the accuracy of the predictions. Although the authors did not explicitly call out Random Sampling as a way to reduce risk, it can be used for that purpose by routing a certain percentage of automated predictions to humans for validation. One unanswered question regarding Random Sampling, is how to choose the appropriate ratio for a given process and ML model.

Regarding point 3., there are two other ways in which Random Sampling can be used beyond validating predictions. First, random sampling can be used before any automated processing occurs, to force business process to be worked completely manually by people. This can counteract potential loss in worker skill as some cases must be processed manually. The next way that random sampling can be used is at the very end of the process. This can be used to force a percentage of automated processing to be verified by a human. This can reduce the impact of “time lag”-based errors.

Thresholding can be used when the ML algorithm reports confidence intervals in addition to its predicted values or labels. It requires an understanding of the model’s predicted values and the confidence levels that are returned from the model. So far, the papers examined have only brought up a fixed threshold values, which seem arbitrarily chosen. For example, in [72], thresholds of <90% is chosen to stop automated processing completely in favour of manual processing, 90%- 99.5% is used for the range of human validation (that can still lead to automated processing) and >99.5% is used to allow for full automated processing. There can be multiple thresholds for different predicted classes. From [132], an example is given where the algorithm predicts 73 classes. If the final prediction is of a class with known low accuracy (<75%), the threshold is set to a high level to make human validation more likely. If the prediction is of a class with high accuracy (> 90%), the threshold is set to 0, allowing for full automated processing. Similar to Random Sampling, there does not seem to be any guidance on how to choose appropriate thresholds for the model predictions – only that thresholding can differ per class and be used to route between fully automated and manual processing. The distinction between Random Sampling and Thresholding is that Thresholding can depend on the predicted class, whereas Random Sampling ignores which class was predicted.

No paper in the SLR has explicitly mentioned this, but HITL also presents a natural “assignment of liability” to the person who is responsible for validating the prediction. While this

may simplify liability concerns, it could also complicate the situation if validation is outsourced to contractors.

4.5.4 Operations

“Self-learning” [29] (not to be confused with self-supervised or unsupervised learning) refers to the ability for ML models to regularly improve themselves over time through retraining and redeployment into production, without the need for human oversight. Although positioned as a risk mitigation technique, the use of self-learning also leads to risks of its own. The viability of choosing self-learning likely depends on the riskiness of what the ML algorithm does in the process. If the accuracy of the ML prediction is highly critical, then it would require significant governance, to ensure that all of the new data has been already verified as clean and free of bias and that the newly updated model passes numerous tests before being used in IA. **“Avoid Self-learning”** [85] argues against the automatic deployment of updated models into production, as new models need to be vetted to be free from issues. “Avoid Self-learning” is the default state of an IA solution as it would require a very specific configuration to achieve self-learning. However, it is worth keeping as an explicit risk mitigation technique in case self-learning becomes more readily usable in the future.

RPA governance is a framework that encourages desirable behaviors in RPA use in firms, and specifies the rights and accountabilities of those involved. Proper RPA governance has been studied to have similar effects to good IT governance. First, the ROI of RPA projects can by increase by up to 40%. Even between firms with similar overall strategies, firms with proper governance have 20% more profitability than those that do not [107]. While research on IA governance is still nascent, **“Governance”** is viewed as a way to prevent loss of process knowledge through documentation [126] and a way to reduce the effects of data flaws, which include bias and quality [111].

IA governance as a field of study is essentially non-existent. The two parts of IA governance – RPA and AI governance are also only just receiving academic attention. Searching for the term “AI Governance” on Google Scholar shows that the most highly cited paper has fewer than 150 citations [133]. Given that IA is a combination of RPA and ML fields, IA governance will likely require the fusion of RPA and AI governance concepts. ML governance is able to reduce three classes of risks [134]. The first is model integrity, or the correctness of the prediction output. Some of the identified reasons for integrity issues come from outliers, data drifts and adversarial attacks. The next class of risk that governance can reduce is privacy, or access and storage of data. The third class of risk that can be controlled is fairness, or bias issues. Fairness issues largely stem from the data collection and data processing steps of ML.

[133] proposes a three-layered AI governance model. The top most layer is for society and law to define the regulations and legislations concerning the use of AI. Below that is the ethical layer, which can act at the firm level. Here, companies can define the ethical standards of using AI throughout the firm. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems group has published a comprehensive report [135] that can be used as a starting point

for the design of such ethical IA guidelines. The bottom-most layer in the governance model is the technical layer, which is where the algorithms and data reside. While other conceptual firm-level AI governance models exist [136], [137], none go into implementation details or seem to have been tested in real-life. Regardless, proper IA governance, combining aspects from RPA, AI and data governance will play a big role in addressing the risks of IA.

Governance seems to be a catch-all risk mitigation technique that can be molded to fit almost every risk. The specifics of what governance must include to address specific risks will be discussed in **Chapter 5: Risk Register Development**.

Two important activities that belong in IA governance include “**Monitor Data**” [114], to detect changes to the data distributions and “**Monitor Models**” [138], [114] to ensure that models are fit for use given changed data, new deployments and new business needs. One challenge of monitoring data drifts is that it must be automated to be effective. First the baseline statistics on the data is collected, typically based on the training data. Then, new statistics are constantly calculated from data streams and presented back to the IA team for regular evaluation. Typical values that are graphed over time per feature are the mean, maximum and minimum. Monitoring the model performance is less straightforward than monitoring the data, because the most it requires finding out retroactively what the correct prediction is. This would require sampling and human verification of the model’s predictions. If this is not possible, the proportions of predicted labels (for classification problems) or the mean of numerical predictions can be visualized and compared to those during previous time periods.

“**Process Runtime Controls**” [112] allows for the real time assignment of who will perform continued processing after a ML model prediction , for example a human vs. a robot and under what supervisory conditions. This allows for risk reduction in a few ways. First, if there is an issue detected in the ML model, most or all of the continued processing can be routed to a human instead. Next, if there are SLA requirements that risk being breached, financial risks can be reduced by increasing the amount of work done by robots compared to humans, or by lowering supervisory requirements. This risk mitigation technique interacts closely with the three Human Interaction Design techniques. This implies that additional business logic must be added into the RPA process steps whenever ML is used, to allow for the tuning of the thresholds and random sampling at run time, so that changes to the parameters take immediate effect on the IA process.

“**Staged Deployment**” strategies of AI models [25], such as blue-green deployments, canary deployments, multi-armed bandit services and A/B testing can be used to reduce service interruption (and thus performance breaches) and is a part of achieving self-learning. The choice of a strategy largely depends on the risk appetite and cost-sensitivity of the firm. Blue-green deployments duplicate the existing ML infrastructure so that both new models and old models can be queried, but traffic is directed to the new models. Once live-testing is complete, the old model and related infrastructure can be decommissioned, or prediction traffic can be immediately reverted back to the old model if issues are discovered. The advantages of blue-green deployments are no downtime and straightforward rollbacks. The major disadvantages are the infrastructure costs needed to maintain a second environment and the risks of directing all predictions to the new model. Canary deployments are similar to blue-green, the main difference

being only a small percentage of production traffic is sent to the new model. A/B and Multi-armed bandit schemes can be used when firms are unsure which model is better, and it requires testing against production data.

4.6 Chapter 4 Summary

This chapter began with an examination of how much risk is present in the implemented IA use cases (Relevance Score of 2). I found that there is a strong tendency for firms to implement low risk use cases. Firms are already performing some form of risk management, although not necessarily in a formalized or explicit manner.

Some of the identified risks centred around the lack of trust in predictions, or algorithms having low predictive power. This required an understanding on how performance is measured in ML, leading to a discussion on Confusion Matrices and Confidence Intervals. Confusion Matrices are used in classification and they give us the exact proportion of incorrect predictions and how they were incorrect. The Confusion Matrix is available for inspection when an algorithm is developed in-house. 29% of use cases were developed using commercial services, meaning that Confusion Matrices would not be available. Instead of Confusion Matrices, Confidence Intervals representing how confident a model is of the prediction can be used instead. These Confidence Intervals can be used together with Human-in-the-Loop mitigation techniques to control risk.

ML interpretability was also discussed. Many people believe that individual predictions will need to be explainable to affected customers, managers and in courts of law. This implies that local interpretable methods for ML must be part of the IA team's skillset. The two most popular locally interpretable methods are LIME and SHAP. However, from a practical standpoint, they cannot be used against commercial models in their current state of development and the explanations still require translation into layman's terms by someone with specialized knowledge in ML or statistics. The main use for explainability methods currently is as an auditing step before a model is released for use in production. Python was found to be the dominant programming language, with every other technology having a Python interface. This includes the interpretability methods of LIME and SHAP.

Next, the 36 risks were categorized in to two main sections. The first category was Socio-Organizational risks and contained 22/36 of the risks. The next category was Operational risks. Socio-Organizational risks arise due to tensions between different actors in the business ecosystem. We have external interactions between the firm and the environment and the firm and third parties. There are also internal risks within the firm, between the enterprise, management and its people. After categorization, each of the 22 risks were examined one by one in more detail to uncover their impact and to gain an idea on how they might be addressed.

The remaining 14 risks belong to the Operational category. These appear during the planning, implementation or day-to-day operations of an IA project. The risks here were further sub-divided into four categories: Project, Process, ML Model and Data. These risks were also

individually examined in more depth to understand their impact and potential mitigation measures.

Finally, I discussed the 15 Risk Mitigation Measures that were surfaced during the coding process. They were split into two categories: Pre-Implementation and Post-Implementation. Each mitigation technique was then discussed in more detail to determine which risks they could potentially address. In the following chapter, I build on the findings in this chapter to develop a risk register that can be used in practice.

Chapter 5: Risk Register Development

After analysing the risks and risks mitigation techniques in more depth, we turn to the development of the risk register itself. Risk registers are risk management tools that are used to communicate risk to relevant stakeholders. Common items included in a risk register are: risk category, name, description, impact, likelihood, a quantitative risk rating, common mitigation steps and owner. The work in **Sections 4.4 Risks** and **4.5 Risk Mitigation Techniques** provide categories, names, descriptions and mitigation methods. Risk registers are typically visualized as tables or scatterplots. The use of risk registers for managing organizational and project risks are recommended by the Project Management Institute [139], PRINCE2 [140] and ISO 31000 [141].

5.1 Risk to Risk Mitigation Mappings

Many of the risks have been mentioned without any risk mitigation technique and vice versa in the literature. However, 15 of the papers in the SLR have already suggested a mapping of some risk mitigation measures to specific risks (see **Appendix: J. Implied Risk to Risk Mitigation Mappings in the Literature**). This initial mapping of suggested techniques for each risk is shown in the “Direct Risk Mitigation Mapping from SLR Papers” column. In this column, HITL is mentioned nine times. Governance, Explainable AI, Thresholding and Staged Deployments are all mentioned twice. After analysing the risks and risk mitigation methods in **Section 4.4 Risks** and **Section 4.5 Risk Mitigation Techniques** in more depth, additional mappings are suggested, shown in **Appendix: K. Risk to Risk Mitigation Mappings** in column “Risk Mitigation Mapping from Discussion”.

5.2 Governance

Governance is one of the “universal” risk mitigation techniques, which helps mitigate 16 of the 36 risks. This section aims to call out specific elements of governance policies that are needed to address the risks where governance is cited as a mitigation method.

5.2.1 Define IA Specific Process Selection Criteria

Quantitative scoring is frequently used to rank and prioritize potential business processes for automation [142]. A number of the IA risks suggest additional scoring criteria that can be used to help determine whether a process should be automated or not. For “Attract Competitive Response”, take careful consideration of how visible IA will be to competing firms and customers. This could take form of a penalty being added to the scoring criteria if IA is publicly facing can be a way to incorporate this risk. In order to prevent “Control Flow Drifts”, choose processes that are mature, stable, and less prone to business changes. Score and penalize processes based on these dimensions.

Switching from cost benefit analysis to a value-based analysis can prevent “Low ROI” and “Unmeasurable ROI” IA projects from being deprioritized. A SLR study focused on the measurement of RPA benefits, containing a list of measurable attributes to consider beyond just “time saved” can be found in [143]. The value analysis of an IA project will have elements from both RPA and ML.

Part of the governance board’s work is to select which processes should be automated next. In order to combat “Information Asymmetry”, the impact of concentrating IA benefits in certain teams or departments should be carefully considered and potentially avoided.

5.2.2 Leverage Existing Policies

From personal experience, it is surprising how often internal policies are not consulted before setting up an IA capability in an organization. Barriers to overcome include discovering which policies exist, what policies are relevant and how they apply to the IA project at hand. Some key policies that should be surfaced for an IA project include data privacy, data retention, information security and IT. Following these internal policies during the setup and operation of IA should reduce the risk of “Compliance” and “Regulatory” issues.

Having strict security standards can reduce the risk of “Adversarial Attacks”. Securing access to production credentials, including the URLs or IP addresses of model endpoints and authentication keys will prevent internal employees from knowing where the model is hosted. Limit querying of the model to only certain groups of white-listed computers. Enable access logging to track which user accounts have accessed the list of white-listed computers. Enable rate-limiting on the models to prevent the model from being queried too quickly. Physical security controls can also be used to prevent access to production machines inside of a datacentre.

Preventing access to data according to the principle of least privilege will reduce “Data Privacy” risks and prevent data leakage. Keep sensitive data masked or encrypted while at rest. Require encrypted communication channels when transferring sensitive data. Penetration testing can be a part of the IA policy to ensure that data is stored and accessed securely.

While security and data policies are widespread, AI Ethics policies are mostly only found in multinational technology firms [144]. If no policies are present, the governance team should lay out their own guidelines to reduce the risks concerning the “Ethics” of IA. Frameworks that can be used as a starting point for the design of such guidelines include IEEE [135], the Future of Life Institute’s Asilomar AI Principles [145], and the Montreal Declaration for responsible AI [146].

5.2.3 Measure Employee Impact

A number of risks deal with the impact of IA on employees post-automation. This is often overlooked as IA are often executed on a project basis with consulting firms that focus solely on implementing business logic and not human needs. Governance of employee impact should be done in conjunction with HR, where feedback is solicited from employees throughout the automation lifecycle. Work overload more generally can be measured using four survey questions from [147]. Work overload can be used as a proxy for “Cognitive Work Overload”, as it is more specific. While governance can measure work overload, actual risk mitigation measures will need to be addressed at the individual level.

In **Section 4.4.1.2 Enterprise Risks**, I discussed that employee turnover is caused by numerous reasons. Organizational support, such as education on IA and training on new skills can reduce employee’s intention to quit. Governance should include participation with HR in the monitoring of employee turnover intention, which can be measured using the TIS-6 scale [148]. Measurements should be taken before and at regular intervals after IA implementation to assess to what degree “Employee Turnover” remains a risk.

5.2.4 Set Baselines and Monitor Data

Proper data and AI governance can help reduce the data risks identified in this SLR. Provide training to data scientists on the five categories of “Data Bias”: historical, representation, measurement, aggregation and evaluation. This training should include how to identify them in the datasets and in the models. Ensure that the class frequencies in the training and testing data are balanced. “Data Quality” can be assessed in parallel with data bias. Establish baselines to ensure that the data is fit for use for a specific ML problem. For example, that there are at minimum 1000 data samples of each class, that it can be readily converted into the proper format for consumption, that there are fewer than 5% missing column values etc. Some frameworks for assessing data quality can be found in [149], [150] and [151].

“Data Drift” requires active and regular monitoring of data statistics after they are free of bias and quality issues. These statistics include minimum, maximum, standard deviation, mean and average values etc. Keep track of these statistics for each feature over time. The three major ML cloud platforms, Google, AWS and Azure offer graphical tools to help monitor data drift [152], [153], [154]. Broader governance principles for the management of data, extending past data quality, bias and drift can be found in [155].

5.2.5 Document

The documentation of work, roles and responsibilities is a key part of effective governance. Extensive documentation of both the process steps and the ML model should be performed to counteract the risk of “Reduced Understanding of Business Logic”. Effort should be spent explaining the purpose of the process and why it needs to be performed as it is, since the mechanical steps needed to execute the process can be inferred from the IA tool itself. Change requests to the process and ML model should also be put into place to document the reasons leading up to a change. Documentation on whom is responsible for correcting predictions in HITL should be done for clear “Prediction Accountability”.

5.3 Mitigating Unaddressed Risks

After performing the initial risk to mitigation mapping (see **Appendix: K. Risk to Risk Mitigation Mappings**), six risks without any corresponding risk mitigation remained. Despite not surfacing any mitigation methods for those six risks during the SLR and data discussion, it would be amiss for me to ignore them. In this section, I examine these six risks in more depth, to uncover potential risk mitigation measures, quantification methods and practical guidance for use by industry practitioners.

5.3.1 Unaddressed Risk 1 - Departmental Resistance

Studies have been conducted to understand the reasons behind management resistance in the face of office automation [156]. The top reason for opposing changes, especially for senior management was a “lack of knowledge”. For the rest of management staff, the reasons in order of importance were: “not being convinced”, “insufficient training” and “individual job importance threatened by change”. The notion of “job importance threatened by change” is exactly what the authors in [75] reported in their study. Proposed ways to address resistance to change from [156] include “increasing knowledge”, “convincing company employees”, “gradual introduction of changes” and “training”. Increasing knowledge and convincing company employees can be done through training and participation in the process. Education materials should contain the data from numerous studies, showing how RPA rarely reduces headcount, since only tasks are removed from workers as opposed to complete jobs. This fear of losing headcount due to automation was directly called out as a major concern of managers.

Resistance to IA adoption at the department or managerial level shares much in common with resistance to traditional RPA adoption. The main difference lies in the replacement of knowledge tasks vs. manual tasks. Specific to addressing the risk of losing headcount and budget, if the amount of time saved by automating knowledge-work is predictable, the affected department should brainstorm and explicitly define to upper management value-adding ways in which that time will be filled. This will likely help prevent headcount from being removed.

5.3.2 Unaddressed Risk 2 - Loss of Job Meaning

Numerous scales to measure and discover the factors underlying “meaningful work” have been proposed, including WAMI, which contains five factors [157] and CMWS, which contains 28 [158]. A comparison of the various meaning work scales can be found in [159]. The conclusion from [159] is to use WAMI to examine the relationship between experiencing meaningful work and outcomes. CMWS should be used when examining how work tasks and organizational practices create meaningful work. For the purposes of IA, WAMI would be the appropriate to use as a questionnaire to assess whether job meaning has diminished among employees after IA implementation.

[160] provides a four-level framework that can be used to foster meaningful work in a firm. The levels in the framework are: individual, job, organizational and societal. As examples, if we believe that IA will necessarily reduce job meaningfulness, the firm can provide more autonomy to affected staff (individual level), more significant work (job level) and increase corporate social responsibility endeavours (organizational level) to counteract this risk.

Perhaps more immediately practical would be to use common sense and avoid automating tasks that we believe are core to someone’s job meaning. Invite those affected by IA and hear their opinions on whether certain tasks out of their day-to-day work should be automated.

5.3.3 Unaddressed Risk 3 - Loss of Job Security

Job security can be quantified using two distinct measures [161]. The Job Security Index measures an individual’s perception of how stable their job is. The second, Job Security Satisfaction scale measures the attitude of that person regarding their level of job security. Probst in [162] proposes a five-layered model of job security. The first part of this model involves worker characteristics, including tenure, education level, absenteeism etc. Next are job characteristics, such as contract type (temporary vs. permanent, part-time vs. full-time etc.) and union affiliation. Third are organizational change factors, such as official announcements of layoffs or mergers. The fourth is proximity to the “core” functions of the organization, such as being in a technology role in a technology firm, or a legal role at a law firm. Positions located outside of the “core operating areas” of the firm are often viewed as less secure. The final layer of this model is technological change, meaning as the company matures, so do the complexity of their systems, which increases worker requirements. Undergoing IA would fall under this level of the model.

A number of the factors making up this five-layered model can be applied to reduce the sense of job insecurity. If the purpose behind IA is to not reduce headcount, as it often is not [163], publicize this and make the actual measurable goals of IA known to alleviate fears. Commit to not cutting jobs, which is a clear message from the organizational change level of the job security model. If the purpose is to actually cut jobs in one area, prepare training plans and new job role definitions for those whose tasks will be automated by IA with HR. This acts on the

education level and organizational commitment factors, which both reduce the perception of job insecurity.

5.3.4 Unaddressed Risk 4 - Mistrust in Management

The factors affecting manager trustworthiness are examined in [99], [164] and [165]. Three factors are most commonly cited: ability, benevolence and integrity. “Ability” or competence is the belief that management have good technical skills and an understanding of the business. Next is “benevolence”, which includes courtesy during employee interactions and the responsibility to inform. Examples of benevolence include views that management discussions are fruitful, that management keeps employees informed and that they provide constructive feedback. “Integrity” relates to compliance and procedural fairness, for example, the view that job performance ratings are done fairly and accurately.

From an “ability” perspective, managers should carefully consider the impact of IA on employees and follow a demand generation approach for adopting IA. A top-down approach, or forcing teams to adopt IA could lower the view that management has “task competence”. “Benevolence” implies that there is openness in communication with affected employees regarding the justification, goals and impact of implementing IA, and that feedback from employees will be carefully considered and addressed. Measuring organizational trust can be done through Shaw’s organizational assessment. A questionnaire implementing this assessment can be found in [166]

5.3.5 Unaddressed Risk 5 - Reduced Work Preparedness

Automation will make staff less prepared as they become more hands-off during data processing. This can be counteracted by making access to processed information easier, such as dashboarding and summary reports. These dashboards or reports should have user friendly interfaces with search and links to individually processed automation work cases, so that workers can quickly find the information they need.

5.3.6 Unaddressed Risk 6 - Transfer Learning Bias

If transfer learning must be used, then the biases of the original model must be understood. But this is extremely difficult unless the creator of the model publishes what biases exist, or if access to the data is provided for analysis. The reason for using transfer learning is usually because the original model has already been trained on a massive dataset, such as GPT-3 that would be impossible to analyse for bias due to the amount of computing power needed. The only realistic way to completely avoid transfer learning bias when the model or data is not accessible is to not use transfer learning at all, in favour of developing models from data that is controlled by the firm.

In this sub-section, I examined the six risks that had no clearly implied risk mitigation strategies from the list of 15 risk mitigation techniques. Each of these risks were examined in more depth through additional searching in literature, to discover their underlying factors leading to this risk and how those factors can be measured and tracked over time.

5.4 Final Risk Register

The final risk register that can be used as a basis for assessing and controlling the risks of an IA project can be found in **Appendix: L. IA Risk Register**. Some of the columns traditionally associated with risk registers, such as owner, dates, impact etc. are context specific and omitted to facilitate customization.

5.5 Chapter 5 Summary

An initial mapping between risk mitigation techniques to the risks was already present in 15 of the SLR literature papers. This was used as a starting point for developing the risk register. Then, through the in-depth understanding of the different risks and risk mitigation techniques that was developed in Sections **4.4 Risks** and **4.5 Risk Mitigation Techniques**, additional links between risk mitigations and risks were added to the risk register. 30 risks had at least one associated risk mitigation technique by the end of this work.

Governance was found to be a generic and widely applicable way to reduce risk, but the specific components of Governance that are required to be implemented for each risk were not clear. Each risk that had Governance listed as a risk mitigation method was examined in more depth to surface which areas Governance needs to address to specifically reduce risk. This resulted in the definition of five areas of focus that should be added to an organization's Governance model in order to control IA-specific risks.

Next, I looked at the six remaining risks that did not have any risk mitigation technique to address them. This work outlined other ways to reduce the risks that were not found during the SLR, and ways to assess quantify the risk so that it can be measured over time. The final risk register template was then created.

Chapter 6: Conclusion and Final Discussion

"Technology, through automation and artificial intelligence, is definitely one of the most disruptive sources of our age." – Alain Dehaze, CEO of Adecco on the future of work [167].

IA is believed to be among this generation's most disruptive technologies, and a key component in the future of work, where full work automation is achieved by replacing human decision-making with AI algorithms. Numerous media publications, global organizations, politicians and celebrities are actively debating the impact of IA on our lives, today. Despite this

belief and media chatter, uptake of IA technology at the firm-level remains slow, or limited to low-risk use cases. Why is there still a gap between the market hype and reality, when technology-wise we are largely there? A few reasons why come to mind, for example, a lack of data scientists on IA teams and lack of suitable data to build models with. But perhaps another more fundamental reason for the lack of impactful IA is that businesses still do not have the confidence to try implementing it. IA implementation should be transformational and strategic, requiring careful change and risk control. But without the groundwork to discover what risks actually exist, firms will remain reluctant to truly transform their organizations.

Research looking at the impact of IA at the firm-level is almost non-existent; much of the existing research focus is placed on discovering and pushing out positive narratives and what benefits IA will bring. This thesis aimed to bridge the gap between academia and practice to facilitate the adoption of IA in industry. The current state of the IA industry is examined through four RQs. These research questions aim to discover 1) what use cases exist, 2) what technologies and algorithms are being used, 3) what risks exist and 4) what risk mitigation techniques exist. A SLR was conducted with peer-reviewed academic publications up to and including 2021; 77 research papers were analysed thoroughly and data pertaining the four RQs were coded. The final goal of thesis is to create a risk register that firms can use as a template to tailor according to their unique environment.

6.1 Research Summary

Although many use cases were captured through the SLR, the majority of cases were not directly studied by the corresponding authors. Many cases were theoretical or references to other papers. Even though I did not go through all of the referenced use cases to check to see if they were implemented, it is likely that many of them are theoretical as well. This lack of fine-grained understanding and connection of use cases to technologies and algorithms prevents us from developing good risk mitigation strategies and pushing the IA field further. Currently, IA is still in its infancy as an industry and a field of academic research.

Over half of the use cases centred around document processing, with ML positioned at the very beginning of the process. This includes the use of OCR, document classification and classifying the meaning of words into entities. This is unsurprising as one key way to enable a business process to be automatable is to digitize the input data. The next largest use case was the use of Chatbots as an interface to capture customer or employee commands. This again is a way to digitize unstructured data so that it can be properly processed by a robot.

Overwhelmingly, companies are using ML to structure unstructured data, so that it can be deterministically processed by RPA. In terms of cognitive difficulty, these ML tasks could likely be performed by someone with a high-school level of education. What was largely missing from the use cases – and what is being advertised by consulting firms and RPA vendors are attempts at replacing specialized human-knowledge and decision making with IA.

After the use cases, the ML algorithms and technologies were examined. Document processing and chatbots can largely be implemented through commercial software services; configuration rather than development is needed. But, my study of the literature did find many instances of custom ML development using Python, the only programming language encountered in this study. Regardless of whether an algorithm is custom developed or off-the-shelf, the human-in-the-loop techniques of thresholding and random sampling can be used to greatly reduce the risk of automated decision-making.

Commercial prediction services are black-box, and nearly 80% of the chosen algorithms are as well. The most popular algorithms were Random Forests followed by neural network-based algorithms, neither of which are interpretable. For almost every single IA case, the predictions made cannot be interpreted by business users. When algorithms are black-box, the explainability of algorithms becomes a key concern. Society expects that decisions made through automated processing can be explained for fairness and liability reasons. Regulation is inevitable [128] and legal experts expect explainable AI to play an important role in liability outcomes. Research into explaining ML predictions is still on-going. Current explainable methods, such as LIME and SHAP, have ready to use libraries in Python, making them available to many IA processes which exist today. These methods however have drawbacks, for example they require access to the underlying data, huge computing power and significant expertise to interpret. We are still quite far away from having an average business user being able to explain ML predictions to management or to customers.

36 risks were identified through coding and categorized into two main groups: Socio-Organizational and Operational risks. The Socio-Organizational risks were further broken down into four categories: Environmental, Enterprise, Third-Party and Employee. Operational risks were separated into four categories as well: Project, Process, ML Model and Data. Much of the IA research examined in the SLR did not focus on risks nor on risk mitigation, mentioning them only in passing. Because of this, each risk was examined in more depth to understand what they entail. Each risk mitigation technique was examined in detail as well to see how they could be put into practice.

Finally, a mapping between the risks and risk mitigations were made, in two steps. First through the links directly present in the SLR research papers themselves, then second through the intuition developed by the understanding of risks and risk mitigations in more depth. This mapping was then converted into a risk register that can be used by practitioners in the field, to implement new IA projects.

6.2 Research Limitations

Major limitations of this research include the lack of validation of the risks and risk mitigation techniques. Some of the papers even suggested that some of the risks may not be valid without additional research. Both steps of the two-step mapping between the risks and risk mitigation techniques also need validation to ensure that there are indeed links between the two.

As the source of data for this thesis was peer-reviewed, published academic research, the data captured and used to conduct the analysis is likely not current and may not paint an accurate picture of the use cases, technologies and algorithms used today. IA is industry led with academia lagging behind by a few years, and the publication process takes time as well. Anecdotally, as a consultant who worked with dozens of companies, helping firms develop their IA practice immediately prior to writing this thesis, I believe that the research findings do reflect reality.

Although the coding of fields was done multiple times to ensure consistency, additional authors or industry experts were not consulted. This could limit the reproducibility of coded data as the author may be biased by being an industry practitioner.

6.3 Future Research

The lists of risks and risk mitigations that I developed are far from complete. Additional sources of data can be used to discover both new risks and risk mitigation measures, such as industry-led publications, grey-literature, interviews and case studies. Next, the risks and risk mitigation methods need real-life validation, as some may not be valid. The links between the risks and mitigation techniques also need real-life tests to confirm that the mitigations are able to measurably reduce the change or impact of the risks. As a continuation of my research in the IA field after returning to industry, I will look at these additional data sources to further develop the list of risks and risk mitigation measures, and assess their usefulness in real life projects.

Many publications imply links amongst the risks themselves, and similarly for the risk mitigation techniques. An example of this for the risks are that a loss of job security and a loss of job meaning can lead to employee turnover. Research into discovering these links may lead to a prioritization scheme on which risks to focus on mitigating, and which mitigation techniques have the greatest impact on reducing risks. I would expect the Operational Risks to have links that largely follow the same top (Project risks) to bottom (Data) structure as seen in **Figure 3**. In terms of priority or which mitigation measures to put into place, having a governance model that implements controls for the items identified in **Section: 5.2 Governance**, choosing an appropriate Human-in-the-Loop scheme and developing Explainable AI capability within the team would likely cover the majority of the risks.

Taking a System Dynamics view of the risks and risk mitigation would be an interesting way to investigate the links between them, surfacing inter-temporal relationships (such as tipping point dynamics) as well as feedback loops that can increase or decrease the risk of financial or reputation loss, having adopted IA.

To explore each risk and mitigation technique in depth, I had to go beyond the academic studies conducted on IA. I had to reach into other fields to uncover the full range of risks and mitigation strategies covered here. Without conducting deep-dive studies on any of the risks or risk mitigations found within the context of IA to understand the risk levels, opinions of staff, and impact on the firm, industry runs the risk of running blind.

By developing this list of risks and risk mitigation measures, I have provided researchers with a starting point for further study on the risks that automation and AI pose on the future of work. There are opportunities for researchers and practitioners to quantify and track risks as IA implementations unfold, which will hopefully lead to more robust, ethical and accurate automation outcomes being delivered to customers. Industries that require highly-reliable AI decision-making, such as autonomous driving, medical care and law may also find this thesis relevant to their fields.

6.4 Final Discussion

“Robots will be able to do everything better than us...I am not sure exactly what to do about this. This is really the scariest problem to me.” – Elon Musk, 2018 [168].

IA is opening up new fields of study, business opportunities and business models. Examples of this include a new class of “programmers” called citizen developers, who are able to create computerized process flows with low-code techniques. Another example is business process outsourcing. Imagine if automated business processes and decision making are even further standardized so that they can be used inside firms in a plug and play manner. Start-up companies could be built using “building blocks”, by choosing the abstracted business processes that are needed and chaining them together, such as order to cash, employee onboarding, employee salary payment, etc. This could spur a new wave of innovation and entrepreneurship worldwide.

Two visions of the future of work emerge from the discussion of intelligent automation, settling at extreme ends of the spectrum. The first envisions a utopia, where automation frees us from the drudgery of our daily jobs. The benefits of automation are democratized, allowing people to focus on higher-value, more fulfilling work, or even the choice to not work at all. The second envisions widescale job-loss and wealth inequality. Here, the benefits of automation are captured by business owners to the detriment of workers, wages are repressed and AI technology is used as an arms race to crush market competition.

The most likely scenario lies somewhere between the two. As we are still in the world of narrow AI, only specific rather than general tasks can be replaced by automation. While the New York Times cautioned us that “the robots are coming for Phil from accounting” [1], the chances of reaching full-workplace automation is unlikely to be achieved within the next decade based on the state of AI as it is today – artificial general intelligence (AGI) is thought to be 20-30 years away at the least by experts [169], and some believe that it will never be achieved [170]. Full-workplace automation might be realized once AGI takes hold, allowing us to see which one of the two extreme scenarios will play out.

Under narrow AI, routine decision-making codified into the ML models through training can be handled well by IA. But edge cases that are less represented or even not present in the data will slip through and enter the automated workstream. ML models are not able to react to data from changing environments without retraining and redeploying the model. Even if

automated model deployments are allowed, there will always be data that cannot be captured and input into the model for consideration. Examples of this include, management sentiment, wider market conditions or the need act conservatively during regulatory scrutiny.

Risk management is needed regardless of whether we are living in a narrow or AGI world. High-reliability organizations have different risk appetites compared to low-reliability ones, for example aircraft and surgery vs. e-commerce. Some firms will require their AI and automations to have reliable, fair and accurate ML outcomes every time. This necessitates a methodology to thoroughly investigate, rank and reduce the risks of implementing office automation with ML. Through the identification of 36 risks, 15 risk mitigation techniques and the development of the risk register, I hope that firms will be able to strike the appropriate level of balance between risk and automated outcomes.

Appendix

A. Exact Search Terms used in Document Search Databases

Search Engine	Search Terms
Web of Science https://www.webofscience.com/wos/woscc/advanced-search	(((((ALL=(rpa "machine learning")) OR ALL=("robotic process automation" "machine learning")) OR ALL=(hyperautomation)) OR ALL=("intelligent automation" rpa)) OR ALL=("intelligent process automation")) OR ALL=("cognitive automation")) NOT SO=(INTELLIGENT AUTOMATION "AND" SOFT COMPUTING)
Science Direct https://www.sciencedirect.com/search	(((((("rpa" "machine learning") OR ("robotic process automation" "machine learning")) OR (hyperautomation)) OR ("intelligent automation" "rpa")) OR ("intelligent process automation")) OR ("cognitive automation"))

For Web of Science, an additional search criterion was added to exclude search results from the “Intelligent Automation & Soft Computing” journal which did not contain any RPA related research results based on a separate search.

B. Summary of SLR Search Results

Description of Search Step	Number of Research Articles
Web of Science articles	127
Science Direct articles	412
Total articles	539
Minus duplicates	-72
Total articles after duplicate removal	467
Minus articles rejected based on title and abstract	-332
Minus articles rejected based on full-text reading	-39 (A)
Remaining articles to read full-text	96
Articles rejected based on full-text reading	82
Articles accepted based on full-text reading	14 (B)
Articles used in forward and backward searching	53 (= -A + B)
Articles found through forward searching	3
Articles found through backward searching	21
Total articles used in the systematic literature review	77

C. Research Question 1 (IA Use Cases) Coding Scheme

Name	Allowed Values	Description
RQ1 Relevance	0, 1 or 2	0 – The article does not contain any use cases 1 – The article describes use cases theoretically, or through references 2 – The article directly examines or implements at least one of the described use cases
ML Positioning	(Empty), Unknown, Start, Middle, End	(Empty) – There is no use case Unknown – Where ML is used in the automated process is unspecified Start – ML is used at the very beginning of the automated process

	Multiple values allowed	Middle – ML is used somewhere in the middle of the automated process (as opposed to at the very beginning or at the very end) End – ML is used at the very end of the automated process If ML is used at multiple places in the automated process, this field can contain multiple values of Start, Middle or End
Industry	(Free text)	The industry name if there is a real use case being analysed by the research paper. Can be empty
Anomaly Detection (AD)	(Empty), Theoretical, Referenced, Real	Anomaly Detection is used to find data points or patterns that differ considerably from previous data points or patterns in order to flag them as irregular or potentially intrusive [171] (Empty) – AD is not mentioned in the use case or if there is no use case Theoretical – AD is described being used in an IA context, but it is theoretical in nature as opposed to a real use case Referenced – AD is described in a use case written by other authors, but is not studied in the paper directly Real – AD is directly used in the studied IA use case If the paper has multiple instances of AD, for example if it has a Real case and a Theoretical case, only the Real case is marked. The precedence is Real > Referenced > Theoretical
Binary Classification	(Empty), Theoretical, Referenced, Real	Binary Classification is the use of ML to classify an object into one of two labels [172], for example “cat” vs. “not cat” [A description of (Empty), Theoretical, Referenced and Real can be found in the “Anomaly Detection” row of this table, with “Anomaly Detection” substituted for the appropriate use case name]
Chatbot and Virtual Agents	(Empty), Theoretical, Referenced, Real	Chatbots and Virtual Agents allow for human to machine conversations to occur either through text or voice [173]
Computer Vision	(Empty), Theoretical, Referenced, Real	Computer Vision is an umbrella term referring to multiple ML techniques and use cases of detecting objects in images and video [174]
Document Classification	(Empty), Theoretical, Referenced, Real	Document classification is the determination of subject content under set of specific headings [175], for example, “Invoice” vs. “Purchase Order” vs. “Shipping Notice”
Email Classification	(Empty), Theoretical, Referenced, Real	Email Classification is used to determine the category of an email based on its contents, such as sender, subject, body contents, attachments, etc. [176]
Facial Recognition	(Empty), Theoretical, Referenced, Real	Facial Recognition is used to detect the presence of human faces in images and video [177]
Forecasting	(Empty), Theoretical, Referenced, Real	Forecasting is used to predict new data based on historical data [178]
Multi-class Classification	(Empty), Theoretical, Referenced, Real	Multi-class Classification is when a classifier must classify between a set of more than 2 disjoint labels [179]
Named Entity Recognition (NER)	(Empty), Theoretical, Referenced, Real	Named Entity Recognition is a form of Natural Language Processing, where entities are explicitly extracted from unstructured text [180], for example people’s names, numbers, stock codes etc.
Natural Language Processing (NLP)	(Empty), Theoretical, Referenced, Real	Natural Language Processing aims to achieve understanding of human language by the use of computers [181]

Object Recognition	(Empty), Theoretical, Referenced, Real	Object Recognition is used to recognize semantic objects belonging to a certain class in images or video [182]
Optical Character Recognition (OCR)	(Empty), Theoretical, Referenced, Real	OCR is when text from documents or images is converted into a machine-readable format [183]
Risk Management	(Empty), Theoretical, Referenced, Real	Risk Mitigation is the use of ML to model and reduce risk for a particular business application [26]
Sentiment Analysis	(Empty), Theoretical, Referenced, Real	Sentiment Analysis is used to determine the opinions of people about a specific topic, typically written in text [184]
Translation	(Empty), Theoretical, Referenced, Real	Translation is a form of NLP, which aims to find a target language sentence that best represents a source language sentence [185]

D. Research Question 2 (ML Algorithms and Technologies) Coding Scheme

Name	Allowed Values	Description
RQ2 Relevance	0, 1 or 2	0 – The article does not contain any ML algorithms nor technologies 1 – The article describes ML algorithms or technologies theoretically, or through references 2 – The article directly examines or implements at least one of the ML algorithms, or uses a specific ML technology in an IA solution
RQ2 Industry	(Free text)	The industry name if there is a real ML algorithm or technology being discussed by the research paper. Can be empty
Machine Learning Algorithms	(Empty) to one or multiple values from Table 4	(Empty) – No mention of any ML algorithms is present in the research article If not empty, one or more values taken from Table 4
Machine Learning Technologies	(Free text)	Free text list of any specific ML technologies mentioned in use for the IA solution. Can be empty

E. Research Question 3 (IA Risks) Coding Scheme

Name	Allowed Values	Description
RQ3 Relevance	0, 1 or 2	0 – The article does not mention any risks of using IA 1 – The article mentions IA risks without elaboration or detailed discussion 2 – The article discusses one or more IA risks with at least two sentences
RQ3 Industry	(Free text)	The industry name if there are IA risks identified in the research paper. Can be empty
Risks	(Free text)	Free text heading used to describe any risks mentioned due to the implementation of an IA solution. Can be empty

F. Research Question 4 (IA Risk Mitigation Techniques) Coding Scheme

Name	Allowed Values	Description
RQ4 Relevance	0, 1 or 2	0 – The article does not mention any risk mitigation techniques for IA 1 – The article mentions IA risk mitigation techniques without elaboration or detailed discussion, or if a measure is used that has the effect of reducing risk without explicitly labelling as a risk mitigation method 2 – The article discusses one or more IA risk mitigation techniques with at least two sentences
RQ4 Industry	(Free text)	The industry name if there are IA risk mitigation techniques identified in the research paper. Can be empty
Risk Mitigation Techniques	(Free text)	Free text heading of any risk mitigation techniques used to manage risk during the implementation of an IA solution. Can be empty

G. Coded Data

G.1 Overall Relevance, Publication Year and Publication Type

Reference	Overall Relevance	Publication Year	Publication Type
[10]	3	2018	Journal Article
[186]	4	2021	Conference Paper
[86]	1	2021	Journal Article
[187]	0	2020	Report
[30]	2	2021	Journal Article
[83]	4	2018	Journal Article
[188]	0	2020	Conference Paper
[25]	5	2020	Conference Paper
[80]	6	2018	Conference Paper
[17]	5	2020	Journal Article
[189]	1	2021	Book
[190]	0	2021	Journal Article
[126]	3	2020	Journal Article
[191]	2	2021	Conference Paper
[138]	4	2020	Journal Article
[192]	3	2021	Journal Article
[193]	1	2021	Journal Article
[194]	0	2019	Journal Article
[14]	3	2021	Conference Paper
[195]	0	2019	Journal Article
[196]	3	2021	Conference Paper
[84]	1	2020	Journal Article
[197]	2	2021	Journal Article
[198]	0	2019	Journal Article
[199]	0	2021	Journal Article

[200]	0	2021	Journal Article
[24]	1	2021	Journal Article
[201]	2	2021	Journal Article
[202]	0	2021	Journal Article
[203]	0	2020	Book
[125]	4	2019	Conference Paper
[204]	1	2020	Journal Article
[205]	3	2020	Conference Paper
[101]	2	2020	Book Section
[111]	2	2021	Journal Article
[114]	5	2021	Conference Paper
[206]	0	2020	Conference Paper
[207]	0	2019	Book Section
[208]	0	2018	Journal Article
[73]	1	2021	Journal Article
[91]	2	2021	Journal Article
[29]	3	2021	Journal Article
[209]	3	2020	Report
[210]	4	2019	Conference Paper
[95]	4	2021	Journal Article
[211]	4	2021	Journal Article
[88]	1	2021	Journal Article
[72]	5	2020	Journal Article
[212]	0	2021	Journal Article
[213]	1	2021	Book Section
[87]	2	2020	Journal Article
[112]	2	2019	Conference Paper
[214]	4	2021	Conference Paper
[215]	3	2021	Conference Paper
[216]	2	2021	Journal Article
[85]	6	2018	Journal Article
[217]	0	2020	Journal Article
[132]	4	2019	Conference Paper
[218]	1	2021	Conference Paper
[7]	1	2020	Journal Article
[22]	0	2021	Journal Article
[219]	3	2020	Conference Paper
[68]	6	2021	Journal Article
[11]	2	2020	Journal Article
[220]	0	2021	Journal Article
[221]	4	2021	Journal Article
[222]	0	2021	Journal Article
[223]	0	2020	Book Section

[74]	6	2021	Journal Article
[224]	0	2020	Conference Paper
[225]	0	2021	Journal Article
[81]	1	2020	Journal Article
[76]	0	2020	Journal Article
[226]	4	2018	Journal Article
[227]	1	2020	Journal Article
[100]	4	2019	Journal Article
[75]	4	2021	Journal Article

G.2 RQ1 Coded Data

The following column names have been abbreviated for ease of presentation in the table below:

- Ref: Reference number from “References”
- Rel: Relevance Score
- DC: Document Classification
- AD: Anomaly Detection
- F: Forecasting
- RM: Risk Management
- CB: Chatbots and Virtual Agents
- NLP: Natural Language Programming
- SA: Sentiment Analysis
- FR: Facial Recognition
- BC: Binary Classification
- EC: Email Classification
- MCC: Multi-class Classification
- OR: Object Recognition
- CV: Computer Vision
- T: Translation

Ref	Rel	Industry	Position	OCR	DC	NER	AD	F	RM	CB	NLP	SA	FR	BC	EC	OR	MCC	CV	T
[10]	2			3		3				3									
[186]	2	Healthcare	Start	3		3				3	3							1	
[86]	0																		
[187]	0																		
[30]	1			2	1	2					1	1			1			1	
[83]	0																		
[188]	0																		
[25]	1								1	1									

[80]	1	Insurance			1													
[17]	1							1			2			1			1	2
[189]	1							1						1			1	1
[190]	0																	
[126]	1	Accounting/Auditing			1	1	2				1	2						
[191]	1				2				2		2	2				2		
[138]	2	Logistics									3			3				
[192]	2	Insurance												3				
[193]	1				1						1							
[194]	0																	
[14]	0																	
[195]	0																	
[196]	2	HR	Start							3	3				3	1		1
[84]	0																	
[197]	2	Accounting/Auditing			3													
[198]	0																	
[199]	0																	
[200]	0																	
[24]	1									2					2			
[201]	0																	
[202]	0																	
[203]	0	Education																
[125]	2	Education												3				
[204]	1	Education									2							
[205]	2				3						3			3				3
[101]	0																	
[111]	0																	
[114]	1								1									
[206]	0																	
[207]	0																	
[208]	0																	

[73]	0																		
[91]	0																		
[29]	1					2		2		2	2	2				2		2	
[209]	2	Insurance	Start, Middle	3						3							3		
[210]	2	HR												3					
[95]	2	Finance					1			3									
[211]	2	Utilities	Middle				3												
[88]	0																		
[72]	2	Finance	Start	3		3	1	1	1	1									
[212]	0																		
[213]	1			2	2														
[87]	0																		
[112]	0	Finance																	
[214]	2	HR		3	3	3													
[215]	1	Insurance	Middle				1												
[216]	1	Healthcare				1												1	
[85]	2	Education								3									
[217]	0																		
[132]	2	IT	Start	3						3							3		
[218]	1			2	2	2				1	1							1	
[7]	1			1							1								
[22]	0																		
[219]	1	IT								1							1		
[68]	2	Healthcare		3	3	3								3					
[11]	1									2	2							2	
[220]	0																		
[221]	2	Utilities	Start					3						3					
[222]	0																		
[223]	0																		
[74]	2			3			3	3	1	1	3	3							
[224]	0																		

[225]	0																		
[81]	0																		
[76]	0																		
[226]	2	Finance		3	3	3													
[227]	1	Accounting/Auditing	Middle			2													
[100]	1	Accounting/Auditing							2	2								1	
[75]	0																		

G.3 RQ2 Coded Data

Ref	Rel	Industry	Machine Learning Algorithms	Machine Learning Technologies
[10]	1			ABBYY
[186]	2	Healthcare		MS Bot Framework
[86]	0			
[187]	0			
[30]	0			
[83]	0			
[188]	0			
[25]	0			
[80]	1	Insurance	Decision Trees	Python, Scikit-learn
[17]	0			
[189]	0			
[190]	0			
[126]	0			
[191]	1		ANN, Deep Learning	
[138]	0			
[192]	1	Insurance	SVM	
[193]	0			
[194]	0			
[14]	0			

[195]	0			
[196]	1	HR	CNN	Python, Tensorflow, Rasa
[84]	0			
[197]	0			
[198]	0			
[199]	0			
[200]	0			
[24]	0			
[201]	0			
[202]	0			
[203]	0			
[125]	1	Education		Azure Face API
[204]	0			
[205]	1		KNN	Jieba
[101]	0			
[111]	0			
[114]	1			Scikit-learn, Tensorflow, Pytorch
[206]	0			
[207]	0			
[208]	0			
[73]	0			
[91]	0			
[29]	1		SVM, Genetic Programming, Deep Learning, Logistic Regression, Random Forests, ANN	
[209]	1	Insurance	Deep Learning	Kommunicate, Google Dialogflow
[210]	2	HR	Random Forests	
[95]	0			
[211]	2	Utilities	Linear Regression, Random Forests	Python
[88]	0			
[72]	1	Finance		Orbograph OrboAnywhere
[212]	0			

[213]	0			
[87]	0			
[112]	0			
[214]	2	HR	LSTM, Naïve Bayes	
[215]	1	Insurance	LDA	
[216]	1	Healthcare	CNN	
[85]	2	Education		Watson Assistant
[217]	0			
[132]	1	IT	Random Forests, Gradient Boosting, LSTM	XGBoost
[218]	0			
[7]	0			
[22]	0			
[219]	1	IT	Random Forests	
[68]	2	Healthcare	ANN, Gradient Boosting	Google Vision, Spacy, Keras, Python, XGboost
[11]	1		Deep Learning	IBM Watson, Azure Machine Learning
[220]	0			
[221]	2	Utilities	Logistic Regression, Decision Trees, Random Forests, Gradient Boosting	Azure Databricks Machine Learning
[222]	0			
[223]	0			
[74]	0			
[224]	0			
[225]	0			
[81]	0			
[76]	0			
[226]	2	Finance	Deep Learning, Random Forest, Linear Regression, SVM, RNN	LibSVM
[227]	0			
[100]	1	Accounting/Auditing		Watson Assistant
[75]	0			

G.4 RQ3 Coded Data

Ref	Rel	Industry	Risks
[10]	0		
[186]	0		
[86]	1	Government	Loss of Job Meaning, Missed Servicing Opportunities
[187]	0		
[30]	1		Data Privacy, Data Quality
[83]	2	Logistics	Assignment of Liability, Performance Agreement Breaches
[188]	0		
[25]	2		Control Flow Drifts, Data Drift, Financial Loss, Low ROI, Mistrust in Model Predictions, Reputation Loss
[80]	2	Insurance	Financial Loss
[17]	2		Assignment of Liability, Financial Loss, Reputation Loss, Worker Deskilling
[189]	0		
[190]	0		
[126]	1	Accounting/Auditing	Data Quality
[191]	0		
[138]	0		
[192]	0		
[193]	0		
[194]	0		
[14]	2		Adversarial Attacks, Compliance, Data Bias, Data Drift, Data Privacy, Loss of Job Security, Mistrust in Management, Transfer Learning Bias
[195]	0		
[196]	0		
[84]	1		Assignment of Liability, Data Bias, Prediction Accountability
[197]	0		
[198]	0		
[199]	0		
[200]	0		
[24]	0		
[201]	0		

[202]	0		
[203]	0		
[125]	1	Education	Data Quality
[204]	0		
[205]	0		
[101]	1		Low Predictive Performance
[111]	1	Forestry	Reduced Understanding of Business Logic
[114]	2		Performance Degradation
[206]	0		
[207]	0		
[208]	0		
[73]	1	Insurance	Ethics
[91]	2		Cognitive Work Overload, Loss of Job Meaning, Loss of Job Security
[29]	0		
[209]	0		
[210]	0		
[95]	2	Finance	Difficult Error Detection, Loss of Job Security, Mistrust in Model Predictions, Reduced Work Preparedness, Reduced Understanding of Business Logic
[211]	0		
[88]	1	Sports	Reputation Loss
[72]	1	Finance	Regulatory, Worker Deskilling
[212]	0		
[213]	0		
[87]	2		Data Bias, Missed Servicing Opportunities, Worker Deskilling
[112]	1	Finance	Data Bias, Time Lag Effects
[214]	0		
[215]	0		
[216]	0		
[85]	1	Education	Attract Competitive Response, Data Quality, Loss of Job Security, Unmeasurable ROI
[217]	0		
[132]	0		

[218]	0		
[7]	0		
[22]	0		
[219]	0		
[68]	1	Healthcare	Compliance
[11]	0		
[220]	0		
[221]	0		
[222]	0		
[223]	0		
[74]	2	Finance	Data Privacy, Difficult Error Detection, Ethics, Financial Loss, Prediction Accountability, Reputation Loss
[224]	0		
[225]	0		
[81]	1		Conflicts of Interest, Data Privacy, Information Asymmetry, Loss of Control
[76]	0		
[226]	0		
[227]	0		
[100]	1	Accounting/Auditing	Worker Deskilling
[75]	2		Data Bias, Difficult Error Detection, Employee Turnover, Departmental Resistance

G.5 RQ4 Coded Data

Ref	Rel	Industry	Risk Mitigation
[10]	0		
[186]	0		
[86]	0		
[187]	0		
[30]	0		
[83]	2	Logistics	AI Liability Terms in Contracts, Contract Renegotiation
[188]	0		

[25]	2		HITL, Explainable AI, Staged Deployments
[80]	2	Insurance	Minimize False Positives
[17]	2		HITL
[189]	0		
[190]	0		
[126]	1	Accounting/Auditing	Governance, Explainable AI
[191]	0		
[138]	2	Logistics	HITL, Monitor Data, Monitor Models
[192]	0		
[193]	0		
[194]	0		
[14]	1		Explainable AI
[195]	0		
[196]	0		
[84]	0		
[197]	0		
[198]	0		
[199]	0		
[200]	0		
[24]	0		
[201]	2		HITL
[202]	0		
[203]	0		
[125]	0		
[204]	0		
[205]	0		
[101]	1		HITL
[111]	1	Forestry	Governance
[114]	1		HITL, Monitor Models
[206]	0		
[207]	0		

[208]	0		
[73]	0		
[91]	0		
[29]	1		Self-learning
[209]	0		
[210]	0		
[95]	0		
[211]	0		
[88]	0		
[72]	1	Finance	Thresholding
[212]	0		
[213]	0		
[87]	0		
[112]	1	Finance	HITL, Process Runtime Controls
[214]	0		
[215]	1	Insurance	Thresholding
[216]	0		
[85]	1	Education	Avoid Self-learning
[217]	0		
[132]	1	IT	Thresholding
[218]	0		
[7]	0		
[22]	0		
[219]	1	IT	HITL
[68]	1	Healthcare	Random Sampling
[11]	0		
[220]	0		
[221]	0		
[222]	0		
[223]	0		
[74]	2	Finance	HITL

[224]	0		
[225]	0		
[81]	0		
[76]	0		
[226]	0		
[227]	0		
[100]	1	Accounting/Auditing	HITL
[75]	2		Understand Employee Sentiment

H. Number of Research Papers Addressing Specific Combinations of RQs

An “X” entry means that the research question had a zero-valued Relevance Score whereas an “O” means that the Relevance Score was either a 1 or a 2.

RQ1	RQ2	RQ3	RQ4	Number of Research Papers
X	X	X	X	22
X	X	X	O	1
X	X	O	X	7
X	X	O	O	6
X	O	X	X	0
X	O	X	O	0
X	O	O	X	0
X	O	O	O	0
O	X	X	X	9
O	X	X	O	1
O	X	O	X	2
O	X	O	O	4
O	O	X	X	14
O	O	X	O	4
O	O	O	X	1
O	O	O	O	6

I. Risk Levels of Real IA Use Cases (Relevance Score of 2)

Reference	Risk Level	Justification
[10]	L	A chatbot is used in HR department to enrol people for training courses. The risk level is low because there is still a manual process to enrol, and the consequences of a failed or incorrect enrolment is easily corrected
[10]	H	Entity extraction is used on documents to extract key data. The extracted key data is input into the company's (FMCG) ERP system. No other details about whether that data is later checked by a human, so the assumption is that the risk level is high
[186]	L	A chatbot is used in a healthcare provider to better enable internal communications, such sending budget reports, booking and checking where meetings are, finding contract documentation, salaries, asking for the status of sick leave etc. As an internal tool, the risk is low
[138]	H	A ship brokering company uses data of ship behaviour, location and cargo to determine in real time the prices of commodities for potential arbitrage. The risk is high as predictions are used in profit making activities
[192]	H	A Title Insurance company expects documents to be divided into six types during processing. It is unknown what kind of processing occurs after the documents are classified, so it is assumed that it is high risk
[196]	N/A	In this case, the IA system was implemented, warranting the relevance score of 2, but only tested, not put into production yet
[197]	L	Invoice documents are scanned and OCR'ed at an accounting firm. OCR'ed documents are corrected by staff afterwards, making the risk level low
[125]	L	Facial Recognition and eye tracking is used in an online education environment. The use case is low risk as there is no financial repercussions of incorrect predictions

[205]	L	Historical documents are classified by AI, and a human is still retained as part of the process, resulting in a low-risk level
[209]	N/A	There are two ML portions. First is a customer-facing chatbot for people to discuss insurance claim details with their insurance company. Second is a module that classifies car damage based on a photograph into low, medium and high damage. The system was built but does not seem to have been used in real life
[210]	N/A	Random Forests are used to predict employee satisfaction. This was only tested in a laboratory setting
[95]	L	A financial services firm automates the investment service and information requests that are submitted online, replacing manual data entry with human validation. As humans remain a part of the process to validate the ML results, it is low risk
[211]	L	AI is used to detect abnormal power meter readings for an electrical utility and provide a substitute reading. All readings are validated by a process manager. This is low risk as a human approves the AI prediction
[72]	H	The “payee” field of a cheque is read using ML. This is compared against a list of valid payees in a database. If the confidence score is low or if the payee is not found in the database, the cheque is sent to a human operator for review. Not all predictions are validated
[214]	N/A	ML is used to assess the quality of job applicants. The system was developed but only tested in a laboratory setting and not operationalized in real life
[85]	L	A chatbot is used in a university to help respond to student queries. As the interaction is used to retrieve information, the risk is low
[218]	L	Email support ticket requests at a plant are classified and redirected to the appropriate person for processing. If it is redirected incorrectly, the person can route the request to the correct person, leading to low risk
[68]	L	Data is extracted from medical prescription images to speed up the digitization process. Digitized prescriptions still undergo two rounds of human screening, making this low risk
[221]	L	A new process is developed at a power utility to forecast power outages and pre-emptively SMS people in the affected region, to reduce the number of complaint and service calls. This is low risk
[74]	L	AI is used to perform adverse media screening as part of the KYC and AML lifecycle at a bank. It flags articles that were found for human review. Given human involvement, this is low risk
[226]	H	A debt collector company must digitize, classify and extract key information from legal documents received by post. No information is provided about human validation

J. Implied Risk to Risk Mitigation Mappings in the Literature

Reference	Risk	Risk Mitigation
[83]	Assignment of Liability	AI Liability Terms in Contracts
[83]	Performance Agreement Breaches	Contract Renegotiation
[25]	Financial Loss	Staged Deployments
[25]	Reputation Loss	Staged Deployments
[25]	Mistrust in Model Predictions	Explainable AI, HITL
[80]	Financial Loss	Minimize False Positives
[17]	Financial Loss	HITL
[17]	Reputation Loss	HITL
[17]	Assignment of Liability	HITL
[126]	Data Quality	Governance
[14]	Compliance	Explainable AI

[101]	Low Predictive Performance	HITL
[111]	Reduced Understanding of Business Logic	Governance
[114]	Performance Degradation	HITL, Monitor Models
[72]	Regulatory	Thresholding
[72]	Worker Deskilling	Thresholding
[112]	Time Lag Effects	HITL
[11]	Compliance	Random Sampling
[74]	Reputation Loss	HITL
[74]	Financial Loss	HITL
[74]	Prediction Accountability	HITL
[74]	Difficult Error Detection	HITL
[74]	Ethics	HITL
[100]	Worker Deskilling	HITL
[75]	Employee Turnover	Understand Employee Sentiment

K. Risk to Risk Mitigation Mappings

Risk	Direct Risk Mitigation Mapping from SLR Papers	Risk Mitigation Mapping from Discussion
Adversarial Attacks	N/A	Governance
Assignment of Liability	Measurable ML Attributes in Contracts, HITL	Explainable AI
Attract Competitive Response	N/A	Governance
Cognitive Work Overload	N/A	Governance
Compliance	Explainable AI, Random Sampling	Governance, Process Runtime Controls
Conflicts of Interest	N/A	Measurable ML Attributes in Contracts
Control Flow Drifts	N/A	Governance
Data Bias	N/A	Governance, Monitor Data
Data Drift	N/A	Governance, Monitor Data
Data Privacy	N/A	Governance
Data Quality	Governance	Governance, Monitor Data
Departmental Resistance	N/A	N/A
Difficult Error Detection	N/A	HITL, Monitor Data, Monitor Models, Thresholding, Random Sampling
Ethics	N/A	Governance, Monitor Data, Monitor Models
Employee Turnover	Understand Employee Sentiment	Governance
Financial Loss	HITL, Minimize False Positives, Staged Deployments	Contract Renegotiation, Measurable ML Attributes in Contracts, Process Runtime Controls
Information Asymmetry	N/A	Governance
Loss of Control	N/A	Measurable ML Attributes in Contracts
Loss of Job Meaning	N/A	N/A
Loss of Job Security	N/A	N/A
Low Predictive Performance	HITL	Explainable AI, Monitor Data, Monitor Models, Thresholding
Low ROI	N/A	Governance
Missed Servicing Opportunities	N/A	Random Sampling

Mistrust in Management	N/A	N/A
Mistrust in Model Predictions	Explainable AI, HITL	Minimize False Positives, Random Sampling, Thresholding
Performance Agreement Breaches	Contract Renegotiation	Measurable ML Attributes in Contracts, Staged Deployments
Performance Degradation	HITL, Monitor Models	Monitor Data, Random Sampling, Self-learning, Thresholding
Prediction Accountability	HITL	Governance
Reduced Understanding of Business Logic	Governance	Explainable AI, HITL, Random Sampling, Thresholding
Reduced Work Preparedness	N/A	N/A
Regulatory	Thresholding	Explainable AI, Governance, HITL, Process Runtime Controls
Reputation Loss	Staged Deployments, HITL	Explainable AI, Minimize False Positives, Monitor Data, Monitor Models, Random Sampling, Thresholding
Time Lag Effects	HITL	Random Sampling
Transfer Learning Bias	N/A	N/A
Unmeasurable ROI	N/A	Governance
Worker Deskilling	HITL, Thresholding	Random Sampling

L. IA Risk Register

Category	Name	Description	Mitigations
Socio-organizational / Environmental	Compliance	IA may complicate the compliance terms of existing regulations. Third-party ML vendors must also be compliant with necessary regulations	Governance (surface and follow all relevant company policies), Process Runtime Controls
Socio-organizational / Environmental	Ethics	Biases in the automated processing may lead to ethical concerns	Governance (follow or create ethical guidelines for IA), Monitor Data. Monitor Models
Socio-organizational / Environmental	Regulatory	Government regulations may appear in the future, governing how and when algorithmic decisions can be used	Explainable AI, Governance (surface and follow all relevant company policies), HITL, Process Runtime Controls
Socio-organizational / Enterprise	Departmental Resistance	Managers may worry that their headcounts or budgets will get frozen or reduced due to IA, leading to non-cooperation or sabotage	Provide IA training and involve managers in the process. Provide statistics on real-life headcount reduction data. Explicitly define how time saved by employees will be filled with value-adding tasks
Socio-organizational / Enterprise	Employee Turnover	Employees may quit the organization in large amounts because of IA	Governance (measure turnover intent over time), Understand Employee Sentiment
Socio-organizational / Enterprise	Financial Loss	Incorrect IA work may lead to financial loss to the firm	Contract Renegotiation, HITL, Measurable ML Attributes in Contracts, Minimize False Positives, Process Runtime Controls, Staged Deployments
Socio-organizational / Enterprise	Loss of Control	If the firm relies on vendors to develop or host the ML predictions, there is a risk that their service will go offline or cease operating	Measurable ML Attributes in Contracts
Socio-organizational / Employee	Cognitive Work Overload	Removing simple cognitive work may leave only difficult cognitive work, leading to cognitive overload and increased job stress	Governance (measure employee work overload over time)
Socio-organizational / Employee	Loss of Job Meaning	IA may be automating work that is meaningful to employees, leaving them less satisfied with their jobs	Measure meaningful work through WAMI over time. For affected staff, provide more autonomy, more significant work or increase corporate social responsibility more broadly at the firm-level
Socio-organizational / Employee	Loss of Job Security	Knowledge-based automation can lead to a much larger group of employees to worry about their jobs, increasing their stress levels and impairing their health	Measure Job Security Index and Job Security Satisfaction scores over time. Make the KPIs of IA known to all. Commit to not cutting jobs.

			Prepare training plans and new job role descriptions for those affected
Socio-organizational / Employee	Mistrust in Management	A push towards IA may lead to mistrust being formed between workers and management	Implement IA based on genuine demand, communicate openly with affected employees and address their feedback. Measure organizational trust over time
Socio-organizational / Employee	Mistrust in Model Predictions	A lack of trust in model predictions may lead employees to actively resist or sabotage IA efforts	Explainable AI, HITL, Minimize False Positives, Random Sampling, Thresholding
Socio-organizational / Employee	Prediction Accountability	Which employee(s) should be held responsible if a prediction or business outcome is incorrect?	HITL, Governance (define and document who is responsible for incorrect predictions)
Socio-organizational / Employee	Reduced Work Preparedness	When IA is in place, employees spend less time looking at work cases meaning that fewer details are known about a case if it needs to be manually worked on	Make access to data processed by IA easier, through dashboards and reports
Socio-organizational / Employee	Worker Deskilling	Worker's skill in performing their knowledge tasks is reduced due to automation	HITL, Random Sampling, Thresholding
Socio-organizational / Third-Party	Assignment of Liability	When multiple companies are involved in the development and operations of ML predictions in IA, liability becomes unclear between the firms if something goes wrong	Explainable AI, HITL, Measurable ML Attributes in Contracts
Socio-organizational / Third-Party	Attract Competitive Response	Publicly investing in IA can trigger responses from competition, either encouraging them to pursue IA themselves or to decry your use of IA	Governance (penalize processes that are visible to competitors and customer during process selection)
Socio-organizational / Third-Party	Conflicts of Interest	AI Vendors may be incentivized towards rent seeking behaviour	Measurable ML Attributes in Contracts
Socio-organizational / Third-Party	Missed Servicing Opportunities	When narrow AI is used to interact with customers, there is no flexibility to discover additional ways in which the customer can be serviced	Random Sampling
Socio-organizational / Third-Party	Performance Agreement Breaches	Existing performance agreements or SLAs may need to be renegotiated after implementing IA	Contract Renegotiation, Measurable ML Attributes in Contracts, Staged Deployments
Socio-organizational / Third-Party	Reputation Loss	Incorrect IA work may lead to reputational loss to the firm	Explainable AI, HITL, Monitor Data, Minimize False Positives, Monitor Models, Random Sampling, Staged Deployments, Thresholding
Operational / Project	Low Predictive Performance	IA predictions may be less accurate or reliable than human predictions, resulting in worse outcomes after automation	Explainable AI, HITL, Monitor Data, Monitor Models, Thresholding

Operational / Project	Low ROI	The ROI of IA may be unattractive compared to RPA due to the additional needs of constant monitoring, retraining of models, and the costs of data scientists	Governance (evaluate processes for automation based on value-based analysis)
Operational / Project	Unmeasurable ROI	The ROI of IA may not be measurable due to being unable to quantify the value of knowledge or decision work	Governance (evaluate processes for automation based on value-based analysis)
Operational / Process	Control Flow Drifts	Changing business process logic and pathing in the control flow may necessitate rebuilding ML models	Governance (penalize processes that are apt to change during process selection)
Operational / Process	Difficult Error Detection	The use of automated ML decisions may make detection of errors in the business process more complicated	HITL, Monitor Data, Monitor Models, Thresholding, Random Sampling
Operational / Process	Reduced Understanding of Business Logic	The overall knowledge about the business process may reduce over time if business decision making is automated	Explainable AI, Governance (require documentation of process steps and explaining why they are performed), HITL, Random Sampling, Thresholding
Operational / Process	Time Lag Effects	There is a time gap between when an error is detected and when the error actually occurred. Automatically processed work during this time gap needs to be reviewed to see if it needs correction or re-doing	HITL, Random Sampling
Operational / ML Model	Adversarial Attacks	ML algorithms used in IA are subject to adversarial attacks, which would propagate in unwanted automated work being processed	Governance (use existing IT security and data security policies)
Operational / ML Model	Performance Degradation	Model predictive performance are known to reduce over time unless actively managed	HITL, Monitor Data, Monitor Models, Random Sampling, Self-learning, Thresholding
Operational / ML Model	Transfer Learning Bias	If transfer learning is used to develop the model, the base model may have hidden biases, with no way to fix it	Develop models in-house before trying transfer learning
Operational / Data	Data Bias	The data may have biases and perform poorly on real life data	Governance (require bias checking of data pre-implementation), Monitor Data
Operational / Data	Data Drift	The underlying nature or distributions of the data may change over time	Governance (establish baseline data distributions), Monitor Data
Operational / Data	Data Privacy	Sending sensitive data to third parties for use or model development may lead to data leaks	Governance (use existing data privacy and data security policies),
Operational / Data	Data Quality	The data quality may lead to poorly performing models	Governance (define data quality standards), Monitor Data

References

- [1] K. Roose, "The Robots Are Coming for Phil in Accounting," *The New York Times*, Mar. 06, 2021. Accessed: Feb. 26, 2022. [Online]. Available: <https://www.nytimes.com/2021/03/06/business/the-robots-are-coming-for-phil-in-accounting.html>
- [2] J. R. Bright, *Automation and management*. Boston: Division of Research, Graduate School of Business Administration, Harvard University, 1958.
- [3] J. Frohm, V. Lindström, M. Winroth, and J. Stahre, "Levels of Automation in Manufacturing," *Ergonomia - International Journal of Ergonomics and Human Factors*, vol. 30, pp. 181–207, Jan. 2008.
- [4] M. C. L. and L. P. Willcocks, "A New Approach to Automating Services," *MIT Sloan Management Review*. <https://sloanreview.mit.edu/article/a-new-approach-to-automating-services/> (accessed Feb. 26, 2022).
- [5] S. Dey and A. Das, "Robotic process automation: assessment of the technology for transformation of business processes," *International Journal of Business Process Integration and Management*, vol. 9, no. 3, pp. 220–230, Jan. 2019, doi: 10.1504/IJBPIIM.2019.100927.
- [6] A. Asquith and G. Horsman, "Let the robots do it! – Taking a look at Robotic Process Automation and its potential application in digital forensics," *Forensic Science International: Reports*, vol. 1, p. 100007, Nov. 2019, doi: 10.1016/j.fsir.2019.100007.
- [7] J. Siderska, "Robotic Process Automation — a driver of digital transformation?," *Engineering Management in Production and Services*, vol. 12, no. 2, pp. 21–31, May 2020, doi: 10.2478/emj-2020-0009.
- [8] L. P. Willcocks, M. Lacity, and A. Craig, "Robotizing global financial shared services at Royal DSM," *Journal of Financial Transformation*, Nov. 27, 2017. <https://www.capco.com/Insights/Capco-Institute> (accessed Mar. 01, 2022).
- [9] "Announcing RPA, enhanced security, no-code virtual agents, and more for Microsoft Power Platform," *Microsoft Dynamics 365 Blog*, Nov. 04, 2019. <https://cloudblogs.microsoft.com/dynamics365/bdm/2019/11/04/announcing-rpa-enhanced-security-no-code-virtual-agents-and-more-for-microsoft-power-platform/> (accessed Feb. 26, 2022).

- [10] S. Anagnoste, "Robotic Automation Process - The operating system for the digital enterprise," *PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON BUSINESS EXCELLENCE*, vol. 12, no. 1. SCIENDO, BOGUMILA ZUGA 32A, WARSAW, MAZOVIA, POLAND, pp. 54–69, May 2018. doi: 10.2478/picbe-2018-0007.
- [11] R. Syed *et al.*, "Robotic Process Automation: Contemporary themes and challenges," *Computers in Industry*, vol. 115, p. 103162, 2020, doi: <https://doi.org/10.1016/j.compind.2019.103162>.
- [12] "Q&A: Everything you need to know about intelligent automation," *World Economic Forum*. <https://www.weforum.org/agenda/2021/09/what-is-intelligent-automation-how-help-us/> (accessed Mar. 01, 2022).
- [13] A. Yang, "Money for nothing: The case for universal basic income," *nydailynews.com*. <https://www.nydailynews.com/opinion/ny-oped-money-for-nothing-universal-basic-income-20180830-story.html> (accessed Mar. 26, 2022).
- [14] L.-V. Herm, C. Janiesch, H. A. Reijers, and F. Seubert, "From Symbolic RPA to Intelligent RPA: Challenges for Developing and Operating Intelligent Software Robots," in *Business Process Management*, Cham, 2021, pp. 289–305. doi: 10.1007/978-3-030-85469-0_19.
- [15] M. Schmitz, C. Stummer, and M. Gerke, "Smart Automation as Enabler of Digitalization? A Review of RPA/AI Potential and Barriers to Its Realization," in *Future Telco: Successful Positioning of Network Operators in the Digital Age*, P. Krüssel, Ed. Cham: Springer International Publishing, 2019, pp. 349–358. doi: 10.1007/978-3-319-77724-5_31.
- [16] "IEEE Guide for Taxonomy for Intelligent Process Automation Product Features and Functionality," *IEEE Std 2755.1-2019*, pp. 1–53, Jul. 2019, doi: 10.1109/IEEESTD.2019.8764094.
- [17] C. Coombs, D. Hislop, S. K. Taneva, and S. Barnard, "The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review," *The Journal of Strategic Information Systems*, vol. 29, no. 4, p. 101600, 2020, doi: <https://doi.org/10.1016/j.jsis.2020.101600>.
- [18] U. Inc, "Introducing UiPath AI Fabric Now In The Cloud | UiPath." <https://www.uipath.com/blog/product-and-updates/uipath-ai-fabric-in-cloud-ga-release> (accessed Feb. 26, 2022).
- [19] "IQ Bot: Again a Winner," *Automation Anywhere*. <https://www.automationanywhere.com/company/blog/company-news/iq-bot-again-a-winner> (accessed Feb. 26, 2022).

- [20] B. Prism, "Blue Prism Shares Vision of the Future Workforce: Digital First, People Enriched." <https://www.prnewswire.com/news-releases/blue-prism-shares-vision-of-the-future-workforce-digital-first-people-enriched-301425204.html> (accessed Feb. 26, 2022).
- [21] "Gartner Forecasts Worldwide Hyperautomation-Enabling Software Market to Reach Nearly \$600 Billion by 2022," *Gartner*. <https://www.gartner.com/en/newsroom/press-releases/2021-04-28-gartner-forecasts-worldwide-hyperautomation-enabling-software-market-to-reach-nearly-600-billion-by-2022> (accessed Feb. 26, 2022).
- [22] J. Siderska, "The Adoption of Robotic Process Automation Technology to Ensure Business Processes during the COVID-19 Pandemic," *Sustainability*, vol. 13, no. 14, Art. no. 14, Jan. 2021, doi: 10.3390/su13148020.
- [23] "Defense.gov Transcript: DoD News Briefing - Secretary Rumsfeld and Ge...", *archive.ph*, Mar. 20, 2018. <http://archive.ph/8k6bU> (accessed Mar. 07, 2022).
- [24] M. Lacity and L. Willcocks, "Becoming Strategic with Intelligent Automation," *MIS QUARTERLY EXECUTIVE*, vol. 20, no. 2. INDIANA UNIV, OPER & DECISION TECHNOL DEPT, KELLEY SCH BUS, E 10 ST, BLOOMINGTON, IN 47405-1701 USA, pp. 169–182, Jun. 2021. doi: 10.17705/2msqe.00047.
- [25] T. Chakraborti *et al.*, "From Robotic Process Automation to Intelligent Process Automation: Emerging Trends," *arXiv:2007.13257 [cs]*, Jul. 2020, Accessed: Jan. 10, 2022. [Online]. Available: <http://arxiv.org/abs/2007.13257>
- [26] S. Aziz and M. M. Dowling, "AI and Machine Learning for Risk Management," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3201337, Jul. 2018. doi: 10.2139/ssrn.3201337.
- [27] M. Leo, S. Sharma, and K. Maddulety, "Machine Learning in Banking Risk Management: A Literature Review," *Risks*, vol. 7, no. 1, Art. no. 1, Mar. 2019, doi: 10.3390/risks7010029.
- [28] A. Gaur and M. Kumar, "A systematic approach to conducting review studies: An assessment of content analysis in 25years of IB research," *Journal of World Business*, vol. 53, no. 2, pp. 280–289, Feb. 2018, doi: 10.1016/j.jwb.2017.11.003.
- [29] K. K. H. Ng, C.-H. Chen, C. K. M. Lee, J. (Roger) Jiao, and Z.-X. Yang, "A systematic literature review on intelligent automation: Aligning concepts from theory, practice, and future perspectives," *Advanced Engineering Informatics*, vol. 47, p. 101246, 2021, doi: <https://doi.org/10.1016/j.aei.2021.101246>.

- [30] D. Baviskar, S. Ahirrao, V. Potdar, and K. Kotecha, "Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions," *IEEE Access*, vol. 9, pp. 72894–72936, 2021, doi: 10.1109/ACCESS.2021.3072900.
- [31] V. J. Duriau, R. K. Reger, and M. D. Pfarrer, "A Content Analysis of the Content Analysis Literature in Organization Studies: Research Themes, Data Sources, and Methodological Refinements," *Organizational Research Methods*, vol. 10, no. 1, pp. 5–34, Jan. 2007, doi: 10.1177/1094428106289252.
- [32] M. Williams and T. Moser, "The Art of Coding and Thematic Exploration in Qualitative Research," *undefined*, 2019, Accessed: Mar. 15, 2022. [Online]. Available: <https://www.semanticscholar.org/paper/The-Art-of-Coding-and-Thematic-Exploration-in-Williams-Moser/c0a0c26ac41cb8beb337834e6c1e2f35b91d071d>
- [33] T. O. Ayodele, *Types of Machine Learning Algorithms*. IntechOpen, 2010. doi: 10.5772/9385.
- [34] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, Mar. 2016, pp. 1310–1315.
- [35] V. K. Ayyadevara, *Pro Machine Learning Algorithms*. Berkeley, CA: Apress, 2018. doi: 10.1007/978-1-4842-3564-5.
- [36] C. Molnar, G. Casalicchio, and B. Bischl, "Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges," 2020, pp. 417–431. doi: 10.1007/978-3-030-65965-3_28.
- [37] Y. Lou, R. Caruana, and J. Gehrke, "Intelligible models for classification and regression," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, Aug. 2012, pp. 150–158. doi: 10.1145/2339530.2339556.
- [38] "Using Confidence Scores - Amazon Lex." <https://docs.aws.amazon.com/lex/latest/dg/confidence-scores.html> (accessed Feb. 26, 2022).
- [39] aahill, "Confidence score - question answering - Azure Cognitive Services." <https://docs.microsoft.com/en-us/azure/cognitive-services/language-service/question-answering/concepts/confidence-score> (accessed Feb. 26, 2022).
- [40] "Document AI | Google Cloud." <https://cloud.google.com/document-ai/docs/reference/rest/v1/Document> (accessed Feb. 26, 2022).

- [41] sanjeev3, "Capabilities and limitations of optical character recognition (OCR) - Computer Vision - Azure Cognitive Services." <https://docs.microsoft.com/en-us/legal/cognitive-services/computer-vision/ocr-characteristics-and-limitations> (accessed Feb. 26, 2022).
- [42] "1.16. Probability calibration," *scikit-learn*. <https://scikit-learn/stable/modules/calibration.html> (accessed Feb. 26, 2022).
- [43] "tf.keras.activations.softmax | TensorFlow Core v2.8.0," *TensorFlow*. https://www.tensorflow.org/api_docs/python/tf/keras/activations/softmax (accessed Feb. 26, 2022).
- [44] "Python API Reference — xgboost 1.5.2 documentation." https://xgboost.readthedocs.io/en/stable/python/python_api.html (accessed Feb. 26, 2022).
- [45] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine Learning Interpretability: A Survey on Methods and Metrics," *Electronics*, vol. 8, no. 8, Art. no. 8, Aug. 2019, doi: 10.3390/electronics8080832.
- [46] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy*, vol. 23, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/e23010018.
- [47] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat Mach Intell*, vol. 1, no. 5, Art. no. 5, May 2019, doi: 10.1038/s42256-019-0048-x.
- [48] C. Molnar, *Interpretable Machine Learning*. Accessed: Feb. 26, 2022. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/>
- [49] A. Deeks, "The Judicial Demand for Explainable Artificial Intelligence," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3440723, Aug. 2019. Accessed: Feb. 26, 2022. [Online]. Available: <https://papers.ssrn.com/abstract=3440723>
- [50] F. Doshi-Velez *et al.*, "Accountability of AI Under the Law: The Role of Explanation," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3064761, Nov. 2017. doi: 10.2139/ssrn.3064761.
- [51] "index | TIOBE - The Software Quality Company." <https://www.tiobe.com/tiobe-index/> (accessed Feb. 28, 2022).
- [52] S. Junyi, *jieba*. 2022. Accessed: Feb. 26, 2022. [Online]. Available: <https://github.com/fxsjy/jieba>

- [53] *Keras: Deep Learning for humans*. Keras, 2022. Accessed: Feb. 26, 2022. [Online]. Available: <https://github.com/keras-team/keras>
- [54] C.-J. Lin, *cjlin1/libsvm*. 2022. Accessed: Feb. 26, 2022. [Online]. Available: <https://github.com/cjlin1/libsvm>
- [55] *microsoft/botframework-sdk*. Microsoft, 2022. Accessed: Feb. 26, 2022. [Online]. Available: <https://github.com/microsoft/botframework-sdk>
- [56] *pytorch/pytorch*. pytorch, 2022. Accessed: Feb. 26, 2022. [Online]. Available: <https://github.com/pytorch/pytorch>
- [57] *scikit-learn/scikit-learn*. scikit-learn, 2022. Accessed: Feb. 26, 2022. [Online]. Available: <https://github.com/scikit-learn/scikit-learn>
- [58] *Rasa Open Source*. Rasa, 2022. Accessed: Feb. 26, 2022. [Online]. Available: <https://github.com/RasaHQ/rasa>
- [59] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, *spaCy: Industrial-strength Natural Language Processing in Python*. 2020. doi: 10.5281/zenodo.1212303.
- [60] M. Abadi *et al.*, *TensorFlow, Large-scale machine learning on heterogeneous systems*. 2015. doi: 10.5281/zenodo.4724125.
- [61] *eXtreme Gradient Boosting*. Distributed (Deep) Machine Learning Community, 2022. Accessed: Feb. 26, 2022. [Online]. Available: <https://github.com/dmlc/xgboost>
- [62] M. T. C. Ribeiro, *lime*. 2022. Accessed: Feb. 26, 2022. [Online]. Available: <https://github.com/marcotcr/lime>
- [63] S. Lundberg, *slundberg/shap*. 2022. Accessed: Feb. 26, 2022. [Online]. Available: <https://github.com/slundberg/shap>
- [64] “Lime - basic usage, two class case.” <https://marcotcr.github.io/lime/tutorials/Lime%20-%20basic%20usage%2C%20two%20class%20case.html> (accessed Feb. 26, 2022).
- [65] “API Reference,” *scikit-learn*. <https://scikit-learn/stable/modules/classes.html> (accessed Feb. 26, 2022).
- [66] M. T. C. Ribeiro, *lime*. 2022. Accessed: Feb. 28, 2022. [Online]. Available: <https://github.com/marcotcr/lime/blob/fd7eb2e6f760619c29fca0187c07b82157601b32/doc/notebooks/Tutorial%20-%20Image%20Classification%20Keras.ipynb>

- [67] M. T. C. Ribeiro, *lime*. 2022. Accessed: Feb. 28, 2022. [Online]. Available: <https://github.com/marcotcr/lime/blob/fd7eb2e6f760619c29fca0187c07b82157601b32/doc/notebooks/Tutorial%20-%20images%20-%20Pytorch.ipynb>
- [68] K. Soeny, G. Pandey, U. Gupta, A. Trivedi, M. Gupta, and G. Agarwal, “Attended robotic process automation of prescriptions’ digitization,” *Smart Health*, vol. 20, p. 100189, 2021, doi: <https://doi.org/10.1016/j.smhl.2021.100189>.
- [69] A. D. Selbst and J. Powles, “Meaningful information and the right to explanation,” *International Data Privacy Law*, vol. 7, no. 4, pp. 233–242, Nov. 2017, doi: 10.1093/idpl/ix022.
- [70] M. E. Kaminski, “The Right to Explanation, Explained,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3196985, Jun. 2018. doi: 10.2139/ssrn.3196985.
- [71] B. Goodman and S. Flaxman, “European Union regulations on algorithmic decision-making and a ‘right to explanation,’” *AIMag*, vol. 38, no. 3, pp. 50–57, Oct. 2017, doi: 10.1609/aimag.v38i3.2741.
- [72] P. Polak, C. Nelischer, H. Guo, and D. C. Robertson, “‘Intelligent’ finance and treasury management: what we can expect,” *AI & SOCIETY*, vol. 35, no. 3. SPRINGER, ONE NEW YORK PLAZA, SUITE 4600, NEW YORK, NY, UNITED STATES, pp. 715–726, Sep. 2020. doi: 10.1007/s00146-019-00919-6.
- [73] M. Mullins, C. P. Holland, and M. Cunneen, “Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market,” *Patterns*, vol. 2, no. 10, p. 100362, 2021, doi: <https://doi.org/10.1016/j.patter.2021.100362>.
- [74] A. S. Villar and N. Khan, “Robotic process automation in banking industry: a case study on Deutsche Bank,” *J BANK FINANC TECHNOL*, vol. 5, no. 1, pp. 71–86, Jun. 2021, doi: 10.1007/s42786-021-00030-9.
- [75] Y.-Q. Zhu, J. ueline Corbett, and Y.-T. Chiu, “Understanding employees’ responses to artificial intelligence,” *Organizational Dynamics*, vol. 50, no. 2, p. 100786, 2021, doi: <https://doi.org/10.1016/j.orgdyn.2020.100786>.
- [76] L. Willcocks, “Robo-Apocalypse cancelled? Reframing the automation and future of work debate,” *JOURNAL OF INFORMATION TECHNOLOGY*, vol. 35, no. 4. SAGE PUBLICATIONS LTD, 1 OLIVERS YARD, 55 CITY ROAD, LONDON EC1Y 1SP, ENGLAND, pp. 286–302, Dec. 2020. doi: 10.1177/0268396220925830.

- [77] “Why Robots and AI May Not Herald a Job Apocalypse,” *Stanford HAI*. <https://hai.stanford.edu/news/why-robots-and-ai-may-not-herald-job-apocalypse> (accessed Feb. 26, 2022).
- [78] H. Ongori, “A review of the literature on employee turnover,” *African Journal of Business Man*, vol. 1, pp. 49–54, Jun. 2007.
- [79] J. (Justin) Li, M. A. Bonn, and B. H. Ye, “Hotel employee’s artificial intelligence and robotics awareness and its impact on turnover intention: The moderating roles of perceived organizational support and competitive psychological climate,” *Tourism Management*, vol. 73, pp. 172–181, Aug. 2019, doi: 10.1016/j.tourman.2019.02.006.
- [80] E. Chalmers, “Machine Learning With Certainty: A Requirement For Intelligent Process Automation,” *2018 17TH IEEE INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND APPLICATIONS (ICMLA)*. IEEE, 345 E 47TH ST, NEW YORK, NY 10017 USA, pp. 299–304, 2018. doi: 10.1109/ICMLA.2018.00051.
- [81] D. N. Wagner, “The nature of the Artificially Intelligent Firm - An economic investigation into changes that AI brings to the firm,” *Telecommunications Policy*, vol. 44, no. 6, p. 101954, 2020, doi: <https://doi.org/10.1016/j.telpol.2020.101954>.
- [82] “Amazon Web Services’ third outage in a month exposes a weak point in the Internet’s backbone,” *Washington Post*. Accessed: Feb. 26, 2022. [Online]. Available: <https://www.washingtonpost.com/business/2021/12/22/amazon-web-services-experiences-another-big-outage/>
- [83] K. Burden, “Impact of disruptive technologies on sourcing and outsourcing transactions,” *Computer Law & Security Review*, vol. 34, no. 4, pp. 886–889, 2018, doi: <https://doi.org/10.1016/j.clsr.2018.05.022>.
- [84] A. Jain and S. Ranjan, “Implications of emerging technologies on the future of work,” *IIMB Management Review*, vol. 32, no. 4, pp. 448–454, 2020, doi: <https://doi.org/10.1016/j.iimb.2020.11.004>.
- [85] R. Scheepers, M. C. Lacity, and L. P. Willcocks, “Cognitive Automation as Part of Deakin University’s Digital Strategy,” *MIS QUARTERLY EXECUTIVE*, vol. 17, no. 2. INDIANA UNIV, OPER & DECISION TECHNOL DEPT, KELLEY SCH BUS, E 10 ST, BLOOMINGTON, IN 47405-1701 USA, pp. 89–107, Jun. 2018.
- [86] C. Andersson, A. Hallin, and C. Ivory, “Unpacking the digitalisation of public services: Configuring work during automation in local government,” *Government Information Quarterly*, p. 101662, 2021, doi: <https://doi.org/10.1016/j.giq.2021.101662>.

- [87] S. Richardson, "Cognitive automation: A new era of knowledge work?," *Business Information Review*, vol. 37, no. 4, pp. 182–189, Dec. 2020, doi: 10.1177/0266382120974601.
- [88] R. Plattfaut and J. Koch, "Preserving the legacy – Why do professional soccer clubs (not) adopt innovative process technologies? A grounded theory study," *Journal of Business Research*, vol. 136, pp. 237–250, 2021, doi: <https://doi.org/10.1016/j.jbusres.2021.07.024>.
- [89] J. A. Mattu Jeff Larson, Lauren Kirchner, Surya, "Machine Bias," *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=kEFLF-0TKt0pC8fA7TbUZGFrm2mn5Ihm> (accessed Feb. 26, 2022).
- [90] N. Martin, "13 Best Quotes About The Future Of Artificial Intelligence," *Forbes*. <https://www.forbes.com/sites/nicolemartin1/2019/06/27/13-greatest-quotes-about-the-future-of-artificial-intelligence/> (accessed Mar. 26, 2022).
- [91] L. Nazareno and D. S. Schiff, "The impact of automation and artificial intelligence on worker well-being," *Technology in Society*, vol. 67, p. 101679, 2021, doi: <https://doi.org/10.1016/j.techsoc.2021.101679>.
- [92] N. Chalofsky, "An emerging construct for meaningful work," *Human Resource Development International*, vol. 6, no. 1, pp. 69–83, Mar. 2003, doi: 10.1080/1367886022000016785.
- [93] L. Firth, D. Mellor, K. (Kate) Moore, and C. Loquet, "How Can Managers Reduce Employee Intention to Quit?," *Journal of Managerial Psychology*, vol. 19, pp. 170–187, Mar. 2004, doi: 10.1108/02683940410526127.
- [94] H. Hur, "Job security matters: A systematic review and meta-analysis of the relationship between job security and work attitudes," *Journal of Management & Organization*, pp. 1–31, Mar. 2019, doi: 10.1017/jmo.2019.3.
- [95] H. Parker and S. E. Appel, "ON THE PATH TO ARTIFICIAL INTELLIGENCE: THE EFFECTS OF A ROBOTICS SOLUTION IN A FINANCIAL SERVICES FIRM," *SOUTH AFRICAN JOURNAL OF INDUSTRIAL ENGINEERING*, vol. 32, no. 2. SOUTHERN AFRICAN INST INDUSTRIAL ENGINEERING, UNIV PRETORIA, DEPT INDUSTRIAL SYSTEMS ENGINEERING, PRETORIA, 0001, SOUTH AFRICA, pp. 37–47, Aug. 2021. doi: 10.7166/32-2-2390.
- [96] A. Bhargava, M. Bester, and L. Bolton, "Employees' Perceptions of the Implementation of Robotics, Artificial Intelligence, and Automation (RAIA) on Job Satisfaction, Job Security, and Employability," *J. technol. behav. sci.*, vol. 6, no. 1, pp. 106–113, Mar. 2021, doi: 10.1007/s41347-020-00153-8.

- [97] H. Ismail, "Job Insecurity, Burnout and Intention to Quit," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2889704, Apr. 2015. Accessed: Feb. 26, 2022. [Online]. Available: <https://papers.ssrn.com/abstract=2889704>
- [98] D. J. Stanley, J. P. Meyer, and L. Topolnytsky, "Employee Cynicism and Resistance to Organizational Change," *J Bus Psychol*, vol. 19, no. 4, pp. 429–459, Jun. 2005, doi: 10.1007/s10869-005-4518-2.
- [99] R. Hodson, "Organizational Trustworthiness: Findings from the Population of Organizational Ethnographies," *Organization Science*, vol. 15, no. 4, pp. 432–445, Aug. 2004, doi: 10.1287/orsc.1040.0077.
- [100] C. (Abigail) Zhang, "Intelligent Process Automation in Audit," *JOURNAL OF EMERGING TECHNOLOGIES IN ACCOUNTING*, vol. 16, no. 2. AMER ACCOUNTING ASSOC, 5717 BESSIE DR, SARASOTA, FL 34233 USA, pp. 69–88, 2019. doi: 10.2308/jeta-52653.
- [101] G. Mahala, R. Sindhgatta, H. Dam, and A. Ghose, "Designing Optimal Robotic Process Automation Architectures," 2020, pp. 448–456. doi: 10.1007/978-3-030-65310-1_32.
- [102] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans, "Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts," *Journal of Artificial Intelligence Research*, vol. 62, pp. 729–754, Jul. 2018, doi: 10.1613/jair.1.11222.
- [103] "Automation 360 IQ Bot." <https://docs.automationanywhere.com/bundle/enterprise-v2019/page/enterprise-cloud/topics/iq-bot/cloud-iqb-process-overview.html> (accessed Mar. 01, 2022).
- [104] "Intelligent Document Processing: Decipher IDP," *Blue Prism*. <https://www.blueprism.com/products/decipheridp/> (accessed Mar. 01, 2022).
- [105] U. Inc, "Document Understanding - AI Document Processing | UiPath." <https://www.uipath.com/product/document-understanding> (accessed Mar. 01, 2022).
- [106] U. Inc, "RPA & AI Integration with AI Center | UiPath." <https://www.uipath.com/product/rpa-ai-integration-with-ai-center> (accessed Mar. 01, 2022).
- [107] "Keys to RPA Success – Part One: Becoming Strategic With RPA," *Blue Prism*. <https://www.blueprism.com/resources/white-papers/keys-to-rpa-success-part-one-becoming-strategic-with-rpa/> (accessed Mar. 01, 2022).
- [108] "What is the Cost to Deploy and Maintain a Machine Learning Model?," *phData*, May 20, 2021. <https://www.phdata.io/blog/what-is-the-cost-to-deploy-and-maintain-a-machine-learning-model/> (accessed Mar. 01, 2022).

- [109] R. Incze, "The Cost of Machine Learning Projects," *Cognifeed*, Sep. 14, 2019. <https://medium.com/cognifeed/the-cost-of-machine-learning-projects-7ca3aea03a5c> (accessed Mar. 01, 2022).
- [110] S. Anagnoste, "Setting Up a Robotic Process Automation Center of Excellence," *Management Dynamics in the Knowledge Economy*, vol. 6, no. 2, pp. 307–332, 2018.
- [111] P. Marciniak and R. Stanislawski, "Internal Determinants in the Field of RPA Technology Implementation on the Example of Selected Companies in the Context of Industry 4.0 Assumptions," *INFORMATION*, vol. 12, no. 6. MDPI, ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND, Jun. 2021. doi: 10.3390/info12060222.
- [112] M. Romao, J. Costa, and C. J. Costa, "Robotic Process Automation: A case study in the Banking Industry," *2019 14TH IBERIAN CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGIES (CISTI)*. IEEE, 345 E 47TH ST, NEW YORK, NY 10017 USA, 2019.
- [113] "Wild patterns: Ten years after the rise of adversarial machine learning - ScienceDirect." <https://www.sciencedirect.com/science/article/pii/S0031320318302565> (accessed Mar. 03, 2022).
- [114] A. Martínez-Rojas, J. Sánchez-Oliva, J. M. López-Carnicer, and A. Jiménez-Ramírez, "AIRPA: An Architecture to Support the Execution and Maintenance of AI-Powered RPA Robots," in *Business Process Management: Blockchain and Robotic Process Automation Forum*, Cham, 2021, pp. 38–48. doi: 10.1007/978-3-030-85867-4_4.
- [115] R. Bommasani *et al.*, "On the Opportunities and Risks of Foundation Models," *arXiv:2108.07258 [cs]*, Aug. 2021, Accessed: May 03, 2022. [Online]. Available: <http://arxiv.org/abs/2108.07258>
- [116] H. Suresh and J. V. Gutttag, "A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle," *Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–9, Oct. 2021, doi: 10.1145/3465416.3483305.
- [117] R. Ashmore, R. Calinescu, and C. Paterson, "Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges," *ACM Comput. Surv.*, vol. 54, no. 5, p. 111:1-111:39, May 2021, doi: 10.1145/3453444.
- [118] "BOP Statistics: Inmate Race." https://www.bop.gov/about/statistics/statistics_inmate_race.jsp (accessed Mar. 05, 2022).
- [119] "U.S. Census Bureau QuickFacts: United States." <https://www.census.gov/quickfacts/fact/table/US/LFE046219> (accessed Mar. 05, 2022).

- [120] J. Travis, B. Western, and F. Redburn, "The Growth of Incarceration in the United States: Exploring Causes and Consequences," *Publications and Research*, Jan. 2014, [Online]. Available: https://academicworks.cuny.edu/jj_pubs/27
- [121] C. Sáez, N. Romero, J. A. Conejero, and J. M. García-Gómez, "Potential limitations in COVID-19 machine learning due to data source variability: A case study in the nCov2019 dataset," *Journal of the American Medical Informatics Association*, vol. 28, no. 2, pp. 360–364, Feb. 2021, doi: 10.1093/jamia/ocaa258.
- [122] V. J. Burroughs, R. W. Maxey, and R. A. Levy, "Racial and ethnic differences in response to medicines: towards individualized pharmaceutical treatment.," *J Natl Med Assoc*, vol. 94, no. 10 Suppl, pp. 1–26, Oct. 2002.
- [123] L. Baier, F. Jöhren, and S. Seebacher, *CHALLENGES IN THE DEPLOYMENT AND OPERATION OF MACHINE LEARNING IN PRACTICE*. 2019.
- [124] P. Jain, M. Gyanchandani, and N. Khare, "Big data privacy: a technological perspective and review," *J Big Data*, vol. 3, no. 1, p. 25, Nov. 2016, doi: 10.1186/s40537-016-0059-y.
- [125] G. Lasso-Rodriguez and R. Gil-Herrera, "ADVANCED HUMAN-ROBOT INTERACTION FOR LEARNING WITH ROBOTIC PROCESS AUTOMATION," *12TH INTERNATIONAL CONFERENCE OF EDUCATION, RESEARCH AND INNOVATION (ICERI2019)*. IATED-INT ASSOC TECHNOLOGY EDUCATION & DEVELOPMENT, LAURI VOLPI 6, VALENICA, BURJASSOT 46100, SPAIN, pp. 7718–7723, 2019.
- [126] M. Gotthardt, D. Koivulaakso, O. Paksoy, C. Saramo, M. Martikainen, and O. Lehner, "Current State and Challenges in the Implementation of Smart Robotic Process Automation in Accounting and Auditing," *ACRN Journal of Finance and Risk Perspectives*, vol. 9, pp. 90–102, May 2020, doi: 10.35944/jofrp.2020.9.1.007.
- [127] "Sample Size in Machine Learning and Artificial Intelligence – Perioperative Data Science." <https://sites.uab.edu/periop-datascience/2021/06/28/sample-size-in-machine-learning-and-artificial-intelligence/> (accessed Mar. 08, 2022).
- [128] I. Giuffrida, "Liability for AI Decision-Making: Some Legal and Ethical Considerations," *Fordham Law Review*, vol. 88, no. 2, p. 439, Nov. 2019.
- [129] P. Čerka, J. Grigienė, and G. Sirbikytė, "Liability for damages caused by artificial intelligence," *Computer Law & Security Review*, vol. 31, no. 3, pp. 376–389, Jun. 2015, doi: 10.1016/j.clsr.2015.03.008.

- [130] H.-L. Truong and T.-M. Nguyen, "QoA4ML - A Framework for Supporting Contracts in Machine Learning Services," in *2021 IEEE International Conference on Web Services (ICWS)*, Sep. 2021, pp. 465–475. doi: 10.1109/ICWS53863.2021.00066.
- [131] U. Bhatt *et al.*, "Explainable machine learning in deployment," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona Spain, Jan. 2020, pp. 648–657. doi: 10.1145/3351095.3375624.
- [132] K. Shanmugalingam, N. Chandrasekara, C. Hindle, G. Fernando, and C. Gunawardhana, "Corporate IT-support Help-Desk Process Hybrid-Automation Solution with Machine Learning Approach," *2019 DIGITAL IMAGE COMPUTING: TECHNIQUES AND APPLICATIONS (DICTA)*. IEEE, 345 E 47TH ST, NEW YORK, NY 10017 USA, pp. 359–365, 2019.
- [133] U. Gasser and V. A. F. Almeida, "A Layered Model for AI Governance," *IEEE Internet Computing*, vol. 21, no. 6, pp. 58–62, Nov. 2017, doi: 10.1109/MIC.2017.4180835.
- [134] V. Chandrasekaran, H. Jia, A. Thudi, A. Travers, M. Yaghini, and N. Papernot, "SoK: Machine Learning Governance," Sep. 2021, doi: 10.48550/arXiv.2109.10870.
- [135] "IEEE Ethics In Action in Autonomous and Intelligent Systems | IEEE SA," *Ethics In Action / Ethically Aligned Design*. <https://ethicsinaction.ieee.org/> (accessed Mar. 14, 2022).
- [136] S. Reddy, S. Allan, S. Coghlan, and P. Cooper, "A governance model for the application of AI in health care," *J Am Med Inform Assoc*, vol. 27, no. 3, pp. 491–497, Mar. 2020, doi: 10.1093/jamia/ocz192.
- [137] J. Schneider and C. Meske, "AI GOVERNANCE FOR BUSINESSES," p. 19.
- [138] T. Grønsund and M. Aanestad, "Augmenting the algorithm: Emerging human-in-the-loop work configurations," *The Journal of Strategic Information Systems*, vol. 29, no. 2, p. 101614, 2020, doi: <https://doi.org/10.1016/j.jsis.2020.101614>.
- [139] "Risk analysis and management." <https://www.pmi.org/learning/library/risk-analysis-project-management-7070> (accessed Mar. 18, 2022).
- [140] "Best practice for managing organizational risk | Axelos." <https://www.axelos.com/resource-hub/blog/best-practice-for-managing-organizational-risk> (accessed Mar. 18, 2022).
- [141] "ISO - ISO 31000 — Risk management," *ISO*. <https://www.iso.org/iso-31000-risk-management.html> (accessed Mar. 18, 2022).

- [142] J. Viehhauser and M. Doerr, "Digging for Gold in RPA Projects – A Quantifiable Method to Identify and Prioritize Suitable RPA Process Candidates," in *Advanced Information Systems Engineering*, Cham, 2021, pp. 313–327. doi: 10.1007/978-3-030-79382-1_19.
- [143] A. Meironke and S. Kuehnel, "How to Measure RPA's Benefits? A Review on Metrics, Indicators, and Evaluation Methods of RPA Benefit Assessment," *Wirtschaftsinformatik 2022 Proceedings*, Jan. 2022, [Online]. Available: <https://aisel.aisnet.org/wi2022/bpm/bpm/5>
- [144] "What's Next for AI Ethics, Policy, and Governance? A Global Overview | Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society." <https://dl.acm.org/doi/abs/10.1145/3375627.3375804> (accessed Mar. 22, 2022).
- [145] "AI Principles," *Future of Life Institute*, Aug. 11, 2017. <https://futureoflife.org/2017/08/11/ai-principles/> (accessed Mar. 22, 2022).
- [146] "The Declaration - Montreal Responsible AI," *respaideclaration*. <https://www.montrealdeclaration-responsibleai.com/the-declaration> (accessed Mar. 22, 2022).
- [147] Buckingham and D. A, "Associations Among Stress, Work Overload, Role Conflict, and Self-Efficacy in Maine Principals," May 2004.
- [148] B. Fc and G. (Gert) Roodt, "The validation of the turnover intention scale," *SA Journal of Human Resource Management*, vol. 11, Jan. 2013, doi: 10.4102/sajhrm.v11i1.507.
- [149] L. L. Pipino, Y. W. Lee, and R. Y. Wang, "Data quality assessment," *Commun. ACM*, vol. 45, no. 4, pp. 211–218, Apr. 2002, doi: 10.1145/505248.506010.
- [150] "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era." <https://datascience.codata.org/articles/10.5334/dsj-2015-002/> (accessed Mar. 21, 2022).
- [151] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," *ACM Comput. Surv.*, vol. 41, no. 3, p. 16:1-16:52, Jul. 2009, doi: 10.1145/1541880.1541883.
- [152] "Monitor feature skew and drift | Vertex AI," *Google Cloud*. <https://cloud.google.com/vertex-ai/docs/model-monitoring/using-model-monitoring> (accessed Mar. 21, 2022).
- [153] "Amazon SageMaker Model Monitor | ML Model Accuracy | Amazon Web Services," *Amazon Web Services, Inc.* <https://aws.amazon.com/sagemaker/model-monitor/> (accessed Mar. 21, 2022).

- [154] buchananwp, "Detect data drift on datasets (preview) - Azure Machine Learning." <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-datasets> (accessed Mar. 21, 2022).
- [155] M. Janssen, P. Brous, E. Estevez, L. S. Barbosa, and T. Janowski, "Data governance: Organizing data for trustworthy Artificial Intelligence," *Government Information Quarterly*, vol. 37, no. 3, p. 101493, Jul. 2020, doi: 10.1016/j.giq.2020.101493.
- [156] C. Harbottle, "Office Automation: A Social and Organizational Perspective," *undefined*, 1986, Accessed: Mar. 20, 2022. [Online]. Available: <https://www.semanticscholar.org/paper/User-Involvement-in-Office-Automation%3A-Overcoming-Kaddah-Gough/c999b296091f127b838e704fe99f759e23636a5e>
- [157] M. F. Steger, B. J. Dik, and R. D. Duffy, "Measuring Meaningful Work: The Work and Meaning Inventory (WAMI)," *Journal of Career Assessment*, vol. 20, no. 3, pp. 322–337, Aug. 2012, doi: 10.1177/1069072711436160.
- [158] M. Lips-Wiersma and S. Wright, "Measuring the Meaning of Meaningful Work: Development and Validation of the Comprehensive Meaningful Work Scale (CMWS)," *Group & Organization Management*, vol. 37, no. 5, pp. 655–685, Oct. 2012, doi: 10.1177/1059601112461578.
- [159] J. M. C. Both-Nwabuwe, M. T. M. Dijkstra, and B. Beersma, "Sweeping the Floor or Putting a Man on the Moon: How to Define and Measure Meaningful Work," *Frontiers in Psychology*, vol. 8, 2017, Accessed: Mar. 22, 2022. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fpsyg.2017.01658>
- [160] E. I. Lysova, B. A. Allan, B. J. Dik, R. D. Duffy, and M. F. Steger, "Fostering meaningful work in organizations: A multi-level review and integration," *Journal of Vocational Behavior*, vol. 110, pp. 374–389, Feb. 2019, doi: 10.1016/j.jvb.2018.07.004.
- [161] T. M. Probst, "Development and validation of the Job Security Index and the Job Security Satisfaction scale: A classical test theory and IRT approach," *Journal of Occupational and Organizational Psychology*, vol. 76, no. 4, pp. 451–467, 2003, doi: 10.1348/096317903322591587.
- [162] J. M. Brett and F. Drasgow, Eds., *The Psychology of Work: Theoretically Based Empirical Research*, 0 ed. Psychology Press, 2002. doi: 10.4324/9781410602411.
- [163] "Artificial Intelligence for the Real World," *Harvard Business Review*, Jan. 01, 2018. Accessed: Mar. 23, 2022. [Online]. Available: <https://hbr.org/2018/01/artificial-intelligence-for-the-real-world>

- [164] C. Caldwell and S. Clapham, "Organizational Trustworthiness: An International Perspective," *Journal of Business Ethics*, vol. 47, pp. 349–364, Nov. 2003, doi: 10.1023/A:1027370104302.
- [165] Y. J. Cho and J. W. Lee, "Perceived Trustworthiness of Supervisors, Employee Satisfaction and Cooperation," *Public Management Review*, vol. 13, no. 7, pp. 941–965, Oct. 2011, doi: 10.1080/14719037.2011.589610.
- [166] P. J. Franta, "A validation study of Shaw's assessment of organizational trustworthiness," Ph.D., University of Missouri - Columbia, United States -- Missouri. Accessed: Mar. 24, 2022. [Online]. Available: <https://www.proquest.com/docview/304613482/abstract/9D02BE5D40684B3BPQ/1>
- [167] L. Brinded, "Boss of the world's largest recruiter: 'One-off education followed by a career will no longer work,'" *Business Insider*. <https://www.businessinsider.com/davos-alain-dehaze-ceo-adecco-interview-tech-skills-jobs-ai-robots-2017-1> (accessed Mar. 26, 2022).
- [168] "Top Ten Best Quotes By Elon Musk On AI," *Analytics India Magazine*, Feb. 29, 2020. <https://analyticsindiamag.com/top-ten-best-quotes-by-elon-musk-on-artificial-intelligence/> (accessed Mar. 26, 2022).
- [169] "When will singularity happen? 995 experts' opinions on AGI," Aug. 08, 2017. <https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/> (accessed Mar. 27, 2022).
- [170] R. Fjelland, "Why general artificial intelligence will not be realized," *Humanit Soc Sci Commun*, vol. 7, no. 1, Art. no. 1, Jun. 2020, doi: 10.1057/s41599-020-0494-4.
- [171] T. Lane and C. E. Brodley, "An application of machine learning to anomaly detection," in *Proceedings of the 20th national information systems security conference*, 1997, vol. 377, pp. 366–380.
- [172] A. Unler and A. Murat, "A discrete particle swarm optimization method for feature selection in binary classification problems," *European Journal of Operational Research*, vol. 206, no. 3, pp. 528–539, Nov. 2010, doi: 10.1016/j.ejor.2010.02.032.
- [173] B. Shawar and E. Atwell, "Chatbots: Are they Really Useful?," *LDV Forum*, vol. 22, pp. 29–49, Jan. 2007.
- [174] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Computational Intelligence and Neuroscience*, vol. 2018, p. e7068349, Feb. 2018, doi: 10.1155/2018/7068349.

- [175] H. Borko and M. Bernick, "Automatic Document Classification," *J. ACM*, vol. 10, no. 2, pp. 151–162, Apr. 1963, doi: 10.1145/321160.321165.
- [176] B. Klimt and Y. Yang, "The Enron Corpus: A New Dataset for Email Classification Research," in *Machine Learning: ECML 2004*, Berlin, Heidelberg, 2004, pp. 217–226. doi: 10.1007/978-3-540-30115-8_22.
- [177] J. Chen and W. K. Jenkins, "Facial recognition with PCA and machine learning methods," in *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug. 2017, pp. 973–976. doi: 10.1109/MWSCAS.2017.8053088.
- [178] N. K. Ahmed, A. F. Atiya, N. E. Gayar, and H. El-Shishiny, "An Empirical Comparison of Machine Learning Models for Time Series Forecasting," *Econometric Reviews*, vol. 29, no. 5–6, pp. 594–621, Aug. 2010, doi: 10.1080/07474938.2010.481556.
- [179] M. Aly, "Survey on Multiclass Classification Methods," p. 9.
- [180] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Linguisticæ Investigationes*, vol. 30, no. 1, pp. 3–26, Jan. 2007, doi: 10.1075/li.30.1.03nad.
- [181] B. Baharudin, L. H. Lee, and K. Khan, "A Review of Machine Learning Algorithms for Text-Documents Classification," *JAIT*, vol. 1, no. 1, pp. 4–20, Feb. 2010, doi: 10.4304/jait.1.1.4-20.
- [182] K. Grauman and B. Leibe, "Visual Object Recognition," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 5, no. 2, pp. 1–181, Apr. 2011, doi: 10.2200/S00332ED1V01Y201103AIM011.
- [183] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," *IEEE Access*, vol. 8, pp. 142642–142668, 2020, doi: 10.1109/ACCESS.2020.3012542.
- [184] R. Feldman, "Techniques and Applications For Sentiment Analysis." <https://cacm.acm.org/magazines/2013/4/162501-techniques-and-applications-for-sentiment-analysis/fulltext> (accessed Feb. 26, 2022).
- [185] J. Zhang and C. Zong, "Deep Neural Networks in Machine Translation: An Overview," *IEEE Intelligent Systems*, vol. 30, no. 5, pp. 16–25, Sep. 2015, doi: 10.1109/MIS.2015.69.
- [186] S. Anagnoste, I. Biclesanu, F. D'Ascenzo, and M. Savastano, "The Role of Chatbots in End-To-End Intelligent Automation and Future Employment Dynamics," in *Business Revolution in a Digital Era*, Cham, 2021, pp. 287–302. doi: 10.1007/978-3-030-59972-0_20.

- [187] N. Ashar, P. Tolar, and K. Bathe, "Sentiment Analysis for Automated Email Response: A Review," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3565490, Apr. 2020. doi: 10.2139/ssrn.3565490.
- [188] M. Cazacu, C. Bodea, and M.-I. Dascalu, "Connecting the Dots Promising directions for developing a robotic cybersecurity analyst," *PROCEEDINGS OF THE 6TH CONFERENCE ON THE ENGINEERING OF COMPUTER BASED SYSTEMS (ECBS 2019)*. ASSOC COMPUTING MACHINERY, 1515 BROADWAY, NEW YORK, NY 10036-9998 USA, 2020. doi: 10.1145/3352700.3352703.
- [189] C. Engel, E. Elshan, and P. Ebel, *Moving Beyond Rule-Based Automation: A Method for Assessing Cognitive Automation Use Cases*. 2021.
- [190] C. Flechsig, F. Anslinger, and R. Lasch, "Robotic Process Automation in purchasing and supply management: A multiple case study on potentials, barriers, and implementation," *Journal of Purchasing and Supply Management*, p. 100718, 2021, doi: <https://doi.org/10.1016/j.pursup.2021.100718>.
- [191] R. Götzen, G. Schuh, V. Stich, and R. Conrad, "Classification of software-based automation technologies: Derivation of characteristics through an empirical investigation," in *2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, Jun. 2021, pp. 1–9. doi: 10.1109/ICE/ITMC52061.2021.9570264.
- [192] A. Guha and D. Samanta, "Hybrid Approach to Document Anomaly Detection: An Application to Facilitate RPA in Title Insurance," *INTERNATIONAL JOURNAL OF AUTOMATION AND COMPUTING*, vol. 18, no. 1. SPRINGER NATURE, CAMPUS, 4 CRINAN ST, LONDON, N1 9XW, ENGLAND, pp. 55–72, Feb. 2021. doi: 10.1007/s11633-020-1247-y.
- [193] A. Haleem, M. Javaid, R. P. Singh, S. Rab, and R. Suman, "Hyperautomation for the enhancement of automation in industries," *Sensors International*, vol. 2, p. 100124, 2021, doi: <https://doi.org/10.1016/j.sintl.2021.100124>.
- [194] J. L. Hartley and W. J. Sawaya, "Tortoise, not the hare: Digital transformation of supply chain business processes," *BUSINESS HORIZONS*, vol. 62, no. 6, SI. ELSEVIER, RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS, pp. 707–715, Dec. 2019. doi: 10.1016/j.bushor.2019.07.006.
- [195] F. Huang and M. A. Vasarhelyi, "Applying robotic process automation (RPA) in auditing: A framework," *International Journal of Accounting Information Systems*, vol. 35, p. 100433, 2019, doi: <https://doi.org/10.1016/j.accinf.2019.100433>.

- [196] P. D. Hung, D. T. Trang, and T. Khai, "Integrating Chatbot and RPA into Enterprise Applications Based on Open, Flexible and Extensible Platforms," in *Cooperative Design, Visualization, and Engineering*, Cham, 2021, pp. 183–194. doi: 10.1007/978-3-030-88207-5_18.
- [197] A. Januszewski, J. Kujawski, and N. Buchalska-Sugajska, "Benefits of and Obstacles to RPA Implementation in Accounting Firms," *Procedia Computer Science*, vol. 192, pp. 4672–4680, 2021, doi: <https://doi.org/10.1016/j.procs.2021.09.245>.
- [198] J. Kokina and S. Blanchette, "Early evidence of digital labor in accounting: Innovation with Robotic Process Automation," *International Journal of Accounting Information Systems*, vol. 35, p. 100431, 2019, doi: <https://doi.org/10.1016/j.accinf.2019.100431>.
- [199] T. Korhonen, E. Selos, T. Laine, and P. Suomala, "Exploring the programmability of management accounting work for increasing automation: an interventionist case study," *ACCOUNTING AUDITING & ACCOUNTABILITY JOURNAL*, vol. 34, no. 2. EMERALD GROUP PUBLISHING LTD, HOWARD HOUSE, WAGON LANE, BINGLEY BD16 1WA, W YORKSHIRE, ENGLAND, pp. 253–280, Mar. 01, 2021. doi: 10.1108/AAAJ-12-2016-2809.
- [200] F. Krieger, P. Drews, and P. Velte, "Explaining the (non-) adoption of advanced data analytics in auditing: A process theory," *International Journal of Accounting Information Systems*, vol. 41, p. 100511, 2021, doi: <https://doi.org/10.1016/j.accinf.2021.100511>.
- [201] M. Lacity, L. Willcocks, and D. Gozman, "Influencing information systems practice: The action principles approach applied to robotic process and cognitive automation," *JOURNAL OF INFORMATION TECHNOLOGY*, vol. 36, no. 3. SAGE PUBLICATIONS LTD, 1 OLIVERS YARD, 55 CITY ROAD, LONDON EC1Y 1SP, ENGLAND, pp. 216–240, Sep. 2021. doi: 10.1177/0268396221990778.
- [202] M. Lahlali, N. Berbiche, and J. El Alami, "How Enterprise must be Prepared to be 'AI First'?", *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, vol. 12, no. 5. SCIENCE & INFORMATION SAI ORGANIZATION LTD, 19 BOLLING RD, BRADFORD, WEST YORKSHIRE, 00000, ENGLAND, pp. 346–351, May 2021.
- [203] G. Lasso R. and R. Gil Herrera, *TRAINING THE TEACHERS WITH ASSISTANCE OF ROBOTIC PROCESS AUTOMATION*. 2020. doi: 10.21125/inted.2020.2373.
- [204] G. Lasso-Rodriguez and K. Winkler, "Hyperautomation to fulfil jobs rather than executing tasks: the BPM manager robot vs human case," *ROMANIAN JOURNAL OF INFORMATION TECHNOLOGY AND AUTOMATIC CONTROL-REVISTA ROMANA DE INFORMATICA SI AUTOMATICA*, vol. 30, no. 3. INST NATL CERCETARE-DEZVOLTARE INFORMATICA-ICI, 8-10

MARESAL A AVERESCU AV, SECTOR 1, BUCHAREST, 011455, ROMANIA, pp. 7–22, 2020. doi: 10.33436/v30i3y202001.

[205] X. Ling, M. Gao, and D. Wang, “Intelligent document processing based on RPA and machine learning,” *2020 CHINESE AUTOMATION CONGRESS (CAC 2020)*. IEEE, 345 E 47TH ST, NEW YORK, NY 10017 USA, pp. 1349–1353, 2020. doi: 10.1109/CAC51589.2020.9326579.

[206] P. Martins, F. Sa, F. Morgado, and C. Cunha, “Using machine learning for cognitive Robotic Process Automation (RPA),” *2020 15TH IBERIAN CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGIES (CISTI'2020)*. IEEE, 345 E 47TH ST, NEW YORK, NY 10017 USA, 2020.

[207] A. Masood and A. Hashmi, “Cognitive Robotics Process Automation: Automate This!,” 2019, pp. 225–287. doi: 10.1007/978-1-4842-4106-6_5.

[208] J. Mendling, G. Decker, R. Hull, H. A. Reijers, and I. Weber, “How do Machine Learning, Robotic Process Automation, and Blockchains Affect the Human Factor in Business Process Management?,” *COMMUNICATIONS OF THE ASSOCIATION FOR INFORMATION SYSTEMS*, vol. 43. ASSOC INFORMATION SYSTEMS, GEORGIA STATE UNIV, 35 BROAD STREET, STE 916-917, ATLANTA, GA 30303 USA, pp. 297–320, 2018. doi: 10.17705/1CAIS.04319.

[209] D. Oza, D. Padhiyar, V. Doshi, and S. Patil, “Insurance Claim Processing Using RPA Along With Chatbot,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3561871, Apr. 2020. doi: 10.2139/ssrn.3561871.

[210] S. Parchande, A. Shahane, and M. Dhore, “Contractual Employee Management System Using Machine Learning and Robotic Process Automation,” *2019 5TH INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION, CONTROL AND AUTOMATION (ICCUBE)*. IEEE, 345 E 47TH ST, NEW YORK, NY 10017 USA, 2019.

[211] A. Pedretti *et al.*, “Robotic Process Automation Extended with Artificial Intelligence Techniques in Power Distribution Utilities,” *BRAZILIAN ARCHIVES OF BIOLOGY AND TECHNOLOGY*, vol. 64. INST TECNOLOGIA PARANA, RUA PROF ALGACYR MUNHOZ MADER 3775-CIC, 81350-010 CURITIBA-PARANA, BRAZIL, 2021. doi: 10.1590/1678-4324-75years-2021210217.

[212] J. Ribeiro, R. Lima, T. Eckhardt, and S. Paiva, “Robotic Process Automation and Artificial Intelligence in Industry 4.0 – A Literature review,” *Procedia Computer Science*, vol. 181, pp. 51–58, 2021, doi: <https://doi.org/10.1016/j.procs.2021.01.104>.

- [213] J. Ribeiro, R. Lima, and S. Paiva, "Document Classification in Robotic Process Automation Using Artificial Intelligence—A Preliminary Literature Review," 2021, pp. 211–221. doi: 10.1007/978-981-16-1089-9_18.
- [214] N. Roopesh and C. N. Babu, "Robotic Process Automation for Resume Processing System," in *2021 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT)*, Aug. 2021, pp. 180–184. doi: 10.1109/RTEICT52294.2021.9573595.
- [215] N. S. Patil, S. Kamanavalli, S. Hiregoudar, S. Jadhav, S. Kanakraddi, and N. D. Hiremath, "Vehicle Insurance Fraud Detection System Using Robotic Process Automation and Machine Learning," in *2021 International Conference on Intelligent Technologies (CONIT)*, Jun. 2021, pp. 1–5. doi: 10.1109/CONIT51480.2021.9498507.
- [216] S. Sarker, L. Jamal, S. F. Ahmed, and N. Irtisam, "Robotics and artificial intelligence in healthcare during COVID-19 pandemic: A systematic review," *Robotics and Autonomous Systems*, vol. 146, p. 103902, 2021, doi: <https://doi.org/10.1016/j.robot.2021.103902>.
- [217] S. Seguin and I. Benkalai, "Robotic Process Automation (RPA) Using an Integer Linear Programming Formulation," *CYBERNETICS AND SYSTEMS*, vol. 51, no. 4. TAYLOR & FRANCIS INC, 530 WALNUT STREET, STE 850, PHILADELPHIA, PA 19106 USA, pp. 357–369, May 18, 2020. doi: 10.1080/01969722.2020.1770503.
- [218] G. Shidaganti, S. Salil, P. Anand, and V. Jadhav, "Robotic Process Automation with AI and OCR to Improve Business Process: Review," in *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Aug. 2021, pp. 1612–1618. doi: 10.1109/ICESC51422.2021.9532902.
- [219] R. Sindhgatta, A. H. M. ter Hofstede, and A. Ghose, "Resource-Based Adaptive Robotic Process Automation Formal/Technical Paper," *ADVANCED INFORMATION SYSTEMS ENGINEERING, CAISE 2020*, vol. 12127. SPRINGER INTERNATIONAL PUBLISHING AG, GEWERBESTRASSE 11, CHAM, CH-6330, SWITZERLAND, pp. 451–466, 2020. doi: 10.1007/978-3-030-49435-3_28.
- [220] P. Ulrich and V. Frank, "Relevance and Adoption of AI technologies in German SMEs – Results from Survey-Based Research," *Procedia Computer Science*, vol. 192, pp. 2152–2159, 2021, doi: <https://doi.org/10.1016/j.procs.2021.08.228>.
- [221] B. Vajgel *et al.*, "Development of Intelligent Robotic Process Automation: A Utility Case Study in Brazil," *IEEE Access*, vol. 9, pp. 71222–71235, 2021, doi: 10.1109/ACCESS.2021.3075693.

- [222] W. M. P. van der Aalst, "Hybrid Intelligence: to automate or not to automate, that is the question," *IJISPM-INTERNATIONAL JOURNAL OF INFORMATION SYSTEMS AND PROJECT MANAGEMENT*, vol. 9, no. 2. SCIKA, SCIKA, CORRELHA, 00000, PORTUGAL, pp. 5–20, 2021. doi: 10.12821/ijispm090201.
- [225] B. Vinod, "Artificial Intelligence in travel," *JOURNAL OF REVENUE AND PRICING MANAGEMENT*, vol. 20, no. 3, SI. PALGRAVE MACMILLAN LTD, BRUNEL RD BLDG, HOUNDMILLS, BASINGSTOKE RG21 6XS, HANTS, ENGLAND, pp. 368–375, Jun. 2021. doi: 10.1057/s41272-021-00319-w.
- [226] A. Wróblewska, T. Stanisławek, B. Prus-Zajączkowski, and Ł. Garncarek, "Robotic process automation of unstructured data with machine learning," *Annals of Computer Science and Information Systems*, vol. Vol. 16, 2018, doi: 10.15439/2018F373.
- [227] S. Yoon, "A Study on the Transformation of Accounting Based on New Technologies: Evidence from Korea," *SUSTAINABILITY*, vol. 12, no. 20. MDPI, ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND, Oct. 2020. doi: 10.3390/su12208669.