Szent István University

Doctoral School of Economic and Regional Sciences

The Importance of Learning in the Age of Artificial Intelligence: The Evolution of the Role of the Auditor

Ph.D. Dissertation

Mario Malouin

Gödöllő, Hungary

2020

Szent István University, Hungary

Name of Doctoral School: Doctoral School of Economic and Regional Sciences

Discipline: Management and Business Administration Sciences

Head of Doctoral School: **Prof. Dr. H.c. Popp, József, DSC**
Corresponding member of the Hungarian Academy of Sciences
Szent István University
Faculty of Economics and Social Sciences
Institute of Agribusiness

Supervisor(s): **Prof. Dr. Makó, Csaba, DSC**
Szent István University
Faculty of Economic and Social Sciences

Prof. Dr. Koplyay, Tamas-Michael, PhD
Université du Québec en Outaouais, Québec
Department of Administrative Sciences

TABLE OF CONTENTS

List of Tables

List of Figures

ABBREVIATIONS

AAA: American Accounting Association

AAAI: Association for the Advancement of Artificial Intelligence

ACFE: Association of Certified Fraud Examiners

AI: Artificial intelligence

BDI: Belief, Desire, Intention

BRT: Business Roundtable

CAS: Canadian Auditing Standards

CEO: Chief Executive Officer

CFE: Common Final Examination

CFO: Chief Financial Officer

CPA: Chartered Professional Accountant

DARPA: (US) Defense Advanced Research Projects Agency

DL: Deep Learning

DPA: Deferred Prosecution Agreement

EA: Evolutionary Algorithms

EDC: Export Development Canada

ERP: Enterprise Resource Planning

ESG: Environmental, Societal, Governance

IEEE: Institute of Electrical and Electronics Engineers

FAT: Fairness, Accountability, Transparency

FICO: Fair Isaac Corporation

FT: Financial Times

GAAP: Generally Accepted Accounting Principles

GAAS: Generally Accepted Auditing Standards

GOFAI: Good Old-Fashion Artificial Intelligence

MIT: Massachusetts Institute of Technology

ML: Machine Learning

NRD: Nintendo Research and Development

IP: Intellectual Property

OECD: Organization for Economic Cooperation Development

PEP: Professional Education Program

RCMP: Royal Canadian Mounted Police

R&D: Research and Development

RTS: Real-Time Strategy

VUCA: Volatile, Uncertain, Complex, Ambiguous

WSJ: Wall Street Journal

XAI: Explainable Artificial Intelligence

1. INTRODUCTION

This dissertation is about the audit profession and artificial intelligence. There is a lot of academic research concerning professions. ELIOT FREIDSON (1988), DAVID SCIULLI (2009), ANDREW ABBOTT (1988) and DAVID MAISTER (1993) are leading theorists on the topic. A research gap related to the four key topics of this dissertation was identified: the audit profession in Quebec, artificial intelligence, knowledge and learning.

The audit profession is of great economic significance. The Big Four auditing firms have a combined annual global revenue of more than $148 billion in 2018 (RAPOPORT, 2018). This amount is greater than the GDP of the sixtieth-richest country in the world. Millions of people, pension funds, investment firms, etc, are relying on high-quality audits to help allocate their savings, their pensions and their livelihoods, so it is critical that the auditor's job is of the greatest quality. There are three important reasons users of financial statements demand for audited financial statements. First, complexity. A company's transactions can be complicated and very difficult for readers to understand. Second, remoteness. Users of financial statements are usually separated from a company's accounting records by distance and time, as well as by lack of expertise. Lastly, consequences. Financial decisions are important to the state of investors' and other users' wealth. Decisions can involve large dollar amounts and massive efforts. The consequences are so important that reliable information, obtained through financial reports, and audited by auditors, is an absolute necessity.

Audits cannot be effective unless they are performed ethically. The essence of information risk is the possibility the reporting will be done unethically (for example, to conceal fraud or provide a deceptive and misleading impression of the financial performance and condition of a company). The essential role and responsibility of an auditor is to establish and communicate assurance to users that financial statements are fairly presented, implying that unethical reporting has not occurred. Professional practice in a three-party accountability situation like the auditor gives rise to many conflicts and dilemmas. It is usually fairly easy to do the right thing, but it is often very difficult to know what the right thing is. Personal bias can arise for anyone, including auditors, and are, by definition, hard to see from the inside.

The audit environment in Canada (and worldwide) has undergone profound changes as a result of corporate failures such as Enron and WorldCom, starting in 2001. Until 2002, the profession in

Canada was largely self-regulated: the profession itself established the rules governing audit practice and monitored compliance with them. The reliance on self-regulation changed with the perceived failure of the profession to detect the problems leading to the corporate scandals of 2001/02. The crucial role of auditing in well-functioning capital markets became clear as never before. This process of rapid change continued with the economic crisis of 2008/09 as the integrity of capital markets was being questioned all over the world.

After Enron's bankruptcy, questions were raised about the effectiveness of its auditors, Arthur Andersen, since there was no official indication of serious problems at Enron until mid-October 2001, when it had to restate previously reported earnings. Joe Berardino, managing partner of Arthur Andersen, tried to explain the apparent audit failure by attributing the problems to the vagueness of accounting standards and the complexity Enron's financial statements. The real problem, however, seems to have been that Arthur Andersen lacked independence because it was auditing its own work.

One may expect that the audit profession would have learned and improved over time. Unfortunately, history repeated itself again. In July 2020, the Financial Reporting Council (FRC) in the UK released its latest audit inspection results at the seven largest UK accounting firms. The FRC's Audit Quality Review (AQR) reviewed 88 audits conducted by KPMG, Deloitte, PwC, EY, Grant Thornton, BDO and Mazars and concluded only two thirds of the audits were of a good standard or required limited improvement (FINANCIAL REPORTING COUNCIL, 2020). Deficiencies included auditors becoming too cozy with clients, disorganized audit paperwork, and a failure by auditors to collect sufficient evidence to support their opinions on a company's books (FINANCIAL REPORTING COUNCIL, 2020). 2020 seems to be another year showing poor records of auditors in uncovering apparent fraud highlighted by scandals such as German payments processing company Wirecard and at FTSE 100 medical group NMC Health.

1.1. Problem Statement

Technology lies at the core of the changes that the audit profession is currently facing. Artificial intelligence (or what is called in the computer science field – intelligent agent) is one of the technologies that will impact the audit profession. As artificial intelligence (AI) will gradually be integrated in the audit profession, this will impact how an audit is performed, the knowledge required to practice the audit profession and the learning requirement. A reasonable assumption is that, as

auditors leverage increasingly intelligent agents to conduct their financial audit work, it should increase the quality of an audit.

The question of whether human-based activity can be executed by an intelligent agent has primarily been addressed using the distinction between routine and non-routine tasks. The effects of automation or computerization of jobs have been studied intensively in economics since the publication of the seminal paper by AUTOR, LEVY & MUNDANE in 2003 (ALM). Since the ALM study, a large and growing strain of literature on task automation has been published. Many of these studies, however, tend to simplify the feasibility of the automation of a task. There are many reasons for this, but two important ones are worth mentioning. First, the methodology used by the authors of these highly cited studies take a high-level view. Second, there is a lack of a common understanding of the term artificial intelligence (AI). An improved task model to assess the potential contribution of intelligent agents in the audit profession is a research gap.

The most fundamental question when conducting research work in the field of artificial intelligence is defining the word intelligence. Intelligence is a concept that we use in our daily lives that seems to have a concrete meaning. We say that our child who received 99% on his calculus test is very intelligent. Although this intuitive notion of intelligence presents us with no difficulties, scientists have not been able to come up with a generally accepted definition of intelligence. "For hundreds of years we have tried to understand and define intelligence and still, we have no agreement on what intelligence is" (TEGMARK, 2017, p. 49). Since there is no generally accepted definition of intelligence, there are many competing ones, including the capacity of logic, understanding, planning, emotional, knowledge, self-awareness, creativity and problem solving. In the context of AI research, LEGG & HUTTER (2007, p. 12) summarized no fewer than 70 definitions on intelligence from the literature into a single statement: "Intelligence measures an agent's ability to achieve goals in a wide range of environments." The lack of a satisfying definition of intelligence is a testament to the immaturity of the research field in AI. If the only success of AI so far has been in developing narrow, task-specific intelligent agents, it is perhaps because only within a very narrow and grounded context have scientists been able to define the goal to be sufficiently precise, and to measure progress in an actionable way. Leveraging on prior work and research, the concept of intelligence will be further analyzed in the context of the audit task execution.

To practise auditing in Quebec, professional auditors must comply with practice standards, which cover the know-what, know-how, know-why, and know-that type of knowledge. Practice standards are general guides for the quality of professional work. Practice standards can be grouped in two broad families: the fitness to practise (know-how) and the technical standards (know-what, know-why, know-that). These standards are part of what we call Canadian Auditing Standards (CAS). Practice standards require that auditors have relevant and sufficient evidence to reach a conclusion on the financial statements prepared by a firm's management. Relevant and sufficient evidence imply that auditors must understand and document appropriately the outcome or recommendations of an intelligent agent. Explainability (known as explainable AI) can refrain or slow down the adoption of intelligent agents to perform audit tasks.

After reviewing the literature, a research gap was identified in the four key topics of this dissertation: the audit profession in Quebec, artificial intelligence, knowledge and learning. In today's digital world, it is obvious that the effectiveness of an audit critically depends on the ability to process the enormous amount of information that is constantly arriving digitally. Auditors need to be well informed about their clients not just at the time to start an audit, but on a continuous basis. The rate of arrival of new information far outstrip the capacity of audit firms and their staff to process such information. With such a diversity and speed of incoming new information, the audit profession in Quebec has to research how it can leverage intelligent agents. To the best of my knowledge, at present, there is a lack of comprehensive conceptual frameworks to assess the feasibility to leverage intelligent agents to accomplish a specific audit task. The results of this study are expected to give new insights to the audit profession in Quebec.

1.2. Significance of the Study

The purpose of this research is to assess the contribution of intelligent agents to execute a specific audit task: the identification of fraud risk factors. The focus of most audit academic technology-related research has been on point solutions and fairly rudimentary use of technology such as email and Microsoft Office, and on specific aspects of the audit function, such as checklist by account and scheduling (JANVRIN et al., 2008). The primary purpose of most commercial software has been to assist in audit management, and not in automating specific audit tasks. Today, financial audits are still largely a manual endeavor. With the advances in business process automation and big data, there is an opportunity to leverage intelligent agents to accomplish specific audit tasks.

The significance of this study is the contribution to the existing literature on task automation or task computerization. Considering the absence of a comprehensive conceptual framework to assess the feasibility to rely on intelligent agents to perform an audit task, this research offers a Task Formula Framework. The Framework was inspired by my audit practical expertise, my professional work in cognitive computing with IBM and my research work during my PhD studies. In 1971, the American Accounting Association Committee on Basic Auditing Concepts prepared a comprehensive definition on auditing as follows: "Auditing is a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between the assertions and established or suitable criteria and communicating the results to interested users" (AMERICAN ACCOUNTING ASSOCIATION, COMMITTEE ON BASIC AUDITING CONCEPTS, 1973).

The definition contains important concepts that this study will apply specifically to financial audit (I refer to financial audit because there are other types of audit). An audit is a systematic process. Auditors execute action that is purposeful, logical, and based on the discipline of a structured approach to decision-making. The audit process involves obtaining evidence, which this study will label as data. Based on computer science terminology, auditors will have access to either structured data (numbers) or unstructured data (mainly text and images). Another important concept in this definition is to ascertain which require the auditor to use their judgment. Finally, due to the volume of transactions in a company, auditors cannot audit every single transaction. Statistical sampling is used, which involved extrapolation and prediction to make inference. Based on the aforementioned definition, I submit the following Audit Task Formula for this dissertation:

$$\text{Auditor's job} = f\ (\text{tasks})$$

$$\text{Task} = f\ (\text{data} + \text{prediction} + \text{judgment} + \text{action})$$

In addition to the Audit Task Formula, I developed the Audit Task Complexity Framework which will contribute to assess the feasibility and the complexity to leverage intelligent agents to accomplish a specific audit task. The ability for an auditor to leverage intelligent agents to perform a specific audit task is complexified by the fact that the auditor has to understand and document properly the outcome of an intelligent agent's work to comply with the Canadian Auditing Standards. Explainability of the outcome of an intelligent agent can refrain the audit profession to leverage intelligent agents. While

the field of explainable machine learning expanded in recent years, much of this work does not take real-world needs into account. A majority of proposed methods use benchmark machine learning problems with generic explainability goals without clear use-case or intended to the end-users or the recipient's community of the explanation. Leveraging on prior work, I developed two important conceptual frameworks in the field of explainable AI for the audit profession in Quebec: The Framework to Reach Artificial Intelligence Explainability (XAI) and the Artificial Intelligence Explainability Recipients' Framework.

To assess the potential contribution of artificial intelligence in the audit profession in Quebec, we must consider the current state of artificial intelligence. This research will focus on machine learning. Technically, machine learning is a subfield of artificial intelligence. At its core and for the purpose of this dissertation, machine learning is about prediction. Prediction is the process of filling missing information. Prediction takes information you have, often called data, and uses it to generate information you don't have. Better data, models, and computers are at the core of progress in prediction. Historically, the core method for predicting churn was statistical techniques such a regression. A regression finds a prediction based on the average of what occurred in the past. Before machine learning, multivariate regression provided an efficient way to condition on multiple things, without the need to calculate dozens of conditional averages. Regression minimizes mistakes on average and punishes large errors more than small ones. It is a powerful method, especially with relatively small data sets and a good sense of what will be useful to predict. Machine learning brings predictions to a significantly higher level. Although explaining the differences between regression/statistical models and machine learning could be the subject of an entire book, it is important to highlight that machine learning is based on a number of earlier building blocks, starting with classical statistics. Statistical inference does form an important foundation for the current implementations of artificial intelligence. But it's important to recognize that classical statistical techniques were developed between the 18th and early 20th centuries for much smaller data sets than the ones we now have at our disposal. Machine learning is unconstrained by the pre-set assumptions of statistics. As a result, it can yield insights that humans do not see on their own and make predictions with ever-higher degrees of accuracy.

This research is exploratory and descriptive in nature as it examines the potential impacts of artificial intelligence to practice the audit profession in Quebec. Findings of this study along with previous

literature available in the area explored in this research is expected to give insights to the audit profession regulatory body in Quebec, computer scientists, audit professionals and universities.

1.3. Objectives of the Study

1. To conceptualize and test four conceptual frameworks to assess the contribution of intelligent agents on the identification of fraud risk factors.
2. To assess the impacts of intelligent agents on the audit profession.
3. To figure out the complexity surrounding explainable artificial intelligence.

1.4. Research Questions and Hypotheses

The emergence of intelligent agents poses a new set of opportunities for the audit profession in Quebec, as well as new challenges. The tasks that can be done by intelligent agents are much broader in scope than previous generations of technology have made possible. The expended scope will change the value employers place on tasks, and the types of skills most in demand. As a result of that, the research questions and the hypotheses hereafter are based on my literature review of previous AI research. Three questions will be addressed:

1. What is the main ethical consideration CPA Quebec should analyze and understand as artificial intelligence will penetrate the audit profession in Quebec?
2. What are, or could be, the impacts of artificial intelligence on both the content of the curriculum to access the chartered professional accountant profession in Quebec and the reskilling requirements?
3. What could be the role of the CPA Quebec ecosystem (government, professional order, universities and firms) in learning in the age of artificial intelligence?

Based upon the stated research questions and objectives, the hypotheses are:

1. Intelligent agents are not a substitute for the audit profession in Quebec and cannot result in a massive employment loss.
2. Intelligent agents cannot assume creative cognitive tasks.
3. The regional audit ecosystem is playing a key role in collective learning and developing ethical and moral regulations for the future of the audit profession in Quebec.

2. LITERATURE REVIEW

When we face a problem for which we have no knowledge, we rely on the expertise and know-how of a professional. Knowledge lies at the heart of professional work. This is how we divide up knowledge in society, so none of us needs to be omniscient. The purpose of an audit is to enhance the degree of confidence on intended users of the financial statements. Those users don't have the knowledge of an auditor. Users of financial statements trust that CPAs in Quebec, whose behaviours are shaped by formal constraints (CPA designation) and informal constrains (e.g., ethics – professional norms of conduct), will not take advantage of what they know and users do not. Membership in Quebec CPA Order is a sort of institutional kitemark, a signal that their members' behaviour is trustworthy, and their insight and guidance is safe to act upon. More formally, under the law in Quebec (The Code of Profession), CPAs owe a fiduciary duty to the shareholders of a corporation, stated differently, an obligation of good faith. It is this strong sense of trust that has been extended and popularized in the phrase of trusted advisor used by the Big Four and other audit firms. It denotes a wealth of knowledge and experience of business issues.

To assume their fiduciary duty, auditors need knowledge. Since knowledge lies at the foundation of this dissertation, it is important to clarify what I mean by knowledge. Knowledge has a number of important characteristics, of which four are relevant to mention. First, knowledge is non-rival. When leveraging on knowledge to solve an audit problem, it does not leave less knowledge for other auditors. Second, knowledge is non-excludable. This means that it can be difficult to prevent other persons from using it. When an auditor issues a report that is public, their knowledge is passed on. It is difficult for the auditor to stop a reader from using the knowledge. Third, the more we tend to use our knowledge, the more valuable it becomes, not less. Finally, knowledge can be digitized. This means that we can convert (some) knowledge into a digital form.

There are various categories of knowledge (MAKÓ & MALOUIN, 2019): know-who (information about who knows most about a given subject within an organization); know-what (substantive technical knowledge, and ideas as well); know-how (procedural knowledge about how to go about some activity); know-where (knowledge of where to go for help, guidance, and expertise on any given topic), know-why (explanations of the rationale behind ideas, activities, processes, and services); know-when (insight into when best to take action or refrain from acting) and know-that (substantive knowledge).

2.1. Trust

What is expected from an auditor is more than the formal knowledge published in a textbook. First, users of the financial statements expect CPAs not just to have substantive knowledge (know-that) at their fingertips but also to have appropriate know-how at their disposal. Sometimes this know-how is tacit; that is, it is not consciously invoked and has not been formally articulated. Often it is procedural and informal. Frequently, it is based on judgment or intuition. Second, users of the financial statements want the knowledge and the know-how of the CPAs to be deep and long-established. They want them to be expert, not just knowledgeable. They also want reassurance that their expertise has been repeatedly applied in the past with considerable success. This track record in the field distinguishes the practitioner from the scholar. Third, there is an applied dimension that requires the CPAs to have the necessary skills, techniques, and methods to apply the expertise and experience effectively. This complex combination of formal knowledge, experience and skills is what I would call practical expertise. This is the foundation of trust in the audit profession.

Since trust lies at the foundation of the audit profession, I will clarify what it means for the purpose of this research. The literature on trust is rich. The importance of trust has been cited in areas such as communication (GIFFIN, 1967), management by objectives (SCOTT, 1980), performance appraisal (CUMMINGS, 1983), leadership (ATWATER, 1988), labour-management relations (TAYLOR, 1989), game theory (MILGROM & ROBERTS, 1992), implementation of a self-managed work team (LAWLER, 1992), and negotiation (BAZERMAN, 1994). Since Arrow remarked that "it can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence" (ARROW, 1972, p. 357), economists have started paying more attention to the effect of trust on economic activity and development. The number of cases where audit firms failed to conduct their audit work in accordance with the Canadian Auditing Standards the past few years (what we can call the post-Enron era) has revived the debate on trust – can we trust the profession? The number of times we can read in a month about a white-collar crime committed in a corporation has revived the debate on trust. The progress made the past decade in artificial intelligence has also revived the debate on trust. When we zoom-out on the progress made in artificial intelligence, we can understand why trust is becoming fundamental: we are gradually shifting the trust in machine from doing something it was specifically dedicated to do, toward a trust in machine for deciding for us what to do and when to do it.

After analyzing many definitions of trust, I realized that it is a confused notion. As the business scholar LARUE TONE HOSMER rightly observes, "there appears to be widespread agreement on the importance of trust in human conduct, but…an equally widespread lack of agreement on a suitable definition on the construct" (HOSMER, 1995, p. 379-380). Such disagreement could be explained by a focus on the search for a generally accepted definition of trust while what we may first understand is why a person might be trustworthy in a particular relationship and context.

"There are more academic papers on the definition of trust than on any other sociological concepts" (BOTSMAN, 2017, p.17) so this research doesn't plan to come up with a generally accepted definition of trust. Based on academic papers on trust and the definition of trust, trust can be viewed as an attitude we have about those who are trustworthy, and that trustworthiness is a property, not an attitude (MCLEOD, 2015). As a result of this interpretation, the need for trust [an attitude…] occurs at two different levels for the purpose of this study.

The first level is at the institutional level: trusting that a firm's management will assume its fiduciary duty, trusting that management prepare and disclose reliable financial statements or trusting that the auditors issue a reliable audit opinion on these financial statements. This type of trust can be labelled as general trust (HARDIN, 2002). It is the trust that we attach to an identifiable but anonymous group of persons. General trust has been subject to many surveys by organizations over the years. For more than fifteen years, the global communication firm Edelman has conducted global surveys of trust in business, government, and the media – the annual Trust Barometer. Its yearly report, released at the World Economic Forum annual meeting in Davos, offers a detailed snapshot of societal [institutional] trust patterns. The headline for the 2019 result reported that the last decade had seen a loss of faith in traditional authority figures and institutions (EDELMAN, 2019). There are several reasons why institutional trust is eroding according to the Trust Barometer; an important one echoes WOLFGANG STREECK's argument in his seminal book *How Will Capitalist End* - corruption. STREECK uses the word corruption in a broad sense, "beyond its definition in criminal law, it means the gross violation of legal rules and the systematic betrayal of trust and moral expectations in pursuit of competitive success and personal or institutional enrichment, as elicited by rapidly growing opportunities for huge material gain in and around today's political economy" (STREECK, 2017, p. 30). Two influential, best-selling books published around the turn of the millennium, ROBERT PUTMAN's *Bowling Alone* (2001) and FRANCIS FUKUYAMA's *Trust* (1995), warned of the fraying of societal and institutional

trust and FUKUYAMA argues that high-trust societies outperform low-trust ones. Business scholars find empirically that companies where trust is high perform better (MAYER et al., 1995). Trust is more than an achievement. It has consequences. Trust shapes interactions. "Trust and be trusted" has economic impact. If we cannot trust other people, we will avoid interacting with them, which will make it hard to build relations, conduct business activities, innovate, etc.

Nobel prize-winning economist RONALD COASE's theory of the firm can be understood as a response to the limitations of generalized trust (COASE, 1937). Firms impose structure, delegation of authority and internal controls because otherwise they cannot trust their executives and their employees to behave reliably. When we trust [generalize], we [voluntarily] reduce two types of friction: information friction (imperfect information, inaccessible information or information risks) and interaction friction (transaction costs, degree of separation and inaccessible marketplaces). As Jack Ma, founder of Alibaba, famously said, "when you trust, everything is simple. If you don't trust, things get complicated" (CNBC, 2014).

The second level of trust is based on relationships or personalized trust. The more we interact with a person over time, the more we get to know them, the more confident we are about how they will behave. Some terms are used synonymously with personalized trust. Trust can be confused with cooperation. There are two schools of thought on the link between trust and cooperation. Scholars such as political scientist ROBERT PUTMAN argue that trust is required to produce cooperation (PUTMAN, 2001). HARDIN argues that cooperation is a general goal and there are many ways to achieve it, some of which do not depend on trust (HARDIN, 2002, p. 11). The prisoner dilemma experiment demonstrates that you can decide to cooperate [by obligation, for certain motives] without trusting the other person (KEE & KNOX, 1970). Relationship between trust and confidence is also obscure. COOK & WALL defined trust as "the extent to which one is willing to ascribe good intentions to and have confidence in the words and actions of other people" (COOK & WALL, 1980, p. 39). Other scholars have not clearly drawn the line between the two concepts (e.g., COLEMAN, 1990; JONES et al., 1975).

Some terms have been used synonymously for generalized trust and personalized trust. Uncertainty, in the context of trust, puts a party in a vulnerable position, forcing the person to predict and take a level of risk. There are many variables that can cause uncertainty. An important one is what economist KENNETH ARROW called information asymmetry (ARROW, 1963). To make decisions in an

uncertain context, we most likely make some predictions. Although the term prediction is related to trust, the association is ambiguous. Both prediction and trust reduce uncertainty (LEWIS & WEIGERT, 1985). If we are highly confident in our prediction ability, we may think that there is no uncertainty, which means we based our decision on data and judgment (the Task Formula variables). This could be a situation that we name in behavioural finance overconfidence. Some scholars have created an overlap between predictability and trust. GABARRO's definition of trust is "the extent to which one person can expect predictability in the other's behaviour in terms of what is normally expected of a person acting in good faith" (GABARRO, 1978, p. 284). The association of the two terms is misleading since someone can be predictable but consistently putting his interest first instead of serving the interest of the trustor.

The association between risk and trust is also ambiguous. Acting on trust involves giving discretion to another to affect one's interest. This action is inherently subject to the risk that the other will abuse the power of discretion. Risk can be seen as the gap between the certain and the uncertain. RACHEL BOTSMAN argues that "trust and risk are like brother and sister" (BOTSMAN, 2017, p. 20). The benefits of trust arise from its ability to stimulate what BOTSMAN describes as a "confident relationship with the unknown" (BOTSMAN, 2017, p. 20). She argues that when we view trust through this lens it starts to explain how it enables us to cope with vulnerability; we make a trust leap.

One may argue that if a person decides to trust another party, it is a rational decision. Assuming that decision is rational, there is some risk that it will prove to be a bad decision. BOTSMAN, like many other researchers, have agreed with DEUTSCH (1980) that risk, or having something invested, is requisite to trust. The need for trust only arises in a risky situation. Other academics recognized the importance of risk to understating trust (e.g.: GIFFIN, 1967; RIKIER, 1974; LUHMANN, 1988), but there is no consensus on the relationship between risk and trust. Is risk a consequence of trust? Is risk a requisite for trust?

Many theorists prescribe greater trust as a necessary antidote for an increasingly litigious and distrustful society (LIEBERMAN, 1981). There is also an opposing view about the need to increase trust. Baroness ONORA O'NEIL, a professor at the University of Cambridge and a crossbench member of the House of Lords, has written extensively about trust and how trust is misplaced. In her TED talk (O'NEIL, 2013), she challenged the conventional, simplistic belief that, as a society, we have lost trust and ought to set about rebuilding it. She argued that we should have more trust in the

trustworthy but not the untrustworthy. This view is consistent with MCLEOD's (2015) definition of trust. Encouraging more general trust simply for the sake of creating a more trusting society is not only meaningless, it's dangerous, Professor O'NEIL argues in her TED talk. For one thing, people are already inclined to want to trust blindly according to her. The Bernie Madoff scandal is a classic case. Why did people trust him? Mostly because Madoff was charming and was in the same social circles as they were. As Professor O'NEIL notes in her TED talk, Madoff is an example of too much trust in the wrong place. She insists that, instead of making decisions about trust, we should be looking at the who, where and why of trustworthiness.

Assessing trustworthiness is an imperfect endeavour. GOOD (1988) suggested that trust is based on expectations of how another person will behave based on that person's current and previous implicit and explicit claims. Similarly, LIEBERMAN (1981) stated that trust in fiduciary relationships is based on a belief in the professional competence and integrity. These authors have suggested that characteristics and actions of the trustee will lead that person to be more or less trusted. Political scientist RUSSELL HARDIN argues that trust is about encapsulated interest, a kind of closed loop of each party's self-interest. He argues that if I trust you [general or personalized trust], it's because I believe that you are going to take my interests seriously – whether it be for friendship, love, money or reputation. Why? "You won't take advantage of me because it benefits you not to do so. You value the continuation of our relationship, and you therefore have your own interests in taking my interest into account", HARDIN writes in *Trust and Trustworthiness* (HARDIN, 2002, p. 6). Many philosophers have debated the concept of trustworthiness and, as with the word trust, there is no generally accepted definition of the terms.

Based on the aforementioned analysis, I will break down trustworthiness into three components:

- Competence: competence comes down to how capable a person is to do something. Does he/she have the skills, knowledge and experience to do a particular role or task. A number of theorists (COOK & WALL, 1980; DEUTSCH, 1980; SITKIN & ROTH, 1993) used the word ability to define a similar construct as O'NEIL.
- Reliability: reliability comes down to a person's consistency in doing what they said they would do for you (the task). Ultimately, it's about you knowing: can I depend on this person?
- Honesty: honesty is about integrity and intentions. What are their interests and motives toward me? Basically, it's whether their intentions are aligned with yours. A number of researchers

have included characteristics similar to honesty as a basis for trust. HOVLAND et al. (1953) described trustworthiness in terms of the trustee's motivation to lie. Others have considered intentions or motives as important to trust (COOK & WALL, 1980; DEUTSCH, 1980; GIFFIN, 1967).

2.2. Bias

The human brain is capable of incredible things, but it's also extremely flawed at times. Science has shown that we tend to make all sorts of mental mistakes that can affect both our thinking and actions. Auditors are not immune from biases. Biases compromise the quality of an audit and worse, they can impact the credibility of the profession and the level of trust that users of the financial statements have in the audit profession.

Because of the subjective nature of accounting and the tight relationships between auditing firms and their clients, even the most honest and meticulous auditors can unintentionally distort the numbers in ways that mask a company's true financial status, thereby misleading investors, regulators, and sometimes management. Indeed, even seemingly egregious accounting scandals, such as Andersen's audits of Enron, may have, at their core, a series of unconsciously biased judgments rather than a deliberate program of criminality.

Bias thrives wherever there is the possibility of interpreting information in different ways. Many accounting decisions and principles such as establishing a proper conversion rate between US dollars and Euros are cut-and-dry; others require interpretations of ambiguous information. Auditors and their clients have considerable leeway, for example, in answering some of the most basic financial questions: What is the useful life of an asset? What are the future economic benefits of an intangible asset? The interpretation and weighting of various types of information are rarely straightforward. As Joseph Berardino, Arthur Andersen's former chief executive, said in an interview to *Frontline*, accounting is not a science where one number, namely earnings per share, is the ultimate number and it's such a precise number that it couldn't be two pennies higher or two pennies lower, accounting is an art (SMITH, 2002). Auditors are exposed to a number of biases and four are important to analyze.

The first one is skewed information processing or confirmation bias. Psychological research (KAHNEMAN, 2011) shows that our desires powerfully influence the way we interpret information,

even when we're trying to be objective and impartial. When we are motivated to reach a particular conclusion, we usually do. That's why most of us think we are better than average drivers, have smarter than average children, and choose stocks or funds that will outperform the market – even if there's clear evidence to the contrary. Without knowing it, we tend to critically scrutinize and then discount facts that contradict the conclusions we want to reach, and we uncritically embrace evidence that supports our positions. Unaware of our skewed information processing, we erroneously conclude that our judgments are free of bias.

The second one is negotiation. Another indication of ambiguity in accounting could be a practice of negotiating about accounting rules. In one study of 93 audit partners working for international accounting firms, GIBBINS et al. (2001) reported that 67% of audit partners commonly negotiated with 50% or more of their clients. These negotiations, for example, might involve the timing of revenue and expenses recognition. Executives are often in a hurry to recognize revenue but prefer to delay recognizing an expense. If there were such a thing as correct timing, these negotiations wouldn't take place.

The third one is opinion shopping. Opinion shopping by clients is an indication of accounting ambiguity. Although there is a fine line to not cross according to the Canadian Auditing Standards, such practice consists of asking different audit firms to interpret specific accounting problems before deciding whom to hire. Because, for some accounting standards, there is flexibility to select an accounting treatment, there is no right conclusion, so different auditing firms can have different opinions.

The last one is audit as a commodity service. Auditors have strong business reasons to remain in clients' good graces and are thus highly motivated to approve their clients' accounts. Under the current system, auditors are hired and fired by the companies they audit, and it is well known that client companies fire accounting firms that deliver unfavourable audit opinions. Even if an accounting firm is large enough to absorb the loss of one client, individual auditors' jobs and careers may depend on success with specific clients. Two decades after the introduction of the Sarbanes-Oxley Act, some accounting firms still treat audits as ways to build relationships that allow them to sell their more lucrative consulting services. In 2019, the UK audit regulator recognized the severity of the problem and, as a result, outlined a plan to break up the Big Four auditing firms.

An audit ultimately endorses or rejects the firm's management's statement about the company's financial information. Self-serving biases become even stronger when people are endorsing others' biased judgments – provided those judgments align with their own biases – than when they are making original judgments themselves. In addition to these structural elements that promote bias, five aspects of human nature can amplify unconscious biases of an auditor.

The first one is proximity. SMITH (1790) identified a fundamental dimension of human nature: our ability to empathize with others is directly related to their proximity to us. The farther away the harm, the more fleeting our emotions. As the saying goes, out of sight, out of mind. According to this theory, people should be more willing to harm strangers than individuals they know, especially when those individuals are paying clients with whom they have ongoing relationships. An auditor who suspects questionable accounting must thus choose, unconsciously perhaps, between potentially harming his client by challenging the company's financial statements or harming shareholders and creditors by failing to object to the possibly skewed numbers. Given this dilemma, auditors may unconsciously lean toward approving the dubious financial statements. This type of bias can grow over time as the audit partner's ties with the CFO and CEO of the company grow over time. The longer an accounting partner serves a particular client, the more biased his judgments will tend to be. The case of GE is a good example. In 2018, after a disastrous year that included a SEC accounting investigation, GE became under fire to dump KPMG as its auditor. Shareholder watchdog groups worried that GE and KPMG had become too cozy during their 109-year-old relationship. Institutional Shareholder Services were urging shareholders not to ratify KPMG as GE's auditor at the company's annual shareholder meeting. This extensive tenure has thrown KPMG's effectiveness and relationship with the company into question (EGAN, 2018).

The second one is temporal gap. In economics, hyperbolic discounting is one of the cornerstones of behavioral economics and it is actively being studied by neuroeconomics researchers (HAMPTON et al., 2017). Basically, the theory says that, given two similar rewards, humans show a preference for one that arrives sooner rather than later. Humans are said to discount the value of the later reward, by a factor that increases with the length of the delay. This explain why people tend to be far more responsive to immediate consequences than delayed ones, especially when the delayed outcomes are uncertain. A good example of that is a financial crime since it lacks instantaneous feedback. The harmful consequences of such crime may follow months, even years, after the initial actions, so it's

easier for the perpetrator to be ignorant of the harm he caused. This bias may explain why auditors can hesitate to issue critical audit reports because of the adverse immediate consequences – damage to the relationship, potential loss of the contract, and possible unemployment. But the costs of a positive report when a negative report is called for – protecting the accounting firm's reputation or avoiding a lawsuit, for example – are likely to be distant and uncertain.

The third one is escalation. It's natural for people to conceal (STOUTHAMER-LOEBER, 1986) or explain away minor indiscretions or oversights, sometimes without even realizing that they're doing it. We all have been late to a meeting and blamed it on traffic while the true reason was bad time management. An auditor's biases may lead him to unknowingly adapt over time to small imperfections in a client's financial practices. Eventually, though, the sum of these small judgments may become large and he may recognize the long-standing bias. But at that point, correcting the bias may require admitting prior errors. Rather than expose the unwitting mistakes, he may decide to conceal the problem. Thus, unconscious bias may evolve into conscious corruption – corruption representing the most visible end of a situation that may have been deteriorating for some time.

The fourth one is the status quo bias. Coined by SAMUELSON & ZECKHAUSER (1988), it is an emotional bias in which professionals like auditors do nothing instead of making a change in the way to work. Auditors, by their conservative nature, are generally more comfortable keeping things the same than with change and thus do not necessarily look for opportunities where changes are beneficial to improve the quality of an audit. If given a situation where one choice is the default choice, people will frequently let the choice stand rather than opting out of it and making another choice.

The last one is the overconfidence bias. It is a bias in which auditors demonstrate unwarranted faith in their own intuitive reasoning, judgments, and/or cognitive abilities (PALLIER et al., 2002). Overconfidence is one example of a miscalibration of subjective probabilities. This overconfidence may be the result of overestimating knowledge levels, abilities, and access to information.

The key to improving audits must be to eliminate incentives that create self-serving biases. The challenge with bias is that, by its very nature, it is typically invisible: you can't review a corporate audit and pick out errors attributable to bias. Often, we can't tell whether an error in auditing is due to bias or corruption. It requires a detailed investigation.

2.3. Ambiguity in Interpreting Irrational Behavior

The audit profession is critical to the proper functioning of capital markets, and if audits are perceived to fail, then the capital markets might do the same. Without effective audits, modern capital markets cannot fulfill their role as efficient economic systems leading to high living standards. A European Commission Green Paper concluded that "auditors, regulators, and corporate governance are key contributors to financial stability and economic growth" (EUROPEAN COMMISSION, 2010, p.3).

Audits are part of a three-party accountability framework. A three-party accountability is a special case of the agency problem of economic theory. Whenever a task is delegated by one party (the principal, i.e., shareholders) to another (the agent, i.e., management), it can create a potential conflict of interest, an agency problem. Agency problem occurs when three conditions are present in an agency relationship: the agent has objectives that are different from those of the principal, the agent has more information than the principal does (information asymmetry) and the contract between the two is incomplete in that not every possible contingency can be anticipated (ARROW, 1984).

In the corporate world, management is the party accountable to the owners. One way that management satisfies its accountability is by preparing financial statements. Financial statements are one way of monitoring how well management is running the company. However, there is a potential problem in that management may bias its statements, making financial statements less credible. The auditor comes in as an outside independent accounting expert to verify the accuracy of financial statements, thereby adding credibility to the statements. The auditor, thus, helps monitor management.

Platform business models like Amazon and Facebook are impacting the three-party accountability framework. Whereas giant industrial-era firms were made possible by supply economies of scale with a focus on short-term profitability to please Wall Street, today's giants are made possible by demand economies of scale – expressed as network effects. A business platform cannot focus only on shareholders' value, it must consider the stakeholder community: users and producers (the network). There is no point to be on Facebook if you are the only user. Positive network effects refer to the ability of a large, well-managed platform community to produce significant value for each party of the platform. Demand economies of scale are driven by efficiencies in social networks, demand aggregation, applications development, and other phenomena that make bigger networks more valuable to their users. Because platform businesses create value using resources that they don't own

or control, they can grow faster than traditional businesses. Equally important, "platform firms benefit from a more patient form of capital" (RAHMAN & THELEN, 2019, p.179).

The business platform model is impacting all industries and was one of the contributing factors to push the influential BUSINESS ROUNDTABLE (BRT), an association of the chief executive officers of nearly 200 of America's most prominent companies, to revisit its formal statement of corporate purpose. In August 2019, the BRT has dropped the shareholder primacy creed that has driven U.S. capitalism for decades, urging companies to consider the environment and workers' wellbeing alongside their pursuit of profits. The new statement is 300 words long (BUSINESS ROUNDTABLE, 2019), and shareholders aren't mentioned until word 250. Before that, the group refers to creating value for customers, investing in employees, fostering diversity and inclusion, dealing fairly and ethically with suppliers, supporting the communities in which we work, and protecting the environment. "Larry Fink, chief executive of BlackRock and a member of the BRT, last year called on businesses to strive to make a positive impact on society in addition to delivering profits. Similar views have been echoed more recently by Ray Dalio, the founder of Bridgewater Associates, and Jamie Dimon, the chief executive of JPMorgan and chairman of the BRT" (HENDERSON & TEMPLE-WEST, 2019). Such a shift from the BRT signals that people are dissatisfied with capitalism. An important question remains: if the party (management) accountable to the owners (shareholders) struggled to assume their fiduciary duty in the past, how can they assume such an important responsibility for a broader group – the stakeholders? Such a statement may reinforce the importance of the role of the auditor in this new three-party accountability framework: management, stakeholders (and not shareholders) and the auditor.

In a three-party accountability environment, one of the key risks that exist is the presence of non-ethical behaviour from management. Irrational behaviour in a three-party accountability framework is not new. A 2018 survey by PwC revealed troubling statistics. Turnover among CEOs at the world's 2,500 largest companies soared to a record high of 17.5% in 2018 – 3 percentage points higher than the 14.5% rate in 2017 and above what has been the norm for the last decade. But the reasons that CEOs were fired in 2018 were different. For the first time in the PwC study's history, more CEOs were dismissed for ethical lapses than for financial performance or broad struggles (PwC, 2018).

The theory of ethics has been the subject of interest to philosophers since the beginning of recorded thoughts. Because philosophers are concerned with the good of all mankind, their discussions have

been concerned with what we may call general ethics rather than ethics of a small group such as the member of a management team. As a result, I cannot rely on philosophers' theories for direct solutions for ethical behaviour for a small group but their work with general ethics is of primary importance to the development of an appropriate concept in any special field. Trying to understand why people commit non-ethical behaviour such as white-collar crime is a very complex undertaking.

The term white-collar crime was first used by EDWIN H. SUTHERLAND in December 1930 during his presidential address in Philadelphia to the American Sociological Society. Since then, there have been constant disputes about what it is. The Dictionary of Criminal Justice Data terminology published by the U.S. Federal Bureau of Justice Statistics, defines white-collar crime as "nonviolent crime for financial gain committed by means of deception by persons whose occupational status is entrepreneurial, professional, or semi-professional and utilizing their special occupational skills and opportunities; also nonviolent crime for financial gain utilizing deception and committed by anyone having special technical and professional knowledge of business and government, irrespective of the person's occupation" (U.S. DEPARTMENT OF JUSTICE, 1981, p. 215).

The debate over just what drives the white-collar criminal has been raging ever since SUTHERLAND's seminal work. On the evening of December 27, 1939, SUTHERLAND took the stage to deliver his presidential address at the fifty-second annual meeting of the American Sociological Society. He was well known by the audience for his popular university textbook *Principles of Criminology*. In SUTHERLAND's view, much of the most serious crime was being committed not by the poor or the delinquent but, instead, by society's most well-known and respected business leaders. Deviance committed by respected business and professional men was simply overlooked because people of high socioeconomic standing were not usually convicted in criminal courts. During his talk, SUTHERLAND even coined a new term for this class of deviance: white-collar crime (SUTHERLAND, 1940). SUTHERLAND's presidential address in 1939 on white-collar criminality was one of the few such addresses that received front-page publicity in the daily newspapers. It was published in the *American Sociological Review* in February 1940 and developed later into the volume on White-Collar Crime.

In December 1944, on the fifth anniversary of his 1939 presidential address, SUTHERLAND planned to give another presentation on white-collar criminality. The address was cancelled due to fuel rationing regulation during wartime, but his content was published in 1945 in the *American*

Sociological Review. SUTHERLAND laid out three impediments that he believed were holding back more aggressive stances toward white-collar misconduct. The first one is the social status of businessmen that seemed to protect them from recrimination. Historically, prosecutors did not want to antagonize business leaders, because of their prominence and respectability in society. Regulators feared the accused. These sentiments shifted over the time. More regulators and politicians sought to define themselves as taking a tougher stance against conduct condemned by their electorates. The second relates to that fact that white-collar offenders, even if convicted, would not be effectively punished. He believed that laws and procedures were designed to prevent placing a negative stigma over executives. Lastly, in SUTHERLAND's eyes, to more vigorous white-collar prosecution was the general indifference of the public toward white-collar offenders (SUTHERLAND, 1945). This has changed over time. Movements like Occupy Wall Street, in which thousands of people took over Zuccotti Park in New York City's Financial District, provide visceral evidence of the change in attitude.

There are several kinds of white-collar crimes. Some are defined in laws, while others are matters of general understanding. Figure 1 summarizes the relevant ones for the purpose this study.

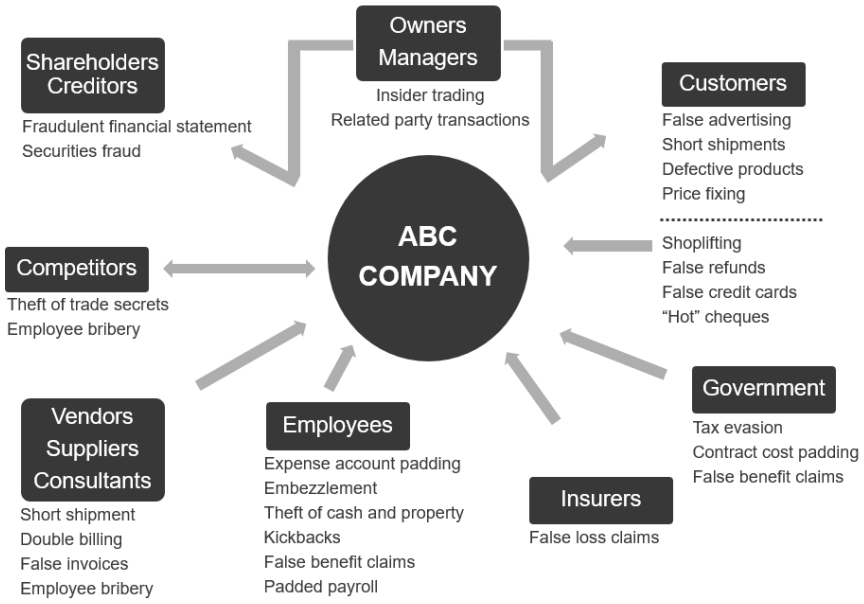

Figure 1. Types of White-Collar Crimes

Source: Author's compilation

Collectively, we can coin them as fraud. Among the various kinds of fraud that organizations might be faced with, occupational fraud is likely the largest and most prevalent threat. Occupational fraud – fraud committed against the organization by its own officers, directors, or employees – constitutes an attack against the organization from within, by the very people who were entrusted to protect its assets and resources. The most recent REPORT TO THE NATION contains some troubling numbers. The total loss caused by white-collar crimes in 2018 is estimated as at least 7.1 billion USD. The median loss for all the case studies is 130 million USD.

Figure 2. Report to the Nation: 2018 Global Study on Occupational Fraud and Abuse

Source: ACFE (2018)

Auditors are concerned with two types of fraud: first, fraudulent financial reporting (fraud involving management making false or misleading claims in financial statements) and second, employees' and management's misappropriation of assets. In response to the growing problems of fraud, and since the fall of Enron, auditors have taken on increased responsibility for detecting fraud. The Canadian Auditing Standards (CAS) set out rigorous requirements relating to fraud in a financial statement audit assignment. In fact, CAS 240 requires auditors to maintain professional skepticism and make no assumptions about management's honesty. CAS 240 paragraph 26, requires auditors to presume there is always a risk of fraudulent revenue recognition, a presumption that is rebuttable by the audit evidence. That is, if the auditors can convince themselves that the risk is appropriately low, then the presumption is rejected. This logic is similar to the burden of proof of concept-in-law and critical thinking. From a critical thinking point of view, the drop in the assumption that management is honest (which took place in 2004 after the fall of Enron) is the most important change in the audit standards.

Auditors assess the risk of fraud through warning signs. To properly assess warning signs, one must understand why people would commit fraud. In the current CPA curriculum, students are not being

taught that important concept. Understanding criminal behaviour is complex. We can't expect CPAs to be criminalists or lawyers, but a basic foundation on understanding why people commit fraud would certainly benefit the auditors in conducting their financial audit assignment. While white-collar crime was recently conceived as elite, upper world, and upper-class offending (COLEMAN, 1989), recent research has shown that white-collar offenders are generally middle to lower-middle class, do not specialize in one form of crime, and also perpetrate street crimes (WEISBURD et al., 2001).

Why do people obey law? In his book *Why People Obey Law*, TOM TYLER studies the two principal types of law-abiding people (TYLER, 1990). According to TYLER, there are two perspectives to answer this question. First, the instrumental perspective which states that people weigh the pros and the cons of compliance with the law and act accordingly. Compliance is associated with the fear of punishment. The second one is the normative perspective, which stipulates that people consider what is just and moral. This implies that when a person believes that compliance is their moral obligation, commitment to the law is voluntary, regardless of fear of punishment. These four elements – pros, cons, just and moral – are the foundation of a number of theories on white-collar crimes.

Some scientifics argue that white-collar crimes can be explained by biological anomalies and that criminal behaviour is not the result of choice (pros/cons analysis), but rather is caused by the physical traits of those who commit the crime. After compiling details of his analysis of thousands of criminals, CESARE LOMBROSO, an Italian doctor, published his findings in his magnum opus *L'Uomo Delinquente*, or *The Criminal Man*. According to LOMBROSO's theory, some people are deeply and fundamentally flawed – they are born criminal. According to LOMBROSO's theory, corporate criminality arises not from an act of mistaken judgment or some situational influence but, rather, from a deviant nature that is innate and simply waiting to exploit an appropriate opportunity (LOMBROSO-FERRERO, 1972). LOMBROSO spent his career measuring the bodies of offenders and concluded that they were marked by a high degree of asymmetry, with such things as sloping foreheads, and other anomalies. Although his theory would be dismissed as pseudoscience today, the theory continues to resonate. As an example, Bernie Madoff's attorney Ira Lee Sorkin conceded that his client "was a deeply flawed individual" at his sentence hearing (THE DAILY BEAST, 2009). LOMBROSO provided a systematic set of causes to identify criminals and, because of his contribution, he is regularly cited as the father of modern criminology.

Biological theories advanced after LOMBROSO published his theory. EARNEST HOOTON, an anthropologist who specialized in the psychological variation between ancient and modern people, came to a similar conclusion as Lombroso. While being a Professor at Harvard, HOOTON completed a major research project to tease out possible physical differences between criminals and non-criminals. HOOTON's study became one of the largest and most ambitious investigations of criminals ever undertaken. More than a decade after beginning the project, HOOTON concluded not only there were differences between criminals and non-criminals but he could "identify physical differences between criminals convicted of different offenses" (HOOTON, 1939, p. 98).

In their book *The Bell Curve: Intelligence and Class Structure in American Life*, psychologist RICHARD J. HERRNSTEIN and political scientist CHARLES MURRAY argue that human intelligence is substantially influenced by both inherited and environmental factors and that it is a better predictor of many personal outcomes, including financial income, job performance and involvement in crime than are an individual's parental socioeconomic status. Scholars and non-scholars in the field highly criticized the book (BROWNE, 1994).

The question of why some people have less control than others has been subject to numerous researches (MISCHEL et al. 1972). While different from external, physically discernable differences proposed by LOMBROSO, HOOTON, HERRNSTEIN AND MURRAY, the idea that crime is caused by insufficient self-control similarly regards managerial misconduct as arising from a biological deficiency. The premise of this theory is that people with lower self-control have greater difficulty resisting temptation and restraining reckless behaviour, and eventually some of this rash and opportunistic behaviour is likely to end up as criminal conduct. The recent story about Andrew Pearse, an investment banker at Credit Suisse, who helped fuel a $2 billion debt fraud in Mozambique is a good example. Mr. Pearse received his first bribe in 2013. The rationale for accepting those bribes was to leave Credit Suisse and start his own financial boutique with his lover, Detelina Subeva, a colleague of his at the bank. In court, he explained "that love and ambition drove him to take $45 million in bribes" (PATRICK & WIRZ, 2019). In some of the most highly cited research by criminologists, MICHAEL GOTTFREDSON and TRAVIS HIRSCHI argue that people with low self-control are more likely to engage in deviant behaviour throughout their entire lives. According to this theory, executives with impaired self-control should not just engage in corporate misconduct, but their low inhibitions should tempt them to engage in all sorts of non-ethical behaviour. AIYESHA DEY,

an associate professor at Harvard University, examined correlation between an overall self-control and a disregard for rules. In one of her research papers, she looked at whether executives' personal legal records – everything from traffic tickets to driving under the influence and assault – had any relation to their tendency to execute trades on the basis of confidential inside information. Using U.S. federal and state crime databases, criminal background checks, and private investigators, DEY identified firms that had simultaneously employed at least one executive with a record and at least one without a record during the period from 1986 to 2017. This yielded a sample of nearly 1,500 executives, including 503 CEOs. Examining executive trades of company stock, DEY and her co-authors found that those trades were more profitable for executives with a record than for others, suggesting that the former had made use of privileged information. The effect was greatest among executives with multiple offenses and those with serious violations (anything worse than a traffic ticket) (DEY et al., 2019). In an earlier study, DEY and her co-authors identified 109 firms that had submitted fraudulent financial statements to the SEC. Comparing those companies' CEOs with the heads of comparable firms that had clean reporting slates, they found that far more leaders in the fraud group had a legal record: 20.2%, versus just 4.6% of those in the control group (DEY et al., 2015).

Some aspects of this theory provides interesting insight about why one could engage in white-collar crimes, but it is very difficult to assess if a person is lacking self-control in committing that type of criminal offense, since white-collar offenses require considerable planning and can take place over a long period of time, as the Madoff case demonstrates. The Bernie Madoff scandal emerged in late 2008 amidst a sharp selloff in the stock market and wide-ranging financial crisis. Madoff, after a long and respected career on Wall Street, turned himself into authorities and allegedly stated that his investment business had been nothing more than an elaborate hoax. Madoff spent his life planning his strategy. He targeted primarily fellow Jewish investors, attracting clients from elite circles in New York City, Palm Beach, Hollywood, Europe and Latin America (MARKOPOLOS, 2010). ADRIAN RAINE, a criminologist at the University of Pennsylvania, believed "that Lombroso was on the path towards a sublime truth" (RAIN, 2013, p.13) and that white-collar criminals have control. RAINES and his co-authors conducted a comprehensive neurobiological study and found that white-collar criminals had significantly better executive functioning, increased electro dermal orienting, increased arousal, and increased cortical gray matter thickness in the ventromedial prefrontal cortex, inferior frontal gyrus, somatosensory cortex, and the temporal-parietal junction compared to controls (RAINE et al., 2012). One brain region suggested that the white-collar criminals had greater cognitive control,

which is useful for creating and acting on goals. The authors concluded that results, while initial, constitute the first findings on neurobiological characteristics of white-collar criminals. It is hypothesized that white-collar criminals have information-processing and brain superiorities that give them an advantage in perpetrating criminal offenses in occupational settings (RAINE et al., 2012). Rather than conceiving these neurobiological characteristics as deficits, the authors argue that they more likely aid them in accommodating the structure, function, and culture of an organization, allowing white-collar offenders to engage in elaborate calculations that consider a wide range of factors (RAINE et al., 2012). Given the evidence that white-collar crime is both a reaction and adaptation to a range of organizational and structural variables (SCHLEGEL & WEISBURD, 1992), and given that there is a rationality to this kind of offending that requires careful calculations, the authors argue that white-collar offenders have superior executive functioning and attentional functioning compared to controls (RAINE et al., 2012). This seems to align with an interesting statistic from the 2018 *Report to the Nation*. The median duration for fraud in the financial statements is 24 months (ACFE, 2018, p. 15). Such type of fraud requires careful planning and understanding of internal control.

A number of important non-biological theories have been developed to try to explain why people commit crime. No single theories can explain it all; humans are very complex creatures. But these theories have one point in common: human beings, unlike inanimate objects, think of themselves. It is very difficult to predict the direction in which a human being might decide to proceed when confronted with a choice.

Circumstances can lead professionals to engage in crime. GABRIEL TARDE, a French sociologist, criminologist and social psychologist in the nineteenth century, took an interest in criminology and the psychological basis of criminal behaviour while working as a magistrate in public service. His theory was based on circumstances and he was critical of the born criminal theory developed by LOMBROSO. TARDE argued that criminal behaviour was learned by observing others. The more frequent and intense the contact with those engaged in criminal conduct, TARDE believed, the greater the likelihood for imitating that behaviour oneself (TARDE, 1962). His theory came to be known as differential association because it suggested that a person's propensity to become a criminal depended on how much one associates with other criminals. Such theory seems to fit the Enron case where both the CEO and the CFO created one of the most important corporate frauds in the U.S. history. This

theory is one of the best-known U.S. theories to explain white-collar crime, but it has been widely criticized on the grounds that it is just about impossible to test it.

Social control theory, unlike the theory of differential association, offers a number of testable propositions. TRAVIS HIRSCHI first articulated the theory in his book *Causes of Delinquency*. Such propositions take the form of if-then statements. If something exists or is done, then it foretells that something will follow. Such formulations allow for experimental testing and rebuttal. Social control theory takes its cue from a classic of sociology, EMILE DURKHEIM's *Suicide*, in which the French theoretician wrote "the more weakened the groups to which [the individual belongs], the less he depends on them, the more he consequently depends on himself and recognizes no other rules of conduct than what are founded on his private interest" (DURKHEIM, 1951, p. 209).

Essentially, social control theory argues that the institutions of the social system train and press those with whom they are in contact into patterns of conformity. As an example, schools educate for adjustment in society. The theory rests on the thesis that, to the extent that a person fails to become attached to the variety of control agencies of the society, his chances of violating the law are increased.

Picking up on TARDE's theory, DONALD CRESSEY, an American penologist, sociologist, and criminologist made innovative contributions to the study of criminology, the sociology of criminal law and white-collar crime. CRESSEY was a graduate student from EDWIN SUTHERLAND, the sociologist who coined the term "white-collar crime". Working on his PhD in criminology, CRESSEY focused his attention on one specific group of criminals, the embezzlers, the ones that have most likely been raised in normal or good conditions. CRESSEY was intrigued by embezzlers, whom he called trust violators. He was interested in the circumstances that led them to be overcome by temptation. For that reason, he excluded from his research those employees who took their jobs for the purpose of stealing. CRESSEY interviewed more than a hundred inmates convicted of embezzlement. Upon completion of his interviews, he developed what still remains the classic model for occupational fraud offences. His research was published in a book named *Other People's Money: A Study in the Social Psychology of Embezzlement*. CRESSEY's final hypothesis was "Trusted peoples become trust violators when they conceive of themselves as having a financial problem which is non-sharable, are aware this problem can be secretly resolved by violation of the position of financial trust, and are able to apply to their own conduct in that situation verbalization which enable them to adjust their

conceptions of themselves as trusted persons with their conceptions of themselves as users of the entrusted funds or property" (CRESSEY, 1973, p. 30).

Over the years, this hypothesis has become known as "the fraud triangle". The probability of occupational fraud is a function of three elements: pressure, opportunity and rationalization. Pressure can take different forms. Economic motive is by far the more common reason, but other pressures can exist, such as egocentric motivations (people committing fraud to achieve more personal prestige) or psychotic motivation (people committing fraud simply for the sake of committing a fraud, which is not frequent). A fraud opportunity is an open door for solving the unshareable problem by violating a trust. The violation maybe a circumvention of internal control policies and procedures, or it may be taking advantage of an absence or lapse of control procedures. Everyone has some degree of trust placed on them in their job, and the higher the position in an organization, the greater the degree of trust, hence the greater the opportunity for larger frauds. Most people in a civilized society can be assumed to know the difference between right and wrong. While unimpeachable integrity is the ability to act according to the highest moral and ethical values all the time, lapses and occasional lack of integrity permit pressure and opportunity to take form as a fraud. People normally do not make deliberate decisions to lack integrity, but they sometimes do find ways to rationalize their act, describing it to themselves in a way that make it acceptable to their self-image. Some form of rationalization would be "I need it more than they do" – the Robin Hood theory. The importance of these three factors have been acknowledged by the General Accepted Auditing Standards in Canada.

TARDE, HIRSCHI, SUTHERLAND and CRESSY focused on circumstances to develop their theory. The decision to commit a white-collar crime is more than just a choice between committing an illicit act or not. It is also a decision about violating a deeply held notion of morally acceptable behaviour. How we develop our moral sense from right or wrong has been the focus of LAURENCE KOHLBERG, an American psychologist best known for his theory of the stages of moral development. KOHLBERG served as a professor in the Psychology Department at the University of Chicago and at the Graduate School of Education at Harvard University. Even though it was considered unusual in his era, he decided to study the topic of moral judgment, extending Jean Piaget and a fascination with children's reactions to moral dilemmas. KOHLBERG hypothesized that individuals advance through six stages of moral development (KOHLBERG, 1958). These stages explain the development of moral reasoning. Individuals begin by evaluating actions on the basis of

avoiding punishment and satisfying their own needs. Later, individuals make judgments based on societal expectations. By the sixth and highest stage of moral development, individuals would respect the rights of others by appealing to the abstract of universal justice. KOHLBERG argues that each successive stage of moral development is superior to the stage before it because it implies more sophisticated distinctions about respecting individual rights. By placing the individual's capacity to reason at the center of moral decision-making, KOHLBERG reaffirmed JOHN DEWEY's idea that development should be the aim of education (KOHLBERG & MAYER, 1972) and that morality could be effectively taught. A fundamental question arises from KOHLBERG's theory – what really makes something a moral decision? Executives make decisions constantly, such as laying off people or raising the price of medication (which impacts peoples' ability to afford them). In fact, most decisions made by executives impact people. The decision-making process involves trade-offs that are moral, economic, legal, etc. One can argue that illegal decisions are being made like any other decision.

Decision-making process has been a field of study of many economists. Within economics, the concept of utility is used to model worth or value. The term was introduced initially as a measure of pleasure or satisfaction within the theory of utilitarianism by moral philosophers such as JEREMY BENTHAM and JOHN STUART MILL. GARY BECKER, an economics professor at Columbia University in the early 1960s, sought a scientific understanding of human behaviour. BECKER argued that many different types of human behaviour can be seen as rational and utility maximizing. His approach included altruistic behaviour of human behaviour by defining individuals' utility appropriately. According to BECKER's theory, individuals pursue activities that are utility increasing and avoid those that are utility decreasing. Viewed from this angle, a criminal is not a criminal by nature but, rather, someone who perceives the cost of illegal decisions differently from other people. From that perspective, the way a criminal may make decisions would be to compare the utility of committing a crime against the utility gained by committing the same amount of time and resources to other legitimate activities. In this cost/benefit analysis, BECKER argues that morality comes into the equation as an additional cost. BECKER's work (1968) *Crime and Punishment: An Economic Approach* became one of the most important papers in economics.

An organization's culture influences behaviour. There are many ways to define organization culture, but typically definitions emphasize a set of values that are shared within a group: what employees believe in and which values influence their behaviour. EDGAR SCHEIN defines organizational

culture as the "basic assumptions and beliefs that are shared by members of an organization that operate unconsciously and define (in basic taken-for-granted fashion) an organization's view of itself and its environment" (SCHEIN, 2004, p. 6). Taken-for-granted can be seen as the ways we do things around here (DEAL & KENNEDY, 1982). An organization's culture can be conceived as consisting of different layers. The four layers proposed by SCHEIN are:

1. Values: May be easy to identify in terms of those formally stated by an organization since they are explicit;
2. Beliefs: Can typically be discerned in how people talk about issues the organization face;
3. Behaviours: Are the day-to-day ways in which an organization operates, they are what can be seen by people both inside and often outside the organization; and
4. Paradigms: Set of taken-for-granted assumptions.

Culture imbues all aspects of workplace life – the communication type (polite, confrontational, etc..), openness to share emotions, innovation mindset and more. Culture starts at the inception of a company and evolves over time. A useful way to think about culture evolution is path dependency, where early events and decisions establish policy paths that have lasting effects on subsequent events and decisions. Path dependence theory was originally developed by economists to explain technology adoption processes and industry evolution. The theoretical ideas have had a strong influence on evolutionary economics (NELSON & WINTER, 1982). Path dependency is not only about technology. It also relates to any form of behaviour that has its origin in the past and becomes so entrenched that it becomes locked-in (HAGE, 2000). Culture and path dependency can contribute to understanding why people commit fraud. The SNC-Lavalin case is a good example[1].

Based in Montreal, SNC-Lavalin is one of the world's largest engineering firms, and has been involved in multibilliondollar construction projects in more than 160 countries. In Canada, it's responsible for projects such as Quebec's James Bay hydroelectric project and the Canada Line transit system in Vancouver. But some of the company's projects, and the methods allegedly used to obtain them, have gotten it into trouble. With nearly 9,000 employees in Canada and many more around the world, SNC has been a big breadwinner for the Quebec economy over the years. But that status is precarious: its legal troubles, leadership changes and political hurdles to its business in Saudi Arabia

[1] The SCN-Lavalin section is based on a series of articles published in *The Globe and Mail* throughout 2019. Each paragraph is an excerpt from these articles. References are provided at the end of each paragraph.

have cost it billions in revenue and left it potentially vulnerable to foreign takeover. SNC is one of 10 companies the Quebec government has deemed strategically important to the province, and Premier François Legault has said he wants to prevent its headquarters from leaving Quebec. That makes the firm politically important to Prime Minister of Canada Justin Trudeau (FINE, 2019).

During the rule of the late dictator Muammar al-Gaddafi, SNC was involved in major public-work projects in the North African country of Libya, including a prison, an irrigation system and a new airport. In 2011, Swiss authorities and the Royal Canadian Mountain Police [or RCMP, the federal Canadian police] began investigating claims that SNC had been bribing Libyan officials to get access to construction contracts. A former SNC executive vice-president, Riadh Ben Aissa, pleaded guilty in Switzerland to bribery and money-laundering in connection with SNC's Libyan projects, which he admitted involved bribes to Mr. Gaddafi's son, Saadi. Federal prosecutors charged SNC in 2015 with attempted bribery and fraud over its activities in Libya from 2001 to 2011. SNC tried to strike a deal with the prosecutors, in what's called a deferred prosecution agreement (FINE, 2019).

In 2018, the Canadian Liberal government's budget bill slipped in the legal concept of deferred prosecution agreements (DPAs). DPAs involve remediation mechanisms that could allow a company to avoid prosecution when charged with committing a serious criminal offence. That may not have seemed like big news at the time, but the issue of DPAs are now right in the middle of a national controversy around ethics. DPAs are not unethical. It's how they are used that matters. DPAs have a longer history elsewhere than in Canada. After the conviction and subsequent implosion of Arthur Andersen, one of the top accounting firms in the world, U.S. prosecutors became convinced that the collateral damage to employees, shareholders and other innocent parties necessitated the use of DPAs; since 2000, the U.S. Justice Department has entered into approximately 400 of them. The belief in the United States is that DPAs strengthen compliance and enforcement, while limiting damage to innocent parties. Britain, meanwhile, established DPAs in 2014, and has only entered into three such agreements. The rarity of their use may be a result of the requirement that such agreements must be concluded under the supervision of a judge who has to be convinced that it is in the interests of justice, and that the terms are fair, reasonable and proportionate. This judicial supervision does not exist in the U.S. scheme for DPAs, but it is in the Canadian legislation. Opponents fear that DPAs could trigger what is known as moral hazard, especially as it relates to criminal corruption, since companies might be less incentivized to avoid criminal corruption if they know a DPA is a likely result. There

are some challenges with DPAs. First, DPAs should be available only to first-time offenders who are cooperative, acknowledge wrongdoing and show remorse through self-imposed compliance systems. Second, the level of pervasiveness of the conduct and level of senior management involved in the misconduct should be taken into account. Finally, the extent of potential effects of conviction on innocent parties should be a factor in deciding whether a DPA is reasonable. There is little doubt that the conviction of SNC-Lavalin would have tremendous consequences on innocent parties. A 10-year loss of access to procurement contracts in Canada and other places could result in major employment and economic losses in Quebec and beyond – perhaps even endangering the company. There can, however, be debates on whether SNC satisfies the other criteria proposed for Canadian DPAs (MENDES, 2019).

Indeed, in 2013, The World Bank announced that it was disqualifying SNC from any of its contracts for 10 years because the company had conspired to pay bribes while bidding for projects in Bangladesh and Cambodia. It was the longest sanctioned period ever agreed to in a World Bank settlement. In 2015, SNC-Lavalin agreed to pay $1.5-million to settle a case filed by the African Development Bank Group. The bank said the settlement resolved allegations, which SNC did not contest, that former employees of SNC-Lavalin International Inc. had ordered illicit payments to public officials to win road contracts in Mozambique and Uganda in 2008 and 2010. SNC was also involved in one of Canada's biggest corruption cases. Its former chief executive officer Pierre Duhaime pleaded guilty in 2019 to breach of trust after he admitted that SNC executives had paid $22.5-million to two top managers of the McGill University Hospital Centre in exchange for information that helped the company win a $1.3-billion contract to build the new hospital (YORK et al., 2019).

In 2015, the RCMP charged SNC-Lavalin and two subsidiaries with paying nearly $48-million to public officials in Libya between 2001 and 2011 to influence government decisions under the Muammar al-Gaddafi regime. The RCMP also charged the Montreal-based company, its construction division and a subsidiary with fraud and corruption for allegedly defrauding Libyan organizations of about $130-million. SNC-Lavalin has sought to avoid criminal trial on fraud and corruption charges stemming from an RCMP investigation into its business dealings with Libya (FIFE et al., 2019).

Given the Trudeau government's repeated refrain that Canada is bound by the rule of law in the context of the Meng Wanzhou (the current CFO of Huawei) extradition proceedings to the United

States, it is troubling to learn that the Prime Minister's Office may have been actively pressuring then Attorney-General Jody Wilson-Raybould to enter into a remediation agreement with SNC. This goes against the constitutional principle that attorneys-general must act independently of partisan concerns (QUAID & TAMAN, 2019). According to Ms. Wilson-Raybould's testimony before a House of Commons committee, she said that, for four months, "I experienced a consistent and sustained effort by many people within the government to seek to politically interfere in the exercise of prosecutorial discretion in my role as the Attorney General of Canada in an inappropriate effort to secure a deferred prosecution agreement with SNC-Lavalin" (BLINCH, 2019).

After the RCMP charged SNC-Lavalin with corruption and fraud in 2015, the company raised the possibility, both internally and to federal officials, that it might sell itself or move its headquarters to Britain if forced to endure a potentially damaging trial ending in a conviction. That would not be easy, however. A $1.5-billion loan agreement with the Caisse stipulates that SNC-Lavalin has to keep its base in Montreal until at least 2024 (BLINCH, 2019). Recently, the company reinforced the fact that it is considering spinning out business units such as WS Atkins ahead of the criminal trial for bribery and fraud charges (VAN PRAET & KILADZE, 2019). Such a breakup is one way to isolate the firm from legal sanctions by exposing only the Canadian component to loss of contracts. In its last investors' meeting, the company said that it aims to cut up to 1,000 jobs (VAN PRAET & KILADZE, 2019). Such a decision is to nominally ensure that the company will deliver on financial guidance for the current fiscal year but, of course, the loss of jobs is the legacy of its past corruption.

In Quebec, the new Legault government, which has pressed Ottawa to give SNC-Lavalin a negotiated settlement, calls the builder one of roughly 10 companies strategic to the Quebec economy (BLINCH, 2019). Its head office includes the type of workers Quebec Premier François Legault covets as he tries to produce more jobs paying $50,000 a year or more for Quebec's economy and close the average wage gap with Ontario. As Quebec Economy Minister Pierre Fitzgibbon put it in an interview with *The Globe and Mail* in January 2020: "This is his obsession" (BLINCH, 2019).

There is a view among political leaders in both Quebec City and Ottawa that Canada needs to level the playing field for homegrown companies competing against rivals in countries that already make use of out-of-court settlements for corporate offenders, or so-called deferred prosecution agreements. That includes the U.K., the United States and France. In a 2017 letter to Ms. Wilson-Raybould, former CEO of the Business Council of Canada John Manley said: "The fact that DPAs already are in use in

several OECD countries puts our firms at a competitive disadvantage". "It's been no secret that the Trudeau government wants to make sure that there is some mechanism available for SNC to deal with this issue," said one lawyer specializing in white-collar crimes, who was granted anonymity because he was not authorized to discuss the matter publicly according to *The Globe and Mail*. There is generally a recognition [in other countries] that this is the way to deal with these issues when you're dealing with companies (BLINCH, 2019).

Interestingly, SNC-Lavalin has been one of the leading recipients of support from the Canadian federal government's export agency, Export Development Canada (EDC), obtaining at least $2-billion in loans over the past two decades, and possibly billions more, according to data obtained by *The Globe and Mail*. Much of the support – at least $800-million in loans and as much as $1.7-billion – was provided to SNC-Lavalin after news broke of an RCMP investigation into alleged corruption at the company in 2011. Since then, the engineering company has faced corruption allegations in Canada, two Asian countries and three African countries. EDC says it suspended its support to SNC-Lavalin between 2014 and 2017 as a result of the corruption allegations against the company – but resumed large scale financing for the company over the past two years after concluding it had improved its ethics (YORK et al., 2019).

DPA is at the center of the ethical-politico dilemma the Canadian government is facing in the SNC-Lavalin bribery case. DPAs were meant as a solution to bribery activity where, in the first occurrence, the firm admits to being guilty and makes restitution. None of these conditions applied to the current case of SNC-Lavalin. The firm demonstrated that it is a serial practitioner of corruption and does it on an industrial scale and when apprehended hides behind the argument that its employees would be punished by a criminal conviction. SNC-Lavalin's culture demonstrates that it tolerated such behaviour for years.

3. MATERIALS AND METHODS

This chapter gives a detailed account of the material and methods used to conduct this research. The chapter begins with the empirical evidence: the audit profession in Quebec. Next, an important discussion on the automation debate and how previous research led me to develop the Task Complexity Framework. Then, I perform a critical analysis of definitions of artificial intelligence and propose a practical definition of artificial intelligence. Such discussion is critical to assess the type of intelligent agents that can contribute to the audit profession in Quebec.

The AAA definition of auditing presented in the Introduction of this study is broad enough to encompass external, internal and governmental auditing. This dissertation focuses on external audit, more specifically, the audit of an entity's financial statements, which summarize the entity's transactions and business events over a period. The Chartered Professional Accountants of Canada (CPA Canada) defines a financial audit as follows: "The purpose of an audit is to enhance the degree of confidence on intended users in the financial statements. This is achieved by the expression of an opinion by the auditor on whether the financial statements are prepared, in all material respects, in accordance with an applicable reporting framework" (CPA Canada Online Handbook, 2020).

The definition contains three important concepts. The first one is to enhance. As stated in the Introduction of this study, there are three key reasons justifying why auditors enhance the degree of confidence: complexity, remoteness and consequence. These three reasons relate to information risk. Information risk refers to the possible failure of financial statements to appropriately reflect the economic substance of business activities and related risks and uncertainties. Information risk gives rise to the misstatements and omissions in information that auditors are hired to detect in a three-party accountability framework and suggest that management is correct. Put differently, information risk from the auditor's perspective is the risk (probability) that the financial statements distributed by a company will be materially false or misleading. Material misstatement is one that would affect a user's decision. The second concept is the degree of confidence. Since most intended users of financial statements don't have the knowledge of an auditor, they have to rely on them. It implies that users have to trust the auditor. Users exhibit an attitude about those who are trustworthy. As I explained previously, trustworthiness carries three properties that auditors must demonstrate: competence, reliability and honesty. Finally, to express an opinion, an audit must be carried out in accordance with

the Canadian Auditing Standards, which require auditors to maintain professional skepticism and make no assumption about management's honesty; stated differently, no blind trust in management.

Three key factors contribute to build the trustworthiness in the audit profession in Quebec: the regulation (since it protects the public), the admission process (since it depends on credentials) and, lastly, knowledge.

3.1. Overview of the Auditing Profession in Quebec

Since auditing is a critical function in the economy, it is extensively regulated to ensure it remains effective. In Canada, the regulation of professional auditors is currently a provincial responsibility, varying somewhat depending on the legislation in different provinces. There are fourteen provincial CPA Orders in Canada. All provincial Orders fall under the umbrella of a national organization called Chartered Professional Accountant (CPA) Canada.

CPA Canada is authorized by legislation to set all auditing standards for Canada, and it also sets the education and examination requirements for students interested in obtaining the CPA designation. The provincial Orders deliver the CPA Canada educational content and define the specific content requirement that universities are expected to teach to the students enrolled in the CPA designation process. The provincial Orders also set the code of ethical conduct that members must follow. They are responsible for the admission of members, as well as inspection of auditors' practices and disciplinary actions to enforce the professional ethics code.

3.2. Quebec CPA Order

The Quebec CPA Order has 40,000 members and 5,000 future CPAs, making it the 3rd largest professional order in Quebec. The CPA Order is a professional order as defined by the Professional Code, that is, a body whose primary mission is to protect the public. It is also an order whose members practice an exclusive profession, such that only individuals who hold the CPA auditor designation may practice public accountancy (perform audit or review engagements, and issue special reports).

Like Quebec's other professional Orders (doctors, lawyers and engineers), the CPA Order must carry out specific functions related to issuing practice permits to candidates for the profession, keeping the

roll of the Order current, supervising the practice of the audit profession and detecting illegal practice. It must also comply with a set of operating rules imposed by the Professional Code.

In order to maintain confidence in the CPA designation – the hallmark of the quality of professional services provided by CPAs – among enterprises, organizations and the general public, the Order provides support and guidance to its members by upholding its core values: integrity, excellence, commitment, innovation and respect. The key roles of CPA Quebec Order can be summarized as follows:

- Monitoring the competency development of aspiring candidates to the profession, whether trained in Quebec, in Canada or abroad;
- Ensuring the continuing competency of members through compulsory continuing education, professional inspection and support for public practice;
- Ensuring member compliance with regulatory and ethical obligations;
- Providing the contact information and confirming the status of Order members; and
- Addressing questions from the public regarding CPAs' obligations and informing it of the rights and remedies available.

3.3. Auditing Practice Standards

To practice auditing in Quebec, professional auditors must comply with practice standards, which cover the know-what, know-how, know-why, and know-that type of knowledge. Practice standards are general guides for the quality of professional work. Practice standards can be grouped into two broad families: the fitness to practice (know-how) and the technical standards (know-what, know-why, know-that). These standards are part of what we call Canadian Auditing Standards or CAS. The rest of this section focus on the fitness to practice standards.

The fitness to practice is critical since the auditor has a fiduciary duty towards intended users of the financial statements. The fitness to practice under the CAS relates to the personal integrity and professional qualifications of auditors. There are four key characteristics to meet the requirement of fitness to practice.

The first characteristic is competence. The rules of professional ethics require competence – adequate technical training and proficiency – for auditors. This competency begins with an education in

accounting/auditing and continues with on-the-job training in applying technical knowledge and in developing and applying professional judgment in real world situations. This stage provides practice in performing the assurance function, in which auditors learn to recognize the underlying assertions being made by management in each element in the financial statements, decide which evidence (data) is relevant for supporting or refuting the assertions, select and perform procedures for obtaining the evidence, and evaluate the evidence and decide whether management assertions correspond to reality and Generally Accepted Accounting Principles (GAAP). This is an important part of learning to practice the profession – learning by doing. This part is critical since technology that used to do the work keeps changing, so skills have to be relearned.

The second characteristic is objectivity and integrity. This can be referred to as intellectual honesty and impartiality. This type of objectivity in audit is achieved by maintaining professional independence, in appearance as well as in fact. Auditors must be unbiased with respect to the financial statements and other information they audit. They are expected to be fair not only to the companies and executives who issue financial information but also the outside persons who use it.

The third characteristic relates to due professional care. This requires observance of the rules of professional ethics and the CAS. Auditors must be competent and independent, exercising proper care in planning and supervising the audit, in understanding the auditee's control structure, and in obtaining sufficient appropriate evidence. Their training should include computer auditing techniques because of the importance and pervasiveness of computers in the business world. Due professional care is a matter of what auditors do and how well they do it. A determination of proper care must be reached based on all facts and circumstances in a particular case. When an audit firm's work becomes the subject of a lawsuit, the question of observance of the rules is frequently cited.

The fourth characteristic relates to ethical behaviour. The essential role and responsibility of an auditor is to establish and communicate assurance to the users that financial statements are fairly presented, implying that unethical reporting has not occurred. Professional practice in a three-party accountability framework can give rise to many conflicts and dilemmas. The auditors must always remain ethical in performing their work. Ethics can be seen as a systematic study of reflective choice and the standards of right and wrong (WHEELWRIGHT, 1959). Two important states of mind are expected from an auditor to act ethically: professional skepticism and professional judgment. Professional skepticism means that an auditor always wants to question the claims made by

management of a company under audit and look for corroborating evidence. Such a state of mind involves critical thinking. A key step in critical thinking is applying logic to reasoning. For an auditor, applying logic means identifying reasons supporting a claim or a conclusion. Critical thinking requires that the truth or substantial truth of the reason should be met before we can say that an audit conclusion is justified by the reason given. Critical thinking also involves questioning the application of the standard, the concepts and the principles underlying it, and the consistency of the standards with one another. Logic is concerned with the link between reasons and conclusion. Professional judgment in auditing is critical thinking. According to CAS 200, professional judgment is "the application of relevant training, knowledge and experience, within the context provided by auditing, accounting and ethical standards, in making informed decisions about the courses of action that are appropriate in the circumstances of the audit engagement" (CPA Canada Online Handbook, 2020).

These four characteristics have a common denominator – learning by doing. Knowledge is required to start practicing the audit profession. Technical knowledge is tough at university, including ethics, but time and experience are required to acquire the mental fitness to practice the profession.

3.4. Auditing Process in Quebec: An Overview

The preliminary stage of determining if an audit engagement can be accepted by an audit firm involves several important tasks. First, it starts with obtaining an understanding of the potential audit's client business and its financial reporting requirements, its corporate governance and ownership structure, and its main financial statement users. Once done, several questions must be answered regarding professional ethics requirements. When these tasks are completed, the auditor will make a decision to accept the client and write an audit engagement letter. Once the letter is received and accepted by the client, the audit engagement begins. We can decompose the financial statement audit process in three steps: i) risk assessment, ii) responding to the assessed risks, and iii) concluding and reporting (issuing the audit opinion). Based on the Canadian Auditing Standards, we can decompose the auditor's job into 32 high level tasks.

Table 1. Audit Steps and Tasks

Steps	Elements	High level audit tasks
Risk assessment	Pre-engagement activities	1. Obtain an understanding of the auditee entity and the audit engagement. 2. Determine if pre-conditions for the audit are present, including management integrity and responsibility. 3. Assess the auditor's independence from the auditee. 4. Identify the auditor's risk arising from the engagement. 5. Make the engagement acceptance or continuance decision.
	Preliminary audit planning: risk identification	6. Understand the auditee's business, environment, risk, management strategy, and controls. 7. Analyse management's draft financial statements. 8. Determine materiality level. 9. Summarize preliminary planning decisions in the overall audit strategy document. 10. Assess the risks of material misstatement at the overall financial statement level. 11. Identify key financial statement assertions and related risks of material misstatement. 12. Determine the audit risk level to be accepted. 13. Understand the accounting information system and financial reporting process.
	Risk assessment procedures to plan audit	14. Use the internal control framework to understand the auditee's internal control. 15. Assess inherent and control risks, combined risk of material misstatement, at the assertions level for account balances, transaction classes, and disclosures. 16. Assess risk of fraud and non-compliance with laws. 17. Identify the evidence procedures available to meet audit objective. 18. Determine the nature and timing of the control tests and substantive procedures required to respond to the assessed risk at the assertion level. 19. Develop the audit plan and detailed program. 20. Document the planning by creating working paper file.
Response to assessed risks	Internal control documentation and testing	21. Identify management controls related to financial statement assertions and any significant control deficiencies. 22. Determine whether control reliance is appropriate for specific assertions. 23. Design control tests to evaluate the effectiveness of the control to be relied on. 24. Apply control evaluations to planning further substantive procedures. 25. Communicate to management regarding any control deficiencies.
	Sampling decisions	26. Apply sampling concepts to determine the extent of control tests and substantive procedures, and design simple control and substantive audit programs.
	Performance of planned audit programs	27. Perform planned control testing and substantive audit procedures for account balances and transactions in all significant accounting processes, including accounting estimates. 28. Perform the audit of revenues, expenses, assets and liabilities. 29. Perform audit completion procedures, review subsequent procedures and contingencies, assess going concern.
Concluding and reporting	Audit findings review	30. Review documentation to assess the sufficiency and appropriateness of the audit evidence obtained. 31. Review accumulated known and likely misstatements to assess financial reporting risk and the adequacy of disclosures.
	Opinion and report	32. Form a conclusion on fair presentation and issue an appropriate opinion and report.

Source: Author's compilation based on CAS.

The audit process can be broken down in three categories: i) audit activities, ii) communication with management, and iii) communication with the governance body.

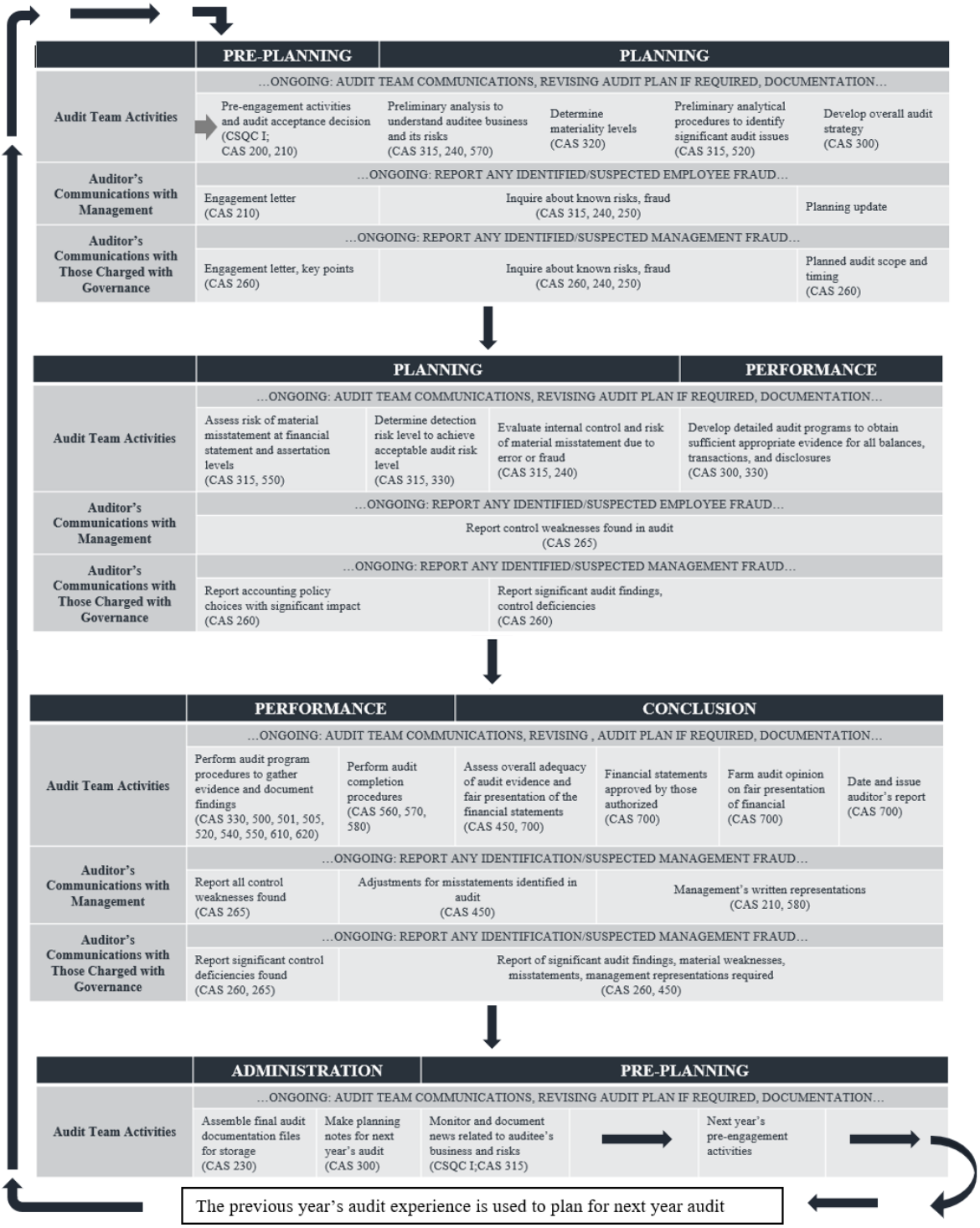

Figure 3. Overview of the Audit Process

Source: Author's compilation based on CAS.

Figure 3 deserves few comments. For the risk assessment step, the audit team must obtain a deep knowledge of the auditee's business operations and the environment the business operates in. This knowledge helps the auditor understand the kinds of transactions, account balances, and disclosures that should appear in the financial statements to fairly portray the underlying economic realities of the business. Key information for this knowledge is prior years' financial statements and current year draft financial statements. Analytical procedures, such as comparative analysis of financial ratios and trends are a useful way to get a preliminary understanding and gain a perspective on risky areas in the financial statements. Other sources include discussions with management, industry reports, research reports, and business news about the company. These sources of information are also used for the first judgments an auditor must make: the materiality level. The term materiality refers to a monetary amount that the auditors believe financial statement users would find significant, i.e., would influence a decision made based on the financial statements. Materiality decisions involve both quantitative and qualitative considerations about what is significant to the users of the financial statements. This number affects all other tasks to be done during the audit work. After the audit work is performed, the materiality concept is applied to reach a final decision on whether the financial statements are materially misstated.

Auditors use their business knowledge to identify the operating and environmental risks that management faces, and the strategies and controls management uses to reduce these risks. Environmental risks include external factors like industry competition, technology changes, regulation, interest rates, supply chain uncertainty, and market price changes. These risks can impact the ability of the company to remain in business, the value of the assets, the pricing strategy, etc. Operating risk arises from internal factors such as inappropriate strategy, weak management systems and controls, inappropriately skilled workers, etc., and auditors need to consider if there are many significant risks that could have a material impact on the transactions, accounts balances, or disclosure in the financial statements.

To identify these risks and assess their potential impact in the financial statements, the auditor needs to apply critical thinking, drawing on business knowledge, analysis of the financial statements and professional judgment to generate hypotheses about things that could go wrong [including any impact in the financial statements]. Auditors must focus their work on key audit areas [based on the

materiality] to ensure they do an effective audit to meet the ethical requirement discussed earlier: due care. The duty of care means that the auditor must prioritize quality of work over efficiency.

Based on ABDOLMOHAMMADI's work, the 32 high level tasks presented in Table 1 can be decomposed into 332 tasks. To identify these tasks, ABDOLMOHAMMADI (1999) consulted auditing texts, professional standards and audit manuals. ABDOLMOHAMMADI breaks down the audit phases into six phases: orientation, understanding the client, control structure, tests of controls, substantive tests, and forming an opinion on the financial statement reporting.

Table 2. Taxonomy of Audit Task Structure

			Task structure		
Audit phase	**No. of tasks**	**Median rank**	**Structured**	**Semistructured**	**Unstructured**
Orientation (OR)	45	Senior auditor	7 (16%)	14 (31%)	21 (53%)
Control structure* (CS)	75	Senior auditor	10 (13%)	58 (77%)	7 (10%)
Substantive tests (ST)	171	Assistant auditor	114 (67%)	54 (32%)	3 (1%)
Forming an opinion (FO)	41	Manager	0 (0%)	9 (22%)	32 (78%)
Total	332		131 (39%)	135 (41%)	66 (20%)

Source: ABDOLMOHAMMADI (1999)

*includes understanding the client and tests of controls

Although ABDOLMOHAMMADI's study grouped differently the audit process (and the related tasks) than the CAS, both groupings capture one of the most important tasks the auditor must perform: assessing the risk of fraud in the financial statements. Based on ABDOLMOHAMMADI's framework, in some tasks, the problem can be well defined with very limited number of alternatives, thus requiring very little judgment to make a final choice. These tasks are considered to be structured. Other tasks with ill-defined problems that have many alternative solutions require considerable judgment and insight to make a choice among alternatives. These tasks are considered to be

unstructured. Lastly, somewhere between the structured-unstructured tasks can be reasonably defined the tasks with a limited number of alternatives, which require a medium level of judgment to make a choice.

The ranking proposed by ABDOLMOHAMMADI is interesting because it associates the profession rank (experience required) to execute a task. Based on the Task Formula, the less structured a task is, the more complex it is, because it requires the auditor to use extensively their judgment to perform the task, which is the application of knowledge and experience.

From these 332 tasks identified in Table 2, ABDOLMOHAMMADI's task OR35 (the assessment of the susceptibility of the assets under audit to material fraud of misappropriation), task OR27 (the assessment of the management attitude about financial reporting) and task OR42 (the assessment of the susceptibility of management to override existing control) are encompassed in task 16 in Table 1. Such differences can be explained by the evolution of the auditing standards between 1999 and 2020. Based on CAS 240 (the auditor responsibilities relating to fraud in an audit of financial statements), I decomposed task 16 into 160 sub-tasks. The list of the 160 sub-tasks is presented in Appendix 2. The list will be used in Section 4 to assess the contribution of intelligent agents to execute task 16.

3.5. Brief Overview of the Mainstream Automation Debate

A large and growing strain of literature on task automation (computerization of jobs, algorithmization of jobs, etc.) has been published over the past decade. Many of these studies tend to simplify the feasibility of the automation of a task. There are many reasons for such situation. Two important ones must be mentioned. First, the methodology used by the authors of these highly cited studies take a high-level view. Second, there is a lack of a common understanding of the term artificial intelligence. Based on the Task Formula, this section aims at proposing a practical definition of artificial intelligence and present a Task Complexity Framework to assess the ability of intelligent agents to identify fraud risk factors (task 16).

3.5.1. ALM Study

The question of whether human-based activity can be executed by an intelligent agent has primarily been addressed using the distinction between routine and non-routine tasks. The effects of automation have been studied intensively in economics since the publication of the seminal paper by AUTOR,

LEVY & MUNDANE in 2003 (ALM). The ALM study is an important contribution in assessing how to build intelligent agents since it focused on task and not job or profession. In fact, a fast-growing literature shows that technological change is replacing labour in routine tasks, raising concerns that labour is racing against machine.

ALM proposed a task-based framework to investigate how adoption of computer technology changes job tasks and employer demand for human skills. ALM discuss the substitution of routine tasks by machines for profit-maximizing firms. Hence, whether substitution takes place hinges not only on technological capabilities, but on the relative price of performing a task by either humans or machines. But classifying single task items into distinct domains leads to a number of challenges and problems.

A wealth of quantitative and case-study evidence documents a striking correlation between the adoption of computer-based technologies and the increased use of college-educated labour within detailed industries, within firms, and across plants within industries. This robust correlation is frequently interpreted as evidence of skill-biased technical change. Yet, as critics point out, this interpretation merely labels the correlation without explaining its cause. As AML stated, it fails to answer the question of what it is that computers do – or what it is that people do with computers – that causes educated workers to be relatively more in demand.

In their model, ALM differentiate between five domains of job tasks: routine manual tasks, routine cognitive tasks, non-routine manual tasks, non-routine cognitive tasks, and analytical and interactive tasks. The most relevant differentiation in their model is between routine and non-routine. A task is routine if it can be accomplished by a machine following explicit programmed rules. ALM assume that different domains of jobs are typically performed by different groups of skilled workers: cognitive non-routine tasks would be typical for high-skilled professional and managerial jobs; routine manual and cognitive tasks constitute most middle-education jobs; and manual non-routine tasks are mostly performed by unskilled workers.

ALM formalize and test a simple theory of how the rapid adoption of computer technology – spurred by precipitous real price declines – changes the tasks performed by workers at their jobs and, ultimately, the demand for human skills. This approach focuses on determining the tasks that computers are best suited to perform and whether computer-performed tasks serve as complements or substitutes for human job skills. ALM affirm that computers substitute for cognitive and manual tasks

that have large routine components. At the same time, computers also complement complex non-routine problem-solving tasks and complex communications tasks. Conversely, low-skilled, manual, non-routine jobs are not directly affected by computerization. Their hypothesis is that computerization leads to a decline in the demand for middle-education workers and leads to an increase in the relative demand for both the most educated and least educated workers.

ALM research is important because it provides a theoretical framework for understanding how computers affect jobs and, thus, explain why new computer technologies are skill-biased. According to ALM, firms react to an exogenous fall in price of computers by changing the task mix towards more non-routine activities, requiring both expert thinking (managing and solving analytical problems) and complex communication skills. This switch in comparative advantage increases the demand for college-educated workers, because they have the knowledge and ability to carry out non-routine analytical and interactive tasks.

The ALM model is deterministic in that it takes the driving force to be exogenous reductions in the price of computers; the effects are presumed to be both direct and indirect. Computer price reductions alter the needed skill mix by directly displacing routine tasks, and they may allow for certain organizational changes to become profitable, such as the delayering of management hierarchies because they complement the new technologies (BRESNAHAN et al., 2002). Those new forms of work organization may then involve more non-routine tasks. As a result, this will impact skills requirements for an organization and skills utilization.

A challenge in the empirical testing of the ALM model is the identification of programmable tasks from descriptions and classifications. Based on my literature review, we can observe that existing studies that make such classifications are not comprehensive and the validity of the categorizations is not always clear-cut. The following example illustrates the potential for misclassification: in the ALM model, “measuring” includes measuring, testing, and quality control tasks. Although measuring can be considered to be a manual routine task, testing and especially quality control might also include non-routine job activities. The definition of measuring use by ALM does not take those distinctions into consideration.

The classification of tasks into distinct domains is, by no means, an exact science (HANDEL, 2008). It involves grouping multiple tasks that are linked in theory, but in practice, many work actions are

indivisible, involving tasks from multiple domains, making their allocation problematic. Another difficulty in classification of tasks relates to interaction. In the ALM model, routine and non-routine tasks do not consider if a task is independent or interactive. Independent work requires little or no collaboration or communication with others, while work performed interactively involves more collaboration and/or communication with others, and relies more on communication skills and empathy, which cannot be automated.

Another challenge with the ALM model is that the model is based on task approach with job task measured at occupational level. The model ignores the fact that tasks may vary significantly among workers based on gender, education and race.

3.5.2. FO Study

FREY & OSBORNE (2013) extended the work done by ALM. As opposed to ALM, FREY & OSBORNE (FO) took a different path to examine how susceptible jobs are to computerization. The Industry 4.0 debate on jobs harkens back to FO, who estimated that, in the coming years, 47 percent of U.S. jobs could be automated through the application of new digital technology. FO begin by implementing a novel methodology to estimate the probability of computerization for 702 detailed occupations, using a Gaussian process classifier. Based on these estimates, FO examine expected impacts of future computerization on U.S. labour market outcomes, with the primary objective of analyzing the number of jobs at risk and the relationship between an occupation's probability of computerization, wages and educational attainment. According to FO estimates, about 47 percent of total US employment is at risk. Another influential analysis sharing the pessimistic tone of FO about the extent of job losses due to automation is the work of FORD (2015) in *The Rise of the Robots*.

In contrast to ALM, FO only assessed the technical capability of substituting a certain task by machines and not its economic feasibility. Contrary to ALM, their analysis is not confined to routine labour inputs. This is because recent developments in machine learning and mobile robotics, building upon big data, allow for pattern recognition, and thus enable computer capital to rapidly substitute for labour across a wide range of non-routine tasks. In fact, FO argues that, beyond some bottlenecks, it is largely already technologically possible to automate almost any task, provided that sufficient amounts of data are gathered for pattern recognition. According to FO, three types of tasks are not susceptible to automation over the next decade or two: perception and manipulation tasks, creative

intelligence tasks, and social intelligence tasks. Creative intelligence tasks are tasks that require the ability to come up with new ideas or artifacts that are novel and valuable. Ideas, in a broader sense, include concepts, poems, musical compositions, scientific theories, cooking and jokes, whereas artifacts are objects such as paintings, sculpture, machinery, and poetry. The challenge according to FO is to find some reliable means of arriving at combinations that make sense. Social intelligence tasks involve negotiation, persuasion and care. To aid the computerization of such tasks, research is being undertaken within the field of Affective Computing according to FO.

The seminal paper by FO is the first to make quantitative claims about the future of jobs. One must recognize the difficulty to perform such a quantitative analysis. FO's model has been an important contribution to assess the impact of automation and artificial intelligence on jobs. Their model, however, has a number of limitations. The following paragraphs present five important limitations.

Instead of using an objective methodology to categorize a task (as AML did), FO used a subjective approach combined with the O∗NET bottlenecks to computerization. For the subjective part, FO, together with a group of ML researchers, subjectively hand-labelled 70 occupations, assigning 1 if automatable, and 0 if not. For their subjective assessments, FO drew upon a workshop held at the Oxford University Engineering Sciences Department, examining the automatability of a wide range of tasks. FO label assignments were based on eyeballing the O∗NET tasks and job descriptions of each occupation. This information is particular to each occupation, as opposed to being standardized across different jobs.

FO model aggregated the 903 O∗NET occupations into 702 occupations by taking the mean of the tasks as reported in O∗NET. In addition, they looked at only 70 occupations out of those 702 aggregated occupations. Their model led them to assess the potential for automation for the remaining 632 occupations. The extrapolation made by the authors lead us to question the overall conclusion of their research.

The labelling of automation was, as the authors admit, a subjective assignment based on eye balling the job descriptions from O*NET. Labels were only assigned to jobs where the whole job was considered to be (non) automatable, and to jobs where the participants of the workshop were most confident. If we know that a job is 100% automatable, we also know that every task of that job must

be completely automatable. But what if a job is 81% automatable? Is every task 81% automatable? Or are 81% of the tasks completely automatable, and 19% not at all?

The O*NET (and DOT used by AML) database suffers from an important deficiency because it only provides information on job characteristics at the level of occupation, not at the worker's level. As a result, the personal and physical dimension of an occupation are not considered. This makes the analysis of within occupation heterogeneity in task demands and its relationship to earnings infeasible. Also, job tasks vary significantly among workers within a given occupation. Differences can be related to gender, education or other factors. DOT approach at the occupation level does not capture those variables.

FO's approach reflects the capabilities of the technologies not the actual uses of them. Since most technologies are rarely use to their full potential, their methodology overestimates the amount of job that could be automated in real life.

Methods to classify tasks items into distinct domains vary substantially. Though this domain is theoretically well defined, it is very hard to identify codifiable routine tasks.

3.5.3. Other Studies

After the publication by FO, other economists struck back. In 2016, a trio of researchers at the Organization for Economic Cooperation and Development (OECD, 2016) used an alternate model to produce an estimate that seemed to contradict FO: just 9 % of jobs in the United States were at high risk of automation. While FO asked machine-leaning experts to judge the automatability of an occupation, the OECD team pointed out that it's not an entire occupation that will be automated but rather specific tasks within those occupations. The OECD report argues that this focus on occupation overlooks the many different tasks an employee performs that an algorithm cannot: working with colleagues in groups, dealing with customers face-to-face, etc.

The OECD team proposed a task-based approach, breaking down each job into its many component activities and looking at how many of those could be automated. The OECD team then ran a probability model to find out what percentage of jobs were at high-risk (i.e., at least 70 percent of the task associated with the job could be automated). They found that, in the United States, only 9 percent of workers fell in the high-risk category. Applying the same model on twenty other OECD countries,

the authors found that the percentage of high-risk jobs ranged from just 6 percent in Korea to 12 percent in Australia.

The OECD task-based approach came to hold sway among researchers, but not all of them agree with the report's conclusion. In early 2017, researchers at PwC (2017) used the task-based approach to produce their own estimate, finding instead that 38 percent of jobs in the United States were at high risk of automation by the early 2030s. It was a striking divergence from the OECD's 9 percent; one that stemmed simply from using a different algorithm in the calculations.

After all those widely divergent conclusions, MCKINSEY GLOBAL INSTITUTE (2017) landed somewhere in the middle. Using the popular task-based approach, the McKinsey team estimated that around 50 percent of work tasks around the world are already automatable. When it came to job displacement, the McKinsey researchers were less pessimistic. If there is a rapid adoption of automation techniques, 30 percent of work activities around the world could be automated by the 2030s but only 14 percent of workers would need to change occupations. Following the MCKINSEY study, BAIN & COMPANY (2018) estimated in their study that by 2030, employers will require 20% to 50% less employees. Experts continue to be all over the map, with estimates of automation on the potential of job losses ranging from just 9 percent to 50 percent. Table 3 summarizes the findings of these key reports.

Table 3. Summary of Key Reports on Job Automation

Frey et Osbourne (Oxford University)	OECD	PwC
▪ "The Future of employment" ▪ 2013 ▪ "Occupation-based study" ▪ 47% of jobs in the United States can be automated	▪ "The risk of automation for jobs in OECD countries" ▪ 2016 ▪ "Task-based study" ▪ 9% of jobs can be automated	▪ "Will robots really steal our jobs?" ▪ 2017 ▪ "Task-based study" ▪ 38% of jobs in the United States have a high risk of being automated
McKinsey Global Institute	**Bain and Company**	
▪ "What the future of work will mean for jobs, skills, and wages" ▪ 2017 ▪ "Task-based approach" ▪ 50% of tasks at the international level can be automated	▪ "Labor 2030: The collision of demographics, automation and inequality" ▪ 2018 ▪ Macro approach ▪ By 2030, employers will require 20% to 25% less employees	**How to interpret these studies?**

Source: Author's compilation

“Empirical studies which would allow us to make reliable statistical generalizations on the future developments of work and employment are still thin on the ground” (ALASOINI, 2018, p. 12). There is no widely shared agreement on the tasks where AI systems excel, and thus little agreement on the specific expected impacts on the workforce and on the economy more broadly.

A novel way to look at potential impact of intelligent agents on jobs is to aggregate the knowledge and ideas of inventors and companies patenting. MICHAEL WEBB, a graduate student in the economics department at Stanford University, created an algorithm to analyze AI patents that had been filed and cross-referenced them with tasks performed in various jobs. WEBB examined a pool of approximately sixteen thousand patents that contained verb-object pairs such as “diagnose disease” and “predict prognosis”, which correlated with descriptions of occupations used by the Department of Labor in the U.S. As a way of testing the effectiveness of this research method, WEBB looked back at the previous thirty years or so of patents in software and industrial robotics to see if the predictions generated by his algorithm about employment and wage decline is what was observed in reality. The results were conclusive.

WEBB then analyzed recent AI patent filings and found them using verbs such as recognize, detect, control, determine, and classify, and nouns like patterns, images, and abnormalities. The jobs that appear to face intrusion by these newer patents are different from the more manual jobs that were affected by industrial robots: intelligent machines may, for example, take on more tasks currently conducted by physicians, such as detecting cancer, making prognoses, and interpreting the results of retinal scans, as well as those of office workers that involve making determinations based on data, such as detecting fraud or investigating insurance claims. People with bachelor’s degrees might be more exposed to the effects of the new technologies than other educational groups, as might those with higher incomes. The findings suggest that nurses, doctors, managers, accountants, financial advisers, computer programmers, and salespeople might see significant shifts in their work. Occupations that require high levels of interpersonal skills seem most insulated according to the study (WEBB, 2019).

3.5.4. The Technology Panic Cycle

Some of the studies in Table 3 referred to what ROBERT ATKINSON, President of the Information Technology and Innovation Foundation, called the “technology panic cycle” (ATKINSON, 2018).

Technology fear is not new. In a 1930 essay, English economist JOHN MAYNARD KEYNES wrote about the onset of a new disease which he named technological unemployment, that is, unemployment due to our discovery of means of economizing the use of labour outrunning the pace at which we can find new uses for labour. Figure 4 presents ATKINSON's technology panic cycle.

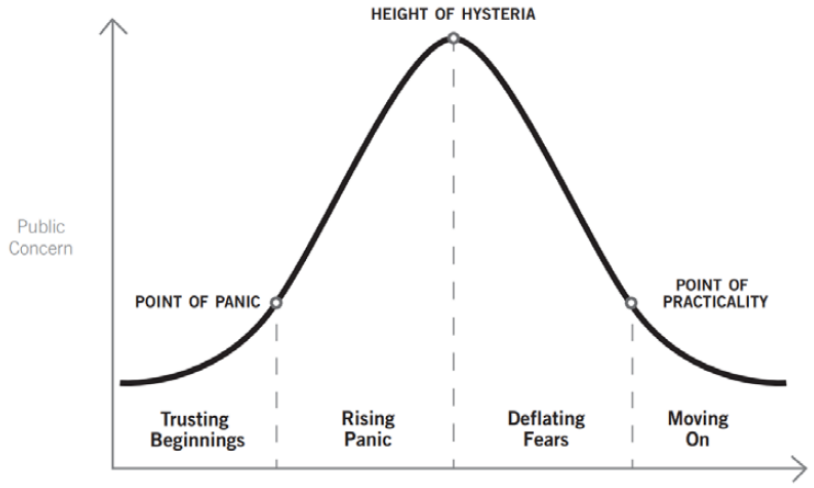

Figure 4. The Technology Panic Cycle

Source: ATKINSON (2018)

According to ATKINSON, during the "Rising Panic" stage, users historically are just beginning to understand the new technology in question and just beginning to see its benefits, making people more susceptible to false statements. In most cases, because they have not yet had direct experience with the technology, antagonists can make almost any claim about the technology without losing credibility. For example, AI antagonists can and do assert that it will be able to do virtually any job. If history is a guide, then fears will continue to climb until public understanding about the technology and its benefits reach a tipping point according to ATKINSON. Various external factors, such as early stages of adoption and use of the technology, or disillusionment when fears never materialize, can affect when this tipping point occurs. At the end of the "Rising Panic" stage, privacy fears eventually will reach their zenith at what we call the "Height of Hysteria".

This is the point where, ATKINSON argues, the fever finally breaks and the public begins to dismiss hyper-inflated fears associated with the technology. It occurs as the technology becomes increasingly commonplace and interwoven into society. Assuming the pattern holds, people's fears will subside as they start to see that AI can be used for X but not for Y, and that it can do some things pretty well and

other things not so well. This period of "Deflating Fears" represents the period during which society comes to embrace the technology and individuals can see for themselves its capabilities and limits. During the "Deflating Fears" phase, new events may cause micro-panics that focus on discrete concerns of a particular aspect of the technology or its integration into society. These micro-panics usually push technology concerns back to the forefront of public attention through media buzz. But the micro-panics quickly disappear or are forgotten as it becomes clear that negative impacts are limited and vastly outweighed by overall societal benefits (e.g., in the case of driverless trucks, safer roads because of less human error and cheaper products because of lower transportation costs) according to ATKINSON.

Techno panic cycles typically end at what ATKINSON calls the "Point of Practicality", at which apocalyptic concerns fade and people move on. At this stage, the majority of the public no longer believes the dystopian claim that antagonists make, and the technology has reached a sufficient level of maturity that most people no longer express concerns about its misuse. The technology is just part of life. And we move on to a new techno-panic cycle for the next big technological innovation.

Intelligent agents have been swept up in the techno-panic cycle for at least three major reasons according to ATKINSON. First, AI is what economists call a "general purpose technology" that can and likely will affect many different aspects of the economy. As such, it is easy to offer doomsday scenarios in which it could affect all occupations, all industries, and all workers. Second, AI is extremely complicated and opaque. But unless someone has a computer science degree, ideally with a specialization in machine learning, they have virtually no understanding of AI. As such, it can and does take on mysterious and ominous powers. As a result, when an AI dystopian suggests that we are only a few short steps away from artificial general intelligence (a computer with intelligence equivalent to human intelligence) or even artificial superintelligence (a computer with vastly superior intelligence), such that Elon Musk can call it our biggest existential threat, the vast majority of people have no common-sense way to judge the validity of his claim. Finally, AI has a perception problem because of its very name according to ATKINSON. The term artificial intelligence implies that the technology has or soon will have intelligence akin to human intelligence. And, ominously, that this will quickly transform into artificial super-intelligence that is beyond human control. But this is wrong argues ATKINSON. AI has very limited intelligence – it can figure out a game of GO or that a picture

of a cat is not a dog, but it can't and won't be able to make the kinds of complex decisions that a human can make.

3.5.5. Technological Revolution

Is the AI panic cycle justified? The technological revolutions and techno-economic paradigms offer interesting insights. Strictly speaking, AI is not a new invention. Scientists have been working in the field of AI for decades. Invention is having an idea, innovation is the other 99 percent of the work. The creative insight of invention can happen in a flash. In contrast, innovation can take years. Technology evolution (innovation) follows a trajectory. CARLOTA PEREZ (2010) called it the techno-economic paradigms. Figure 5 demonstrates the trajectory of an individual technology.

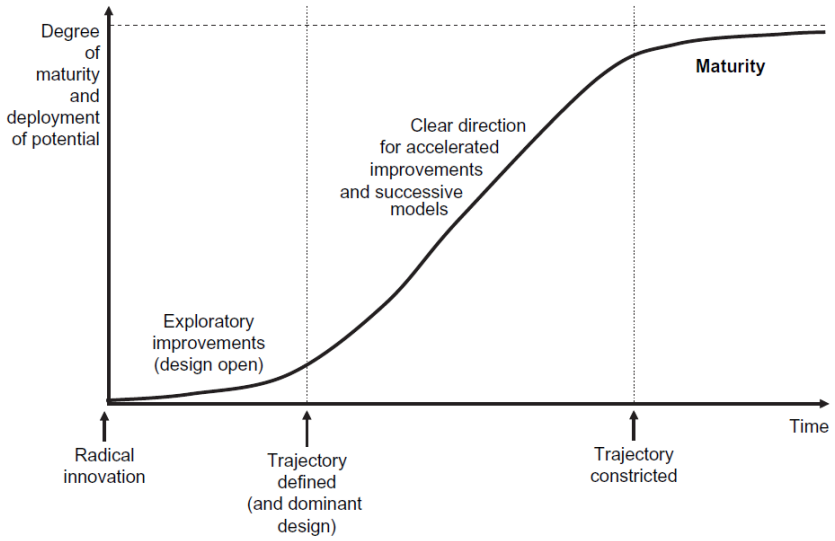

Figure 5. Technological Revolutions and Techno-Economic Paradigms

Source: PEREZ (2010)

Studies of innovation have shown that the introduction of technological change such as AI is not random but path dependent and interdependent with other innovations clustered in systems, which are, in turn, interconnected in revolutions. "Radical individual innovations are usually introduced in a relatively primitive version and, once market acceptance is achieved, they are subject to a series of incremental innovations following the changing rhythm of a logistic curve" (PEREZ, 2010, p. 186). Changes generally occur slowly at first, while producers, designers, distributors and consumers engage in feedback learning processes. Then changes occur rapidly and intensively once a dominant

design (ARTHUR, 1988) has become established in the market; and slowly once again when maturity is reached and WOLF's (1912) law of diminishing returns to investment in innovation sets in.

The notions of trajectory or paradigm highlight the importance of incremental innovations in the growth path following each radical innovation. Though it is true that major innovations have a central role in determining new investment and economic growth, expansion depends on incremental innovation (ENOS, 1962). The numerous minor innovations in product enhancement and process improvement that follow the introduction of any new product have an important impact on productivity increases, market growth and addressing market needs and demands.

The studies summarized in Table 3 miss the mark about the potential impact of intelligent agents. They take a high-level view and look at jobs or tasks with ill-defined practical reality of a task – the complexity and what makes a task complex. They also forget the fundamental aspects of innovation highlighted by ATKINSON's panic cycle theory and PEREZ's techno-economic paradigms.

3.5.6. Task Complexity Framework

This section introduces a new framework: The Task Complexity Framework. The framework can contribute to assessing the potential contribution of intelligent agents for a specific audit task. The framework contributes to performing a more precise assessment than the methodologies used by the authors of the studies summarized in Table 3.

There is little consensus among researchers concerning the properties that make a task complex. Task-complexity has been examined in three bodies of research literature: the information processing and decision-making literature (MACCRIMMON, 1976); the task and job design literature (BEER, 1968; and the goal-setting literature (CAMPBELL & GRINGRICH, 1986). It is not possible to come up with a fit for all answers about task complexity without being subject to the same critics as the ALM's framework, the FO's model and all other studies presented in Table 3.

It is important to distinguish task complexity with task difficulty. LOCK et al. (1984) state the difference as follows: certain tasks can be difficult (i.e., requiring physical effort) without necessarily being complex; in contrast, complex tasks are difficult because they are complex. They also point out that the notion of difficulty represents a person-task interaction. A task of special complexity may be

difficult for one but not for another (e.g., driving a Formula 1 car is easy for an experienced driver but a student driver will find it difficult).

Other scientifics have an opposing view recognizing the importance of both the task-doer and the task when determining the complexity of a task. MARCH & SIMON (1958, p. 58) defined complexity "in terms of the abilities of the task-doer" meaning that tasks are more complex or less complex relative to the capabilities of the individual who performs the task. SHAW (1976, p. 308-324), in his analysis of complexity, includes "intrinsic interest and population familiarity". The task-doer view on task complexity has inherent limitations. Although these two factors are related, they are not identical. A person's familiarity with the task, their memory, ability to concentrate, computational efficiency, time constraints, and so forth, can moderate the complexity of a task.

In this research, my position is that complexity is rooted in the task and not the task-doer. To develop the Task Complexity Framework, I will start with the Task Formula introduced in the Introduction. A task has four components:

$$\text{Task} = f\ (\text{data} + \text{prediction} + \text{judgment} + \text{action})$$

Data have a direct impact on the complexity of a task. STEINMANN (1976) equates complexity with the amount of information (data) involved in a task, the internal consistency of this information, and the variability and diversity of the information itself. SCHRODER et al. (1967) identified three primary properties of a complex task: the number of dimensions of information (data) requiring attention (information load); the number of alternatives associated with each dimension (information diversity); the rate of information change (the degree of uncertainty involved). Researchers on multiple-cue probability learning argued that complexity is built upon six sources: the number of information (data) sources; cue inter-correlations; reliabilities; validities; function forms (linear, etc.) and the principle underlying the integration of the information (STEINMANN, 1976).

The more judgment that is involved in executing a task, the more complex the task. Judgment is significantly involved when accessibility to data is a challenge. Judgment plays a key role when a task has multiple performance dimensions, including non-quantitative ones (LATHMAN & YUKL, 1975) and when a task has several interrelated and conflicting elements to satisfy (CAMPBELL, 1984).

Task execution (action) impacts complexity. Complex tasks are characterized by "unknown or uncertain alternatives to reach an outcome" (MARCH & SIMON, 1958, p. 139-141) and by inexact or unknown means-end connections. Complex tasks are characterized by "the existence of a number of sub-tasks, which may or may not be easily factored into nearly independent parts" (MARCH & SIMON, 1958, p. 151-152). Complex tasks have path-goal multiplicity – i.e., the existence of several ways for accomplishing the task (TERBORG & MILLER, 1978).

Based on the aforementioned analysis, I developed the following Task Complexity Framework to assess the capability of an intelligent agent to accomplish an auditor's task:

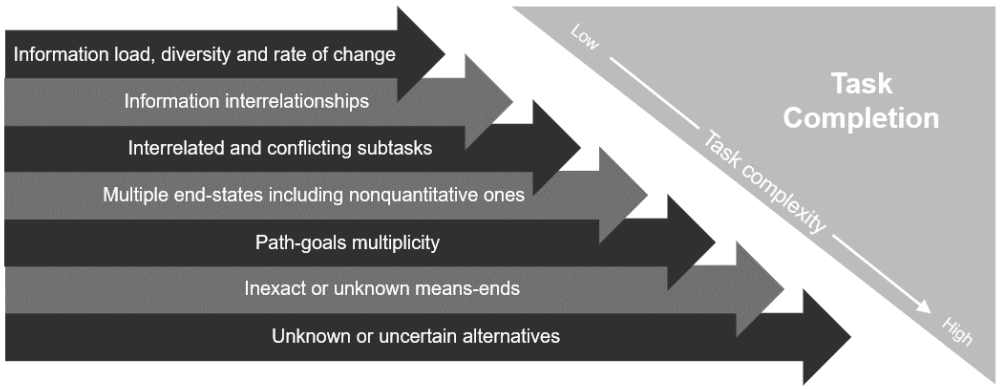

Figure 6. Task Complexity Framework

Source: Author's framework

3.5.7. Artificial Intelligence

The emergence of intelligent agents poses a new set of opportunities – and challenges – for the audit profession. The tasks that can be done by intelligent agents are much broader in scope than previous generations of technology have made possible. The expanded scope will change the value employers place on tasks, and the types of skills most in demand.

The field of AI is exploding. In 2019, the number of published papers related to AI and machine learning was nearly 25,000 in the U.S. alone, up from roughly 10,000 in 2015. And NeurIPS 2019, one of the world's largest machine learning and computational neuroscience conferences, featured close to 2,000 accepted papers from thousands of attendees. Figure 7 shows key trends in research.

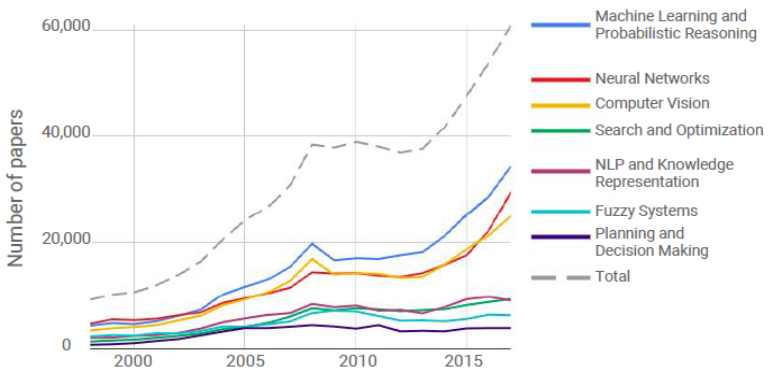

Figure 7. Number of AI Papers on Scopus by Subcategory (1998-2017)

Source: Artificial intelligence index/Elsevier (2018)

There's no question that the momentum reflects an uptick in publicity and funding – and correspondingly, competition – within the AI research community. But some academics suggest the relentless push for progress might be causing more harm than good. ZACHARY LIPTON, an assistant professor at Carnegie Mellon University, proposed a one-year moratorium on papers for the entire community, which he said might encourage thinking without sprinting/hustling/spamming toward deadlines. There's preliminary evidence to suggest the crunch has resulted in research that could mislead the public and stymie future work, as demonstrated with the summary of studies in Table 3.

Since its inception in the 1950s, AI has been falling short of its ideal. Although we are able to engineer systems that perform extremely well on specific narrow tasks, they still have stark limitations. The seven foundational layers of AI allow us to understand its current limitations.

- Philosophy: The Oxford Dictionary defines philosophy as "the study of the fundamental nature of knowledge, reality, and existence, especially when considered as an academic discipline." This definition raises a number of interesting questions when analyzing AI. Where does knowledge come from? How does knowledge lead to action? Can we define and design rules to draw conclusions?
- Mathematics*:* Philosophers stake out some of the fundamental layers for AI, but the leap to a formal science required a level of mathematics in three fundamental areas: logic, computation and probability.
- Computer: For artificial intelligence to succeed, we need two things: intelligence and artifact. The computer has been the artifact of choice. Computing power is at the core of intelligent agent ability to execute a task.

- Economics: Work in economics research has contributed to the evolution of intelligent agents. Complexity of making rational decisions is one of the areas. An important concept in AI is based on Nobel Prize winner HERBERT SIMON (1958) for his work on models based on satisficing. At its core, machine learning is about prediction. The question is how accurate should an intelligent agent be in its prediction? What is good enough?
- Neuroscience: Investigating if machines can think is inextricably connected to how humans think, which is connected to mind which is connected with brain. There is a lot of debate in the scientific community about whether the ultimate goal of AI should be to replicate how the brain works. The challenge is we don't know how the brain works, yet.
- Linguistic: In 1957, B.F. SKINNER published *Verbal Behaviour*. SKINNER believed that reinforcement learning could be used to explain verbal behaviour in humans. This was a comprehensive account of behaviourist approach to language learning. NOAM CHOMSKY criticized SKINNER's theory arguing that the behaviourist theory did not address the notion of creativity in language. Modern linguistic and AI is an intersecting field called natural language processing (NLP).
- Psychology: Cognitive psychology, which views the brain as an information-processing device, influenced the research field in AI. In the United States, the development of computer modelling led to the creation of the field of cognitive science. The MIT workshop held on September 1956 served as the starting point. At the conference, GEORGES A. MILLER presented the result of his work – *The Magic Number Seven* – on several experimenters and concluded that the immediate memory capacity of humans was approximately seven chunks of information. ALLEN NEWELL and HERBERT SIMON presented *The Logic Theory Machine* and NOAM CHOMSKY presented *Three Models of Language*. CHOMSKY claimed that all humans have at birth a universal grammar (or development mechanism for creating one) that accounts for much of their ability to learn languages. These three influential papers demonstrate how computer models could be used to address the psychology of memory, language and logical thinking.

The first challenge in artificial intelligence is that there is no common definition of the term intelligence. "For hundreds of years we have tried to understand and define intelligence and still, we have no agreement on what intelligence is" (TEGMARK, 2017, p. 49). Since there is no generally accepted definition of intelligence, there are many competing ones, including the capacity of logic,

understanding, planning, emotional, knowledge, self-awareness, creativity and problem solving, all of which are related to the foundational layers we discussed previously. The Oxford Dictionary defines intelligence as "the ability to acquire and apply knowledge and skills." From the earliest beginnings of ancient Greek philosophy, the concept of intelligence has been tied to the ability to perceive, to reason, to act and to rationalize. ARISTOTLE studied the notion of successful reasoning – methods of logical deduction (rationality). Swiss mathematician BERNOULLI's introduced the concept of utility – an invisible property – to explain intelligence: rational agents act as to maximize expected utility. SIMON introduced the concept of satisfice – rational agents satisfice when arriving at their decision. These interpretations of human-like intelligence, however, provide no actionable formal AI definition and measurement benchmark.

The lack of a satisfying definition of intelligence is a testament to the immaturity of the research field in AI. If the only successes of AI so far have been in developing narrow, task-specific systems, it is perhaps because only within a very narrow and grounded context have scientists been able to define the goal sufficiently precisely, and to measure progress in an actionable way. Goal definitions and evaluation benchmarks are among the most potent drivers of scientific progress. To make progress towards the promise of an AI field, we need precise, quantitative definitions and measures of intelligence – in particular, humanlike general intelligence.

In the context of AI research, LEGG & HUTTER (2007, p. 12) summarized no fewer than 70 definitions on intelligence from the literature into a single statement: "Intelligence measures an agent's ability to achieve goals in a wide range of environments." This is closely related to Professor TEGMARK's definition of intelligence: "the ability to accomplish complex goals, including learning" (TEGMARK, 2017, p. 49). The reason TEGMARK opted for a broader definition is "to capture all other conflicting definitions of intelligence since they are all complex goals" (TEGMARK, 2017, p. 49). This points to two characterizations, which are nearly universally – but often separately – found in definitions of intelligence: one with an emphasis on task-specific skills (achieving goals), and one focused on generality and adaptation (in a wide range of environments). In this view, an intelligent agent would achieve high skills across many different tasks (for instance, achieving high success in finding the different fraud risk factors in Appendix 2 by executing a number of tasks). Implicitly, the tasks may not necessarily be known in advance: to truly achieve generality, the agent would have to be able to learn to handle new tasks (skill acquisition).

These two characterizations map to CATELL's (1971) theory of fluid and crystallized intelligence (Gf-Gc), which has become one of the pillars of the dominant theory of human cognitive abilities: the Cattell-Horn-Caroll theory (CHC) (MCGREW, 2005). They also relate closely to two opposing views of the nature of the human mind that have been deeply influential in cognitive science since the inception of the field (SPELKE & KINZLER, 2006): one view in which the mind is a relatively static assembly of special-purpose mechanisms developed by evolution (an idea which originated with Darwin), only capable of learning what it is programmed to acquire (task-specific skills view), and another view in which the mind is a general-purpose muscle capable of converting experience into knowledge and skills, and that could be directed at any problem (learning ability view).

Advances in developmental psychology argue that neither of the two opposing views of the nature of the mind are accurate (SPELKE & KINZLER, 2006). The human mind is not merely a collection of special-purpose programs hard-coded by evolution; it is capable of a remarkable degree of generality and open-endedness, going far beyond the scope of environments and tasks that guided its evolution (SPELKE & KINZLER, 2006). The large majority of the skills and knowledge we possess are acquired during our lifetimes, rather than being innate. Simultaneously, the mind is not a single, general-purpose vehicle system capable of learning anything from experience. Our cognition is specialized, shaped by evolution in specific ways; we are born with priors about ourselves, about the world, and about how to learn, which determine what categories of skills we can acquire and what categories of problems we can solve (SPELKE & KINZLER, 2006).

These two conceptualizations of intelligence – along with many other intermediate views combining elements from each side – have influenced a host of approaches for evaluating intelligence in machines, in humans, and, more rarely, in both at the same time. This may explain why research in AI seems to be divided into four streams/definitions. These streams/definitions differ, on the one hand, as to the objective of AI application (thinking vs. acting), on the other hand, as to the kind of decision-making (targeting a humanlike decision vs. an ideal, rational decision).

Table 4. AI Research Streams/Definitions

Thinking Humanly	**Thinking Rationally**
The exciting new effort to make computers think … machines with minds, in the full in literal senses. (HAUGELAND, 1985) The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning…. (BELLMAN, 1978)	The study of mental faculties through the use of computational models. (CHARNIACK & MCDERMOTT, 1985) The study of computations that make it possible to perceive, reason, and act. (WINSTON, 1992)
Acting Humanly	**Acting Rationally**
The art of creating machines that perform functions that require intelligence when performed by people. (KURZWEIL, 1990) The study of how to make computers do things at which, at the moment, people are better. (RICH & KNIGHT, 1991).	Computational intelligence is the study of the design of an intelligent agent. (POOL et al., 1998) AI…is concerned with intelligent behaviour in artifacts. (NILSSON, 1998)

Source: Based on RUSSELL & NORVIG (2015)

AI can be seen as being about practical reasoning: reasoning in order to do a task – but the scientific community tends to anthropomorphize a statistical model. WATSON (2019) argues that the rhetoric of anthropomorphism in AI may be helpful when explaining complex models to audiences with minimal background in statistics and computer science, but it is misleading and potentially dangerous, however, when used to guide (or cloud) our ethical judgment (WATSON, 2019).

According to WATSON, there is no denying that some of the most innovative achievements in contemporary machine learning are directly or indirectly inspired by prominent theories of neuroscience, cognitive psychology, and social epistemology. Experts and laypeople alike actively promote the notion that these technologies are humanlike in their ability to find and exploit patterns in data. Yet the tendency to focus on structural affinities between biological and artificial neural networks suggests a mechanistic interpretation of intelligence that fails to account for functional complexities. The anthropomorphic tendency in AI is not ethically neutral. The temptation to grant algorithms decision-making authority in socially sensitive applications threatens to undermine our ability to hold powerful individuals and groups accountable for their technologically mediated actions (WATSON, 2019). Algorithms are not just like us and the temptation to pretend they are can have profound ethical consequences when they are deployed in high-risk domains like finance, audit, law

and clinical medicine (EUBANKS, 2018). By anthropomorphizing a statistical model, we implicitly grant it a degree of agency that overstates its true abilities. Algorithms can only exercise their (artificial) agency as a result of a socially constructed context in which we have deliberately outsourced some task to the machine. WATSON brings an important point and specialists have to be careful to manage expectations about what AI can and cannot do.

3.5.8. Toward a Practical Definition of AI

In April 2020, The White House in the U.S. asked researchers to develop machine-learning techniques to quickly analyze nearly 30,000 coronavirus-related studies to better understand the deadly virus. Despite the number of promising projects, however, none of their AI is ready to be widely used today. It will likely take years until the technology is ready to provide tangible results. "I haven't seen anything in which AI has helped us yet, clinically," said Eric Topol, a world-renowned cardiologist, founder and director of the nonprofit Scripps Research Translational Institute and one of the top ten most-cited medical researchers (VANIAN, 2020). What is missing today in intelligent agents that humans do every day is the ability to generalize.

The resurgence of machine learning in the 1980s has led to an interest in formally defining, measuring, and maximizing generalization. Generalization is a concept that predates machine learning, originally developed to characterize how well a statistical model performs on inputs that were not part of its training data. In recent years, the success of Deep Learning (LECUN et al., 2015), as well as increasingly frequent run-ins with its limitations (MARCUS, 2019), have triggered renewed interest in generalization theory in the context of machine learning (see for example NEYSHABUR et al., 2017). The notion of generalization can be formally defined in various contexts but statistical learning theory (VAPNIK, 1999) provides a widely-used formal definition that is relevant for machine learning. We can informally define generalization or generalization power for any AI system to broadly mean the ability to handle tasks that differ from previously encountered situations. The notion of "previously encountered situations" can be broken down as follows:

- System-centric generalization: this is the ability of a learning system to handle situations it has not itself encountered before; and
- Developer-aware generalization: this is the ability of a system, either learning or static, to handle situations that neither the system nor the developer of the system have encountered.

It is also useful to qualitatively define the degree of generalization for information processing systems (CHOLLET, 2019):

- Local generalization, or robustness: This is the ability of a system to handle new points from a known distribution for a single task or a well-scoped set of known tasks, given a sufficiently dense sampling of examples from the distribution (e.g., tolerance to anticipated perturbations within a fixed context). One could characterize it as an adaptation to known unknowns within a single task or well-defined set of tasks. This is the form of generalization that machine learning has been concerned with from the 1950s up to this day.
- Broad generalization, or flexibility: This is the ability of a system to handle a broad category of tasks and environments without further human intervention. This includes the ability to handle situations that could not have been foreseen by the creators of the system. This could be considered to reflect human-level ability in a single broad activity domain (e.g., driving in the real world), and could be characterized as an adaptation to unknown unknowns across a broad category of related tasks. Arguably, even the most advanced AI systems today do not belong in this category, although there is increasing research interest in achieving this level.
- Extreme generalization: This describes open-ended systems with the ability to handle entirely new tasks that only share abstract commonalities with previously encountered situations, applicable to any task and domain within a wide scope. This could be characterized as adaptation to unknown unknowns across an unknown range of tasks and domains. Biological forms of intelligence (humans and possibly other intelligent species) are the only example of such a system at this time.

The intelligent agents that will be analyzed in the next section belongs to the local generalization category. Leveraging on the theoretical concepts analyzed in this section and the Task Formula, intelligence can be defined as a measure of skill-acquisition efficiency over a specific task, with respect to prior knowledge, experience, and local generalization ability. As a result, I submit that artificial intelligence is a non-biological intelligence.

Many possible definitions of intelligence may be valid, across many different contexts, and this study does not pretend that the definition above represents the single version of truth. Nor is that definition meant to achieve broad consensus. Rather, the purpose of this definition is to be actionable for the purpose of detecting fraud risk factors and to serve as a useful perspective to assess the ability of an

intelligent agent to contribute to identify fraud risk factors. This definition captures an important concept: if you consider two systems that start from a similar set of prior knowledge, and that go through a similar amount of experience (e.g., practice time) with respect to a set of tasks not known in advance, the system with higher intelligence is the one that ends up with greater skills (i.e., the one that has turned its prior knowledge and experience into skill more efficiently). This definition of intelligence encompasses meta-learning prior knowledge, memory, and fluid intelligence (CHOLLET, 2019).

3.6. Research Methodology: CPA Quebec Case Study

Qualitative research focuses on understanding how people interpret their experiences, how they construct their world, and the meaning they attribute to their experiences. There are many definitions of qualitative research. A Google search reveals more than 132,000,000 definitions. MAANEN offers one of the most comprehensive definitions of qualitative research: "An umbrella term covering an array of interpretive techniques which seek to describe, decode, translate, and otherwise come to terms with the meaning, not the frequency, of certain more or less naturally occurring phenomena in the social world" (MAANEN, 1979, p. 520).

While researching how intelligent agents can contribute to detect fraud risk factors in artificial intelligence and audit journals, I could not find a case study that encompasses the hypotheses and the research questions in this dissertation for the audit profession in Quebec. Because of the subjective nature of the research questions, I decided to rely on qualitative research, which focuses on understanding the contribution of intelligent agents in the identification of fraud risk factors and the consequences on the learning requirement for the audit profession in Quebec. Because of the exploratory goal, I decided to focus on a case study.

Case study is a significant qualitative strategy, along with critical narrative analysis, phenomenology, ethnography, and grounded theory (MERIAM, 2009). However, case study differs from other research strategies in that it conducts an in-depth analysis of a bounded system. The case in this context is a unit, with defined boundaries, and the bounded system in my research is a business entity. In *Case study research: Design and methods,* YIN (2009) compares case study methods with other forms of research: experimental, survey, archival, analytic, and historical. The author explains that case study research focuses on answering questions that ask how or why, and where the researcher has little

control of events that are happening at present, and when the focus is on contemporary occurrence within some real-life environment. This study meets YIN's criteria. First, based on the Task Formula presented in this dissertation, I want to explore how intelligent agents can improve fraud detection. Second, I want to understand why auditors should leverage intelligent agents to perform their audit work. Third, I do not have control on whether or not an audit firm will decide to leverage intelligent agents to execute audit tasks. Finally, intelligent agents start to impact some tasks of other professions such as doctors and lawyers.

3.7. Unit of Analysis: Algorithms Selection

Based on the research questions, I selected the units of analysis, which is the type of algorithms. When choosing the algorithms, it was important to scope them properly since artificial intelligence is a broad field of research. The current research in artificial intelligence goes well beyond what this study will cover. There is so much active research in all areas of artificial intelligence. There have been and continue to be impressive advances in planning, learning, perception, predictive, natural language understanding, robotics, and other subareas of artificial intelligence. The decomposition of artificial intelligence in subarea is not surprising. The design space is too big to explore all at once. Once a researcher has decided to handle, say, predictive analytics or relational domain and reasoning about the existence of objects, it is difficult to specialize in other areas.

To select the proper algorithms, it is important to understand what an algorithm is. Algorithms are precise and unambiguous instructions that tell computers exactly what to do. Designing algorithms is difficult, time-consuming and often counterintuitive. When programmers and computer scientists succeed in writing good algorithms, they build on each other's work, producing more and more algorithms, which interact like the elements in an ecosystem. Open AI is a good example of an open algorithms platform for computer scientists. This study will focus on one type of algorithms – learning agents or machine learning – which is one of the sub-fields in artificial intelligence.

As stated in Section 3.5.8, intelligence is a measure of skill-acquisition efficiency over a specific task, with respect to prior knowledge, experience, and local generalization ability. At the core of this definition is skill-acquisition or learning. The focus in this dissertation is on learning agents. According to DOMINGOS (2015), we can group learning agents into five categories:

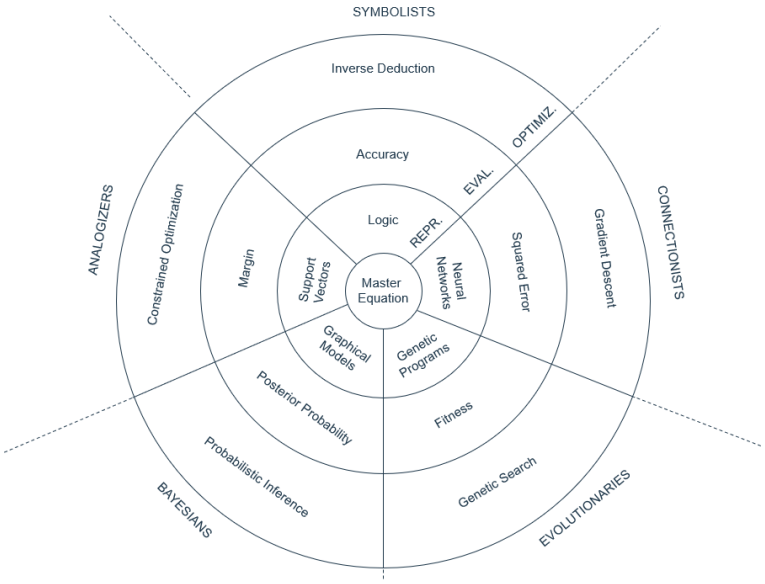

Figure 8. The Five Learning Agent Tribes

Source: DOMINGOS (2015)

The problem domain will affect the kind of algorithm needed. Based on the Task Formula, the problem domain in this dissertation is the ability to predict. The focus in this dissertation is on predictive learning agents:

$$\text{Task} = f\ (\text{data} + \text{prediction} + \text{judgment} + \text{action})$$

Symbolic artificial intelligence is the term for the collection of all methods in artificial intelligence research that are based on high-level symbolic (human readable) representations of problems, logic and search. Symbolists recognize that learning can't start from scratch. Symbolists scientists include pre-existing knowledge in their model. Symbolic AI was the dominant paradigm of AI research from the mid-1950s until the late 1980s. The symbolists' family tree traces back to philosopher David Hume, who asked a profound question: How can you generalize from what you've observed to what you haven't experienced? All learning algorithms seek to find a solution to this query. Some 250 years after Hume asked his question, physicist David Wolpert created the no free lunch theorem, which gave birth to knowledge creation by using what you already know, but also includes random chance. It offers positive examples of each concept for the learner to follow and negative examples of things that don't illustrate the concept. To get a learner to identify cats, you'd add positive examples of cats and negative examples of animals that are not cats, such as dogs. One popular form of symbolic AI is expert systems, which uses a network of production rules. Production rules connect symbols in a

relationship similar to an if-then statement. The expert system processes the rules to make deductions and to determine what additional information it needs, i.e., what questions to ask, using human-readable symbols. JOHN HAUGELAND gave the name GOFAI (Good Old-Fashioned Artificial Intelligence) to symbolic AI in his 1985 book *Artificial Intelligence: The Very Idea*, which explored the philosophical implications of artificial intelligence research.

Connectionism presents a cognitive theory based on simultaneously occurring, distributed signal activity via connections that can be represented numerically, where learning occurs by modifying connection strengths based on experience. Scientist Donald Hebb explained a key element of brain function in 1949, when he showed that repeated activity in one neuron sparks activity in nearby neurons – a principle often summarized as neurons that fire together wire together. Connectionists use algorithms to simulate a brain. Computers don't have as many connections as the brain, so faster processing must compensate. The brain might use 1,000 neurons, but computers would use the same wire a thousand times. Connectionists reverse-engineer the brain to create machine learning. Backpropagation is their main approach. This approach compares the output from a system with the output you want and changes the connections one layer of neurons at a time, improving the output each time. The success of deep learning networks in the past decade has greatly increased the popularity of the connectionists, but the complexity and scale of such networks have brought with them increased interpretability problems. Connectionism is seen by many to offer an alternative to classical theories of mind based on symbolic computation, but the extent to which the two approaches are compatible has been the subject of much debate since their inception.

Evolutionary scientists see natural selection as the engine for learning. Evolutionaries use genetic programming as their main algorithm: they evolve computer programs in much the same way that organisms evolve in nature. Genetic algorithms use mechanisms inspired by biological evolution, such as reproduction, mutation, recombination, and selection. Candidate solutions to the optimization problem play the role of individuals in a population, and the fitness function determines the quality of the solutions. Evolution of the population then takes place after the repeated application of the above operators. Evolutionary algorithms often perform well approximating solutions to all types of problems because they ideally do not make any assumption about the underlying fitness landscape. Techniques from evolutionary algorithms (EA) applied to the modelling of biological evolutions are generally limited to explorations of microevolutionary processes and planning models based upon

cellular processes. In most real applications of EAs, computational complexity is a prohibiting factor. In fact, this computational complexity is due to fitness function evaluation. Fitness approximation is one of the solutions to overcome this difficulty. However, seemingly simple EA can solve often complex problems, therefore, there may be no direct link between algorithm complexity and problem complexity.

Bayesians see learning as a specialized use of the Bayes' theorem. Reverend Thomas Bayes created an equation for incorporating new evidence into existing beliefs. Bayesians recognize the inherent uncertainty and incompleteness of all knowledge. They see learning as a form of uncertain inference. Their challenge is separating data from their surrounding noise and building systems that can deal with incompleteness. If the data support a hypothesis, you give the hypothesis more weight. If the data contradict it, you give the hypothesis less weight. Words are not the best tool for presenting this reasoning, because people neglect key steps in evaluating reasoning. Trying to integrate multiple chunks of evidence adds complexity. People deal with this by compromising and simplifying their evaluation process until it is workable. A machine learner applying Bayes is a Naïve Bayes classifier. The name recognizes a key point: Bayes' theorem starts from a naïve assumption, like how two symptoms of the flu correlate. Search engines use algorithms like Naïve Bayes to make basic assumptions about the terms that people search for most often. Vision learning and spam filtering are also some of the classic problems tackled by the Bayesian approach.

Analogizers see recognizing similarities as central to learning. Their challenge is determining just how alike the two compared things might be by using, amongst others, the support vector algorithm. While neural networks played a larger role in the early years of machine learning, analogy offers exciting possibilities in machine learning. Analogizers offer one of the best learning algorithms: nearest neighbour. This works so well because it does nothing. You don't calculate anything. You just compare the new thing you encounter with records of existing objects in your database. If you want a machine to recognize faces, don't define face. Instead, compare the new image to other pictures of faces. This reasoning works for online recommendations of books or movies. If you like X, you might like Y. You can modify this system to give more weight to some correlations or similarities because your wishes resemble those of one recommender more than they tap into the suggestions of another. The problem with the nearest neighbour algorithm is the curse of dimensionality. The more factors you try to integrate, the more difficult it becomes to use this algorithm.

In the next section, two tribes of algorithms will be put to work: symbolists (rule-based learners) and connectionists (machine learning algorithms). I was given access to nine algorithms developed by MindBridge Ai. For intellectual property reasons, some information cannot be disclosed about these algorithms.

4. RESULTS AND DISCUSSION

To position the problem properly and analyze the contribution of learning agents, my starting point is the Task Formula:

$$\text{Task} = f\ (\text{data} + \text{prediction} + \text{judgment} + \text{action})$$

4.1. Positioning of the Problem

Today, learning agents have their most immediate impact at the prediction level. Before the explosion in AI research, the distinction between prediction and judgment was mainly of academic interest because, as JEFF HAWKINS (2004) explains in his book *On Intelligence*, humans always perform the two together without realizing it. By breaking up the task into components, we can observe where intelligent agents can assist the auditor in performing a task.

To perform their audit tasks, learning agents operate in a unique online environment (real world). Each auditee (a company being subject to a financial audit) is unique. The data subject to an audit by an auditor or a learning agent comes from the general ledger of the auditee. The data can be classified in four categories: assets, liabilities, revenues and expenses. Figure 9 presents a learning agent in a financial audit environment. The analysis in this section will serve as input for the analysis in Section 4.2.

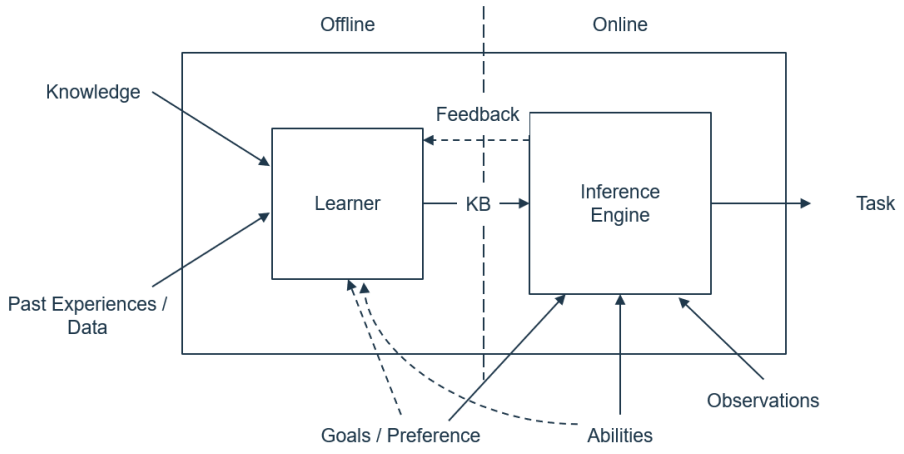

Figure 9. Learning Agent and its Environment

Source: Adapted based on POOL & MACKWORTH (2018)

Offline computation is the computation done by the learning agent before it has to accomplish the task online. The task expected in this dissertation is to find fraud risk factors in the data coming from the general ledger. Offline computation will generally include compilation (or aggregation) and learning.

Compilation (aggregation) happens when the designer takes background knowledge and data and compiles or aggregates them into a usable form called knowledge-base (KB). The data used by the nine algorithms analyzed are structured data (numbers). These data represent assets, liabilities, revenues and expenses. They are the data coming from general ledgers of dummy (fictitious) companies. In a rule-based learning agent, these data will be pushed in. For machine learning algorithms, they will first be pushed in and the learning agent will increase its knowledge-base as it is learning from additional data injected offline.

Knowledge is the information about a domain that can be used to solve tasks in that domain. To solve many tasks requires much knowledge, and this knowledge must be represented in the computer. As part of designing a program to solve tasks, the designer must define how the knowledge will be represented. Typically, we know more about a domain than a database of facts; we know general rules from which other facts can be derived. Which facts are explicitly given, and which are derived is a choice to be made when designing and building a knowledge-base.

Primitive knowledge is knowledge that specifies explicitly in terms of facts. Derived knowledge is knowledge that can be inferred from other knowledge. Derived knowledge is typically specified using rules. The use of rules allows for a more compact representation of knowledge. Derived relations allow for conclusions to be drawn from observations of the domain. This is important because we do not directly observe everything about a domain. Much of what is known about a domain is inferred from the observations and more general knowledge. Building a large knowledge-based system is complex:

- Knowledge often comes from multiple sources and must be integrated. Moreover, these sources may not have the same division of the business environment (or the world). Often knowledge comes from different fields that have their own distinctive terminology and divide the world according to their own needs;
- Systems evolve over time and it is difficult to anticipate all future distinctions that should be made in the business environment; and

- The people involved in designing a knowledge base must choose what relationships to represent in the data.

Online computation is the computation done by the learning agent between observing the data in the general ledger and acting on it – finding anomalies. A piece of information obtained online (in the general ledger of a company) is called observation or a percept. An intelligent agent typically must use its knowledge-base, its beliefs and its observations to determine what to do next with the data – is there a risk of fraud, anomalies or not? When the learning agent is accomplishing the task, it uses its knowledge-base, its observations from the general ledger, and its goals and abilities to choose what to do and use its newly acquired information to update its knowledge base. The knowledge-base is this long-term memory, where it keeps the knowledge that is needed to act in the future. This knowledge is learned from prior knowledge and from data/observations coming from the general ledger and past experiences (data). Online, the information about the particular situation of a company becomes available, and the agent has to act – accomplish the task.

An agent typically has much more time for offline computation than for online computation. During the online computation, it can take advantage of particular goals and particular observations. For a fraud diagnostic, the computational agent has the details of a particular company. Offline, it can acquire more knowledge about how the fraud risk factors interact and do some debugging and compilation/aggregation. It can only do the computation about a particular company online.

To perform its tasks, a learning agent operates in a business environment. An important challenge for the learning agent to operate as expected is the type of environment they will face online. In the computer science world, this is the task environment; the environment in which an intelligent agent acts which range in terms of complexity. RUSSEL & NORVIG (2015) provide the best categorization of task environment characteristics to consider when designing a learning agent.

The environment can be fully observable, partially observable and not observable. A task environment is fully observable if the agent can detect all aspects of the environment that are relevant to the choice of the task; relevance, in turn, depends on the performance measure (objective). A good example of such an environment is playing a chess game. The entire environment – the chessboard – is observable.

If the environment is completely determined by the current state and the action executed by the agent, then we say the environment is deterministic; otherwise it is stochastic. Most real situations are so complex that it is impossible to keep track of all the unobserved aspects; as a result they must be treated as stochastic. Stochastic environment implies that uncertainty about the outcome is quantified in terms of probabilities. A good example of that is a medical diagnostic and the identification of fraud risk factors. A chess game would be considered deterministic. It should be noted that in a nondeterministic environment, actions are characterized by their possible outcomes, but no probabilities are attached to them.

A task environment can be episodic or sequential. In an episodic environment, the agent's experience is divided into atomic periods. In each episode, the agent receives a percept and then performs a single action/task. Crucially, the next episode does not depend on the actions taken in previous episodes. An example would be a computational agent that has to spot defective parts on an assembly line. Each decision on the current part is not impacted by the previous part. In a sequential environment, the current decision could affect all future decisions. A chess game is an example. Episodic environments are much simpler than a sequential environment because the agent does not need to think ahead.

If the environment can change while an agent is deliberating (online computing), then we say the environment is dynamic for that agent; otherwise it is static. Static environments are easy to deal with because the agent needs not to keep looking at the world while it is deciding on an action, nor need it worry about the passage of time. Dynamic environments are continuously asking the agent what it wants to do. If the environment itself does not change with the passage of time but the agent's performance score does, then we say the environment is semi-dynamic. Autonomous taxi driving is a dynamic environment. An environment can also be semi-static like in a chess game or fraud risk factor identification.

Another distinction is whether the environment is discrete or continuous. This applies to the state of the environment, to the way time is handled, and to the percepts and actions of the agent. Taxi driving is a continuous state, continuous time problem and driving actions are also continuous. A chess game is an example of discrete environment. Business operates in a continuous environment, in which an audit takes place.

The last dimension to consider is single versus multiagent environment. Single agent environments are easy to identify. A computational agent solving a crossword puzzle is a single agent environment. Two autonomous vehicles is a multiagent environment, more precisely, a cooperative multiagent environment. Both sides benefit from minimizing the risk of collision. A chess game between two computational agents is a competitive multiagent environment. An audit occurs in a multiagent environment.

The following table summarizes the task environment characteristics. It should be noted that the answer is not always cut and dry. In many cases, the line is blurred when we try to categorize an environment.

Table 5. Task Environment Characteristics

Task environment	Observable	Agents	Deterministic	Episodic	Static	Discrete
Chess game	Fully	Multi	Deterministic	Sequential	Semi	Discrete
Medical diagnostic	Partially	Single	Stochastic	Sequential	Dynamic	Continuous
Taxi driving	Partially	Multi	Stochastic	Sequential	Dynamic	Continuous

Source: RUSSEL & NORVIG (2015)

The quality of the task (output) is another important aspect in AI. Given a well-defined task, the dilemma is whether it matters if the output returned is incorrect or incomplete. For example, if the specification asks for all instances, does it matter if some are missing? There are four common classes of output category (POOLE & MACKWORTH, 2018). An optimal solution to a task is one that is the best outcome according to some measure of solution quality. This is not necessarily desirable since there are significant costs to it and it might be impossible to predict with 100% accuracy. Identifying fraud risk factors with 100% accuracy is not possible as I will demonstrate in Section 4.2. Another option is to develop an intelligent agent that will provide a measure of desirability, known (in economics) as utility. Satisficing outcome is one that is good enough, according to some description of which solutions are adequate. Another option is to have an approximately optimal solution. This is one whose measure of quality is close enough to the best that could theoretically be obtained.

Sometimes, agents do not need optimal solutions to tasks; they only need to get close enough. The last option is to have a probable solution. This is one way to approximate, in a precise manner, a satisficing solution.

4.2. Empirical Analysis

This section addresses the first two hypotheses presented in the Introduction:

1. Intelligent agents are not a substitute for the audit profession in Quebec and cannot result in a massive employment loss.
2. Intelligent agents cannot assume creative cognitive tasks.

4.2.1. Connectionists: Machine Learning Algorithms

The three machine learning algorithms analyzed are all based on supervised learning. In supervised learning, the agent observes some example input-output pairs and learns a function that maps from input to output. More specifically, in supervised learning, there is a set of examples, and a set of features, partitioned into input features and target features. The aim is to predict the values of the target features from input features. A feature is a function from examples into a value. In supervised learning tasks, the learner is given:

- A set of input features, $X_1 \ldots X_n$,
- A set of target features, $Y_1 \ldots Y_n$,
- A set of training examples, where the values for the input features and the target features are given for each example, and
- A set of test examples, where only the values for the input features are given.

The first learning algorithm analyzed is a Rare Flow Control Point. A rare flow is an assessment of how common a flow between different accounts are. Rare flow does not analyze the amount of money between accounts in the general ledger. It examines the frequency of occurrence of the flow relative to all the other flows in a set of data from the general ledger. The rarity of the flow is calculated as follows: each flow or the direction money is moving, is analyzed, grouped and scored based on their rarity in comparison to all other flows. If the money flowing from one account to another account is unusual within a ledger, then flows between these accounts are given a higher score. In contrast, if

money flowing from an account to another account is common within a ledger, then the flows between these accounts are given a low score. This analysis is performed in a company general ledger (online data) – all transactions for a specific fiscal year end. Offline data for this algorithm comes from the designer of the algorithm – called account ontology.

How is flow calculated on a larger transaction? A larger but balanced transaction has ten total entries: one debit and nine credits. The Rare Flow Control Point flattens the transaction into nine separate microtransactions where each credit is associated to the main debit entry. Each microtransaction is then examined as a separate flow and compared at each account level. After each of the nine microtransactions have been analyzed and scored, the microtransactions are rolled up into a rarity score for the original transaction.

In an even more complex example, a transaction has many debits and credits. The rare flow looks to first match debits and credits by dollar values. If matching dollar values are present, the flow(s) are analyzed and set aside. If unmatched entries remain, a composite account is used for the specified debit amount and is analyzed as a microtransaction. Each microtransaction is then rolled up into a rarity score for the transaction.

What is the value of the Rare Flow Control Point during an audit? By analyzing transactions for rare flows, auditors can quickly analyze the uncommon flows which could be caused by fraudulent behaviour or could violate internal controls. Rare flows can indicate unique insights which could require additional investigation or inclusion into the audit. For example:

- Rare Flow and end of year indicate a rare flow at the end of the fiscal year. This transaction includes an uncommon flow at the end of the fiscal year which could be a coincidence or could indicate fraudulent behaviour through an unusual transaction to impact financial statement reporting; and
- Rare flow and cash to bad debt conversion indicate a transaction that contains a rare flow as well as an entry that includes an amount that matches a credit to a cash account with an equal amount debited to the bad debt account. If the rare flow includes the entry associated with the cash to bad debt entry, the transaction is uncommon within the G/L and could also be related to a fraudulent chain of transactions via cash to bad debt.

The second learning algorithm analyzed is looking for outlier anomaly – the Outlier Anomaly Control Point. It uses a combination of nearest-neighbour and stochastic outlier selection (SOS) to create neighborhoods (clusters) for flows within a general ledger and then identifies outliers that aren't grouped within a neighborhood. If a flow cannot be classified in a neighborhood, it is flagged as anomalous and triggers the outlier anomaly control point. The following inputs are used to determine neighborhoods and outliers:

- Dollar amount of the monetary flow;
- Source and destination accounts of the monetary flow;
- Number of flows that occur alongside this flow in this flow's transaction; and
- The proximity of this flow's transaction to the end of the month (time).

In general terms, the nearest-neighbour and SOS techniques find the anomalous points in data sets by asking what points are the closest neighbours to this point? Then they ask, would this point's neighbours consider it to be a neighbour as well? If the answer to the second question is no, then the point is anomalous.

SOS also considered many details of the transaction which can be easily overlooked by a human auditor. The algorithm can find something superficially like other transactions but which differ in some detail that sets that transaction apart and makes it unusual.

In a simplified example, utilizing the SOS and nearest-neighbors methodology described above, the Outlier Anomaly Control Point has grouped a series of flows into a neighborhood based on the flows between accounts, monetary values, number of flows and proximity to end of the month. The neighborhood is centralized on flow 4 and its nearest neighbours include 1, 2, 3, 5, 6 and 7. Flows 1, 2, 3, 5, 6 and 7 will, in turn, classify each other as a neighbour. However, although flow 8 would consider flow 5 a neighbour, flow 5 does not consider itself as a neighbour to flow 8. Flow 8 is then classified as an outlier.

What is the value of the Outlier Anomaly Control Point in an audit? The outlier anomaly provides a powerful lens for identifying flows that contain a combination of uncommon monetary value, flows, frequencies and timing within a month. This control can help pinpoint flows that fall outside of the

normal course of a company's operations or accounting controls. The Outlier Anomaly Control Point in combination with:

- Cash expenditures can indicate an uncommon flow, amount or frequency in which a cash account is being credited. This is important because the way in which the cash is being distributed is uncommon within the general ledger; and
- Manual entry indicates an outlier flow that was also entered manually and was not included in a batch transaction. The origin of the transaction was from an employee and the result was an outlier. If user information is also present in the general ledger data, the transaction information can be compared to other employees to indicate anomalous behaviour.

The third learning algorithm analyzed is looking for unusual amounts. The Unusual Amounts Control Point looks at all entries associated to an account code and determines which amounts are anomalous based on their proximity to neighbouring amounts. The analysis begins at the lowest level of the account structure and is then examined at other levels of the account structure. The assessment of unusual amounts is scoped to the accounts involved in the journal entries. Because of this, the Unusual Amounts Control Point will find amounts that do not normally occur in specific account interactions. This is important because different business processes can cause vastly different amounts in each account so the account focus is essential. Each entry receives a continuous 0-100% anomaly score at each account level, then the maximum score is used in the overall calculation of risk.

What is the value during an audit? The value of the Unusual Amounts Control Point is in determining how uncommon entries are within a ledger and increasing the risk associated with the transactions to increase the likelihood of being sampled. Although an anomalous amount is not suspicious on its own, examining the combination of the unusual amounts control point with the Rare Flow Control Points identify entries that are both anomalous and part of a monetary flow that is uncommon. This ensemble insight is valuable because it identifies both the amount and the flow as being uncommon.

4.2.2. Symbolists: Rule-Based Learners

Six rule-based algorithms have also been analyzed. For these types of algorithms, the developer manually encoded the knowledge required to accomplish a task.

The first algorithm analyzed, the Unbalanced Debits and Credits Control Point, is a transaction level analysis which sums the debits and credits associated with the transaction and identifies differences between the sums. The control point quickly identifies which transactions are unbalanced for further review. Reviewing any unbalanced transactions are an important step in the data validation process as well. Let's analyze the following example. A company uses an ERP system that has limited controls and allows users to enter unbalanced transactions. The unbalanced transactions have caused an issue during previous audits in which determining both sides of a general ledger entry has been difficult. Using the Unbalanced Debits and Credits Control Points can help identify the erroneous transactions and, in the event that the general ledger file balances, can be used to help match the unbalanced transactions to create complete transactions. What is the value during the audit? The Unbalanced Debits and Credits Control Point is critical to understanding if a general ledger is unbalanced and which transactions may be unbalanced. Unbalanced transactions can either be generated by mistake as part of the entry into the ERP system, or could be entered deliberately to hide a portion of a transaction. It is also possible that unbalanced transactions could be caused through the creation of transaction IDs during the MindBridge import process. In either case, unbalanced transactions should be examined and the cause determined. Examining the combination of the Unbalanced Debits and Credits Control Points with end of a period or year-end accounting entries could indicate a deliberate effort to conceal a portion of a transaction which could have a material impact on financial statements.

The second algorithm analyzed is the Zero-Entry Control Point. This algorithm triggers on entries with $0 listed in its debit and credit and can be particularly helpful in identifying the source of unbalanced transactions and determining the quality of the ERP export being used for analysis in the AI Auditor. Zero entries can indicate an accidental or deliberate exclusion from a transaction, but could also be caused during the export or conversion of financial data. If zero entries are included as part of the general ledger data, the entries should be reviewed with the audit company to determine if the cause was within the ERP system. Some ERPs allow rules which can cause additional empty entries; for example, tax as part of a sales journal where no tax was applicable. What is the value during the audit? Zero entry can provide feedback on the quality of the data being analyzed and quickly identifies entries that may have been entered incompletely or by accident. Additional value is present when the Zero-Entry Control Point is triggered with the Unbalanced Debits and Credits Control Points as the zero entry could be the cause of the unbalanced transaction.

The third algorithm analyzed is the Weekend Post Control Point. This algorithm triggers based on the posted date associated with entries occurring during a weekend. The control point is of particular value for organizations that have stricter controls over when transactions can be entered into the general ledger. What is the value during an audit? The weekend posting can identify transactions which are being entered at potentially abnormal times. Although an abnormal posting time does not mean the transaction is suspicious by itself, for organizations with stronger controls or regular business hours, the Weekend Post Control Point can be an important insight when identifying higher risk transactions. Examining the combination of the Weekend Post Control Point with cash expenditures flag transactions where cash or cash equivalents have been credited during a weekend. The timing of these transactions should be considered within the context of the company's normal operations to determine their inclusion in the audit plan.

The fourth algorithm analyzed is statistical by nature: the Benford's Law. Benford's Law was discovered by Frank Benford during the 1930s when examining a book of logarithmic tables. Benford noticed the wear on each page within the book wasn't evenly distributed and that pages beginning with the digit 1 were more worn than pages beginning with digits 2-9. After analyzing the distribution of numbers across a great number of subjects including atomic weights, baseball statistics and the areas of river, he published the article on Benford's Law. Benford's Law has a distribution that includes the following distribution of leading digits: 1 = 30.1%, 2 = 17.6%, 3 = 12.5%, 4= 9.7%, 5 = 7.9%, 6 = 6.7%, 7 = 5.8%, 8 = 5.1%, 9 = 4.6%. The frequency of each first and two digit combination is then counted and mapped against the Benford distribution. How is Benford's Law applied for an audit? Given the complexity of general ledger data, all entries are analyzed within a two-digit distribution of Benford's Law. The observed counts of the first two digits in the general ledger are computed against the expected counts of the first two digits using the Benford probability. If the difference between the observed and expected counts is found to be significant, the Benford's control point is triggered. Although the full analysis of the general ledger will provide a greater likelihood of returning false positives, the application of Benford's Law, in conjunction with the ensemble of other control points, still provides a meaningful analysis and risk profile. What is the value of Benford's Law during an audit? Benford's Law can provide insight into anomalous patterns of entry and transaction data, which can represent falsification of accounting data.

The fifth algorithm analyzed is the cash to bad debt conversion. The Cash to Bad Debt Conversion Control Point flags matching dollar amounts from credited cash and cash-equivalent accounts to debits in the bad expense account. This control point does not look at transactions or flows: it looks at the dollar value of specific entries. If a bookkeeper or an accountant is committing fraud, they may disguise the movement of a monetary asset through multiple transactions as a single transaction from a cash equivalent to a bad debt. The Cash to Bad Debt Conversion Control Point triggers when multiple transactions or a single transaction is used, provided the dollar values are matching. Will this return false positives? In general, a direct conversion from cash to bad debt is a problem requiring review. As this control point is performing number matches, false positives are possible. Potential relationships between triggered entries may not be immediately obvious, which is another important reason to consider the examination of these entries. What is the value of the Cash to Bad Debt Conversion Control Point in an audit? The value is that auditors can quickly identify entries that could be involved in a fraudulent chain of transactions. Although false positives will be present, examining the triggered entries should be considered as part of an audit.

The last algorithm analyzed is the Complex Instrument Control Point. This algorithm flags transactions that appear to be complex in nature by examining the memo field for specific keywords. The default configuration of the Complex Structure Control Point is configured to look for transactions that are complex in nature including forward contracts and options. The default keywords include: fair value, guarantee, embedded derivative, net settlement, fix for fix, forward contracts, swap, option, taps, callers, hedge, hedging, commodities, host contracts, forward options, re-commission, extinguishment, modifications and transaction cost. What is the value during an audit? The combination of the complex instrument with two other algorithms can provide higher value insights:

- The combination of the complex instrument and Rare Flow Control Points indicate an uncommon monetary flow within the transaction that could be unique or significant based on the complex instrument keywords entered; and
- The combination of the complex instrument and high monetary value will quickly identify transactions that are within the top two percentile of the general ledger but also include specified keywords requiring further analysis. The combination ensures the transactions being viewed are material.

4.2.3. Analysis

The following table summarizes the contribution of the nine intelligent agents to the 160 sub-tasks identified in Appendix 2.

Table 6. Contribution of the Nine Intelligent Agents

Contribution to ten sub-tasks	32, 33, 34, 35, 36, 37, 38, 39, 121, and 151
Scope out – no contribution	150 sub-tasks

Source: Author's Compilation

For the specific tasks they are built for, the nine intelligent agents analyzed provide useful insight to the auditor. The quality of the task is considered as satisficing, meaning it is not optimal, but it provides direction to the auditors about what to pay attention to. They cannot make a comprehensive assessment (generalization) like an auditor can do. To do so, the auditor must address the other 150 sub-tasks in Appendix 2. As we have seen in Section 2.3, there are many reasons why a person decides to commit a fraud and cases from Enron to SNC-Lavalin demonstrated that this type of crime is carefully planned and difficult to discover. Discovering a fraud involves going through many of the seven steps in the Task Complexity Framework.

Sub-task 1 in Appendix 2 is a good example. No algorithm can evaluate the values and integrity of an executive in relation to his title and function in a company and assess if these values increase the risk of fraud. To do so, you need a broad range of information (steps 1 and 2 in the Task Complexity Framework), conduct interviews and understand the state of mind of the executive (step 4 in the Task Complexity Framework), then figuring out if, for example, the three components of the fraud triangle – opportunity, justification, pressure – (step 5 and 6 in the Task Complexity Framework) are present and can lead to an increase in the risk of fraud. Based on all the evidence collected, the auditor will have to assess the risk of fraud. The auditor's judgment can direct him to further investigate if some risks exist. Abstraction and generalization play an essential role in the auditor cognitive process. A lot of what we know is fairly abstract. The representations that underlie both cognitive models and common sense are all built on a foundation of a rich collection of such abstract relations combined in complex structure. Humans can abstract just about anything: time, features, theories, space and so forth and use them in a sentence, an explanation, a comparison, stripping hugely complex situations down to their essentials and giving the mind enormous leverage in reasoning broadly about the world.

Machine learning thus far has struggled with open-ended inference. Auditors, as they read texts and read numbers, frequently derive wide-ranging inferences that are both novel and only implicitly licensed. At present, there is no machine learning system that can draw open-ended inferences based on real-world knowledge with anything like human-level accuracy. Basically, they cannot ask the question who, what, why, when, where and how. What is missing in machine learning is common sense reasoning.

The analysis of the nine intelligent agents demonstrates that they are not a substitute for the audit profession in Quebec and they cannot result in a massive employment loss because of the complexity of the audit work. They also cannot assume creative cognitive tasks. Interestingly, intelligent agents don't have to work like humans. Based on the Task Formula:

$$\text{Task} = f\ (\text{data} + \text{prediction} + \text{judgment} + \text{action})$$

intelligent agents working with auditors can improve the quality of an audit. The nine algorithms analyzed are faster and better at finding inference (predict) in structured data – accounts in the general ledger – and auditors are better at contextualizing the outcome. All the time an auditor would have spent on doing the work of the nine algorithms analyzed previously can be reallocated to value-added work (the other 150 sub-tasks) and, therefore, increase the quality of the audit.

One of the promises of intelligent agents in auditing is to provide composite and panoramic views of accounting data which can improve the identification of fraud risk factors. There is, however, an important limit that auditors must understand to properly rely on intelligent agents: bias. The key challenge for auditors is to understand the nature of biases that can be built in connectionist machine learning algorithms.

Bias is not a new problem; rather bias is as old as human civilization. As I presented in Section 2.2, human bias has many facets and bias is known to be an impediment to fair and right decisions and task execution in many domains. Bias is also an old concept in machine learning (MITCHELL, 1997). In the field of machine learning, the term bias seems to be used in different contexts and with different meanings (CAMPOLO et al., 2017). Terminology shapes how we identify, analyze and resolve problems. Since there is no generally accepted definition for bias in AI, this study will analyze bias along the lifecycle of the three connectionist machine learning agents analyzed in Section 4.2.1. The

following analysis builds on the analysis presented in Section 4.1. Figures 10 and 11 summarize the key components of the lifecycle and the most prominent biases I have identified.

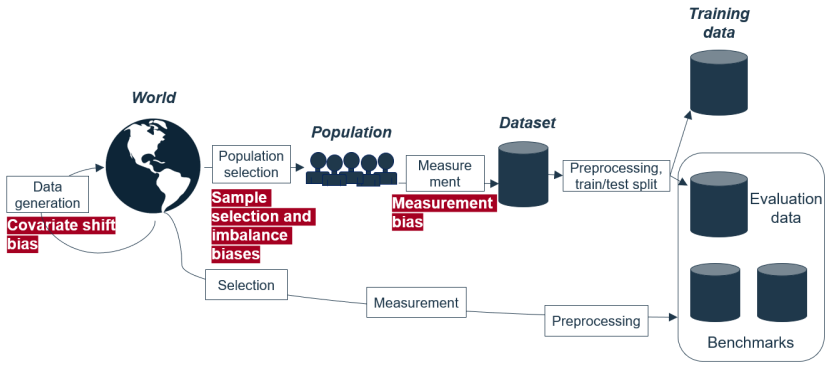

Figure 10. Data Generation

Source: Adapted from SURESH & GUTTAG (2020)

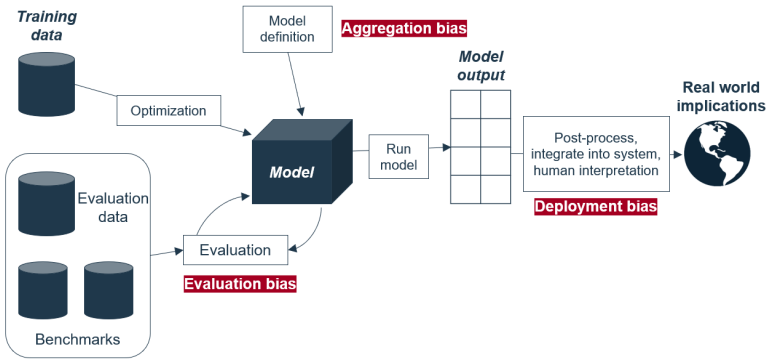

Figure 11. Model Building Implementation

Source: Adapted from SURESH & GUTTAG (2020)

Connectionist learning agents rely heavily on data generated by humans (user-generated content) or collected via systems created by humans. Therefore, whatever biases exist in humans enter in the system and, even worse, they can be amplified depending on the complexity of the model and the sources of data. If the data or the decisions taken on it are biased and the machine uses them as an example, then the machine is going to incorporate this bias into the model. It learns the bias from the examples given to it.

Generally speaking, data is biased if the sampling distribution (the data which we use for training the model) is different from the population distribution (referring to the true situation in the real world). Putting it another way, to avoid bias we have to make sure that the data sample that we use for training

the model resembles as closely as possible the true distribution of the features and the decisions taken on them.

Covariate shift occurs if one of the features is not covered uniformly in the dataset. Stated differently, it refers to the change in the distribution of the input variables present in the training and test data. This bias exists because of the limited real-world data that Mindbridge Ai has access to. There are two important challenges to have access to real-world data: these data are commercially sensitive, and the company can be reluctant to share them. The other challenge relates to data privacy compliance regulation.

Sample selection bias refers to a correlation between a (subset of) feature(s) and the label. If this correlation only occurs in a set of examples but not in the normal population, our dataset is biased. An example might be that a certain combination of abnormal transactions have only been observed in certain industries. Auditing a company in a different industry could pause a challenge for the intelligent agent to find such anomalies.

Imbalance bias denotes the situation in which there are considerably fewer examples for one specific decision (label) than for the other(s).

Note that in reality, these three types of bias do not necessarily occur separately, but often a biased dataset contains a mixture of these. Not every bias necessarily results from unconscious bias of humans. Sometimes bias occurs naturally. This is why the auditor must understand the concept and the type of bias that exists in the learning agent.

In epidemiology, measurement bias, observational bias, and information bias refers to bias arising from measurement errors (ROTHMAN & al., 2008) i.e., errors occurring in the process of making observations of the world. Measurement bias is rarely mentioned in machine learning. IBM defines measurement bias as bias that occurs when the data collected for training differs from the data collected during production (IBM, 2019). VANDERWEELE & HERMAN (2012) argue that measurement bias arises because proxies are generated differently across groups (also known as differential measurement). Two main reasons why measurement bias can arise are that the measurement process varies across groups (BAROCAS & SELBST, 2016) and the quality of data varies across groups (CALDERONE, 1990).

Aggregation is the process of collapsing data point and then using a measure of central tendency to represent that range of data (WALKER & CATRAMBONE, 1992). Aggregation bias arises during model construction when distinct populations are inappropriately combined. In many applications, the population of interest is heterogeneous and a single model is unlikely to suit all subgroups. Aggregation bias can lead to a model that is not optimal for any group, or a model that is fit to the dominant population (if combined with representation bias). If there is a non-linear relationship between group membership and outcome, for example, any single linear classifier will have to sacrifice performance on one or both groups. In some cases, incorporating information about group differences into the design of a model can lead to simpler learned functions that improve performance across groups (DWORK et al., 2017).

Evaluation bias occurs when the evaluation and/or benchmark data for an algorithm don't represent the target population. A model is optimized on its training data, but its quality is often measured on benchmarks (HUANG et al., 2007). A misrepresentative benchmark encourages the development of models that only perform well on a subset of the population. Evaluation bias ultimately arises because of a need to objectively compare models against each other. Applying different models to some set of external datasets attempts to serve this purpose but is often extended to make general statements about how good a model is. Such generalizations are often not statistically valid (SALZBERG, 1997) and can lead to overfitting to a particular benchmark or set of benchmarks.

For a real-world machine learning application, there are many steps that arise when a system is actually deployed. For example, a model may need to be changed based on requirements for interpretability or interactivity, or there may be real-time feedback that should be integrated back into the model. Importantly, there is no guarantee that the population a model sees as input after it is deployed looks the same as the population it saw during training and evaluation. Deployment bias arises when there is a mismatch between the task an intelligent agent is intended to accomplish and the way in which it is actually used. This often occurs when an intelligent agent is built and evaluated as if it were fully autonomous, while in reality, it operates in a complex world moderated by human decision makers. SELBST et al. (2019) refer to this as the framing trap. Each of the three intelligent agents are built for a specific task and if this is not the task actually being carried out after deployment, there is no guarantee that good evaluation performance will carry over. In some cases, a system that produces results that must first be interpreted by human decision makers may actually lead to harmful

consequences because of phenomena such as automation or confirmation bias (GREEN & CHEN, 2019).

Accounting for bias not only requires understanding of the different sources – i.e., data, knowledge-base and algorithms – but more importantly, it demands the interpretation and description of the meaning, potential side effects, provenance, and context of bias. This is one of the reasons that explainability is critical. The concept of explainability will be analyzed in the next section.

4.3. The Social and Moral Dimension of AI: Ethical Challenge

This section addresses the first question of this research presented in the Introduction: What is the main ethical consideration CPA Quebec should analyze and understand as artificial intelligence will penetrate the audit profession in Quebec?

To practice the audit profession in Quebec, auditors must comply with the Code of Professional Ethics. In Canada, all of the Chartered Professional Accountant provincial bodies have their own rules of professional conduct for their members and students. In Quebec, the Code of Ethics has three important components: objective of the audit profession, principles necessary to attain the objective and conformity. The objective of the profession is to serve the public interest. This cannot be accomplished by mere conformity to detailed rule; this dedication to serve is more like a state of mind. The principles necessary to attain the objective can be summarized as follows:

- Integrity;
- Objectivity;
- Professional competence and due care;
- Confidentiality; and
- Professional behaviour (including conformity and technical standards).

The ethical dilemma created by learning agents relates to professional competence and due care. Professional competence and due care require auditors to comply with the Canadian Auditing Standards (CAS). CAS require auditors to obtain sufficient appropriate evidence as the basis for an audit opinion on financial statements.

Appropriateness of evidence relates to its qualitative aspects: is it relevant and reliable? Evidence that is relevant and highly reliable is persuasive as proof regarding the financial statement's assertions. Relevant audit evidence means that it must relate logically to at least one of the financial statement's assertions, otherwise it is not relevant to the auditor. The reliability of audit evidence depends on its nature and source.

Sufficiency considers how much appropriate evidence is enough. The matter of efficiency is an important application of the auditor's professional judgment, as this varies from situation to situation. The standards cannot really set out a specific amount of evidence required. Realistically, however, audit decisions must be based on enough evidence to stand the scrutiny of other auditors and outsiders, judges and inspectors. The real test of sufficiency is whether the body of evidence you have gathered allows someone else to reach the same conclusions you reached. If an auditor has not been able to obtain sufficient appropriate audit evidence, the auditor cannot reach a conclusion.

The first three algorithms analyzed in Section 4.2.1 create a challenge for an auditor with respect to appropriateness, relevance and sufficiency. Auditors face fundamental limits on their ability to trace the inductive reasoning of a complex intelligent system, specifically machine learning. The most significant limitation of the first three algorithms analyzed in Section 4.2.1 is the opaqueness or the lack of explainability, which inherently characterizes them as black box machine learning models. This means that these models' internal logic and inner workings are hidden to the auditors, which is a serious disadvantage as it prevents an auditor from being able to verify, interpret, and understand the reasoning of the system and how particular decisions are made. This is where explainable AI can play a significant role because as long as high-performing models remain opaque, it seems rational to withhold how trustworthy AI really is when conducting an audit, especially in high-risk areas like fraud detection.

Different scientific communities (ABDUL et al., 2018) studied the problem of explaining machine learning decision models. However, each community addresses the problem from a different perspective and provides a different meaning to the explanation. Most of the works in the literature come from the machine learning and data mining communities. The first one is mostly focused on describing how black boxes work, while the second one is more interested in explaining the decisions even without understanding the details on how the opaque decision systems work in general.

Many questions feed the papers in the literature proposing methodologies for interpreting black box systems. What is an explanation? Is it when a model or an explanation is comprehensible? Which is the best way to provide an explanation and which kind of model is more interpretable? Which are the problems requiring interpretable models/predictions? What kind of decision data are impacting? Which type of data records is more comprehensible? How much are we willing to lose in prediction accuracy to gain from interpretability? The running hypothesis is that, by building explainable systems, users will be better equipped to understand and, therefore, trust intelligent agents.

4.3.1. Explainable Artificial Intelligence

To address this challenge, Explainable Artificial Intelligence (XAI) emerged the past few years as a field of study which focuses research on machine learning interpretability and aims to make a shift towards a more transparent AI. There is a large consensus that it is important for AI and machine learning to be interpretable/explainable (see, for example, the 2018 report from the European Commission's High-Level Expert Group on Artificial Intelligence). Since XAI is a relatively new research field, there is no consensus over what is meant by explainable and interpretable (LIPTON, 2016). A recent UK Government report on the state of AI received substantial expert evidence and noted that the terminology used by the experts varied widely (Select Committee on Artificial Intelligence, 2017).

A key challenge in XAI is that the term is connected to numerous other terms such as transparency, accountability, intelligibility, interpretability, fairness, and many more, resulting in defining explainability in various ways. At its core, explainability is about the translation of technical concepts and decision outputs into intelligible, comprehensible formats suitable for evaluation. The T20 Report on the future of work and education, for example, highlights the importance of "clear, complete and testable explanations of what the system is doing and why" (THINK 20, 2018, p. 7). Stated differently, a satisfactory explanation "should take the same form as the justification we would demand of a human making the same kind of decision" (UNIVERSITY OF MONTREAL, 2018, p. 12).

The term XAI was coined by Van Lent et al. (VAN LENT et al., 2004) in order to describe the ability of their system to explain the behaviour of AI-controlled entities in simulation applications. Historically, explanations first appeared in the context of rule-based expert systems and were mostly treated as a system design task. The need for explaining the decisions of expert systems was discussed

as early as the 1970s (SHORTLIFFE & BUCHANAN, 1975). SWARTOUT (1983) described a framework for creating expert systems with explanation capabilities and was one of the first to stress the importance of explanations that are not merely traces, but also contain justifications. LACAVE & DIEZ (2001) present an excellent survey of methods of explanation for probabilistic decision-making systems based on Bayesian Networks (referred to as expert systems and often regarded as successors of earlier rule-based systems). Their work presents an excellent analysis of the methods in terms of several properties of explanation. Of particular interest is their classification of the focus of explanation into an explanation of the reasoning, the model, and the evidence for the decision. Generating explanations for recommendation systems was also a field of research in the early 2000. HERLOCKER et al. (2000) conducted an experiment measuring user satisfaction with a variety of justification types for a collaborative filtering movie recommendation system.

There has been a surge of interest in explainable AI in recent years driven by the U.S. Defense Advanced Research Projects Agency (DARPA). A number of countries have made public the demand for AI explainability/interpretability. This will impact how professions such auditors, lawyers and doctors will use intelligent agents.

The draft version of the Dutch AI Manifesto (created by IPN SIG AI), which focused on explainable AI, states that one of the most important features of AI systems is being not only accurate but also able to explain how the system came to its decision (IPN SIG AI, 2018). The French Strategy for Artificial Intelligence, presented in 2018 by the President of the French Republic, contains a set of proposals in which the first one is to develop algorithm transparency and audits, which entails producing more explainable models (VILLANI, 2018). The Royal Society, which is the United Kingdom's Academy of Sciences, published in 2017 a report on their machine learning project and the report recognizes the importance of interpretability and transparency as well as responsibility and accountability associated with machine learning (ROYAL SOCIETY, 2017). The Portuguese Government, through its National Initiative on Digital Skills, published a draft version titled *AI Portugal 2030* outlining innovation to foster artificial intelligence. The document recognizes the importance of transparent AI as one of the fundamental research lines in the future of AI (PORTUGUESE NATIONAL INITIATIVE ON DIGITAL SKILLS, 2019).

In April 2018, the European Commission published a communication to many official European bodies, such as the European Parliament and the European Council, on Artificial Intelligence for

Europe. In the communication, it stressed the importance of research into the explainability of AI systems to further strengthen people's trust in AI. Furthermore, the communication stressed the importance that AI systems should be developed in a manner which allows humans to understand the basis of their actions in order to increase transparency and to minimize the risk of bias error (EUROPEAN COMMISSION, 2018). In January 2020, as part of the Trump Administration's National AI Strategy – The American AI Initiative – the White House proposed a set of principles to guide how the federal agencies regulate AI in the private sector, characterizing it as an effort to govern AI without stifling innovation. The principles promote the development of trustworthy AI. The principles state that, when considering action related to AI, regulators must consider fairness, transparency, safety and security (PRESIDENT TRUMP, 2020). In April 2019, the High-Level Expert Group on Artificial Intelligence (AI HLEG), which is an independent expert group set up by the European Commission as part of its AI Strategy, published the document Ethics Guidelines for Trustworthy AI (AI HLEG, 2019). The document lists seven key requirements that AI systems should meet in order to be trustworthy: transparency and accountability being two of those key requirements. Explainability is listed as one of the ethical principles in the context of AI systems.

The response by the AI and machine learning communities to address the challenges created by XAI has been strong the past few years with a wide range of conferences and workshops. The following table summarizes the main events since 2016.

Table 7. Scientific Events Focusing on XAI/Interpretability

Conferences/Workshops	Year (s)
FAT Annual Conference	2016-2020
ICML Workshop on Human Interpretability in Machine Learning	2016-18
NIPS Workshop on Interpretable Machine Learning for Complex Systems	2016
NIPS Symposium on Interpretable Machine Learning	2017
XCI: Explainable Computational Intelligence Workshop	2017
IJCNN Explainability of Learning Machines	2017
ICJAI Workshop on Explainable Artificial Intelligence	2017-18
IPMU 2018 – Advances on Explainable Artificial Intelligence	2018
CD-MAKE Workshop on Explainable Artificial Intelligence	2018-19
Workshop on Explainable Smart System (ExSS)	2018-19
ICAPS – Workshop on Explainable AI Planning	2018-19
AAAI-19 Workshop on Network Interpretability for Deep Learning	2019
CVRP – Workshop on Explainable AI	2019

Source: Author's compilation

FAT academics (meaning fairness, accountability, and transparency in multiple artificial intelligence, machine learning, computer science, legal, social science, and policy applications) is a prominent actor pursuing XAI. The group's primary focus is on promoting and enabling explainability and fairness in algorithmic decision-making systems with social and commercial impact. The most recent FAT conference (its seventh annual conference was held in January 2020) attracted more than 500 participants (researchers and practitioners) and published 95 papers.

Another prominent actor in XAI is DARPA. DARPA launched its XAI program in 2017 with the aim of developing new techniques capable of making intelligent systems explainable, the program includes 11 projects and will continue running until 2021. DARPA funded researches focused primarily on increasing explainability in sophisticated pattern recognition models needed for security applications. Even though DARPA is funded by the US Department of Defense, the program involves researchers drawn from various academic institutions and diverse corporate teams.

Increasing interest in XAI has also been observed in the private sector community. Companies on the cutting edge of contributing to make AI more explainable include Microsoft with its next generation of Azure (Azure ML Workbench), Kyndi with its XAI platform for government and financial services such as Fair Isaac Corporation (FICO) with its Credit Risk Models. To push the state of XAI even

further, FICO is running the Explainable Machine Learning Challenge (xML challenge). The goal of this challenge is to identify new approaches for creating machine learning based AI models with both high accuracy and explainability.

In February 2020, The Pontifical Academy for Life, an advisory body to Pope Francis, drafted a charter on artificial intelligence ethics that is being supported by International Business Machines Corp. and Microsoft Corp. The Charter, called the *Rome Call for AI Ethics*, looks to ensure that AI is developed and used to serve and protect people and the environment. The document calls for AI education and regulation and outlines a set of six principles that define the ethical use of AI. The ethical use of AI, according to the document, is defined by six principles, the first one being transparency, which addresses the need that all artificial intelligence systems be explainable.

As we can observe, interest in explainability in different research communities is increasing; however, a large part of that work is rather recent and has often not appeared in peer-reviewed journals. Many research papers in that field are uploaded on the arXiv.org, a distribution service and an open-access archive for 1,653,012 scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics.

In theory, having an algorithm clearly explain how it makes a decision would be the most effective and direct way to ensure an algorithmic system is acting as intended; stated differently, that it is trustworthy. Developing an algorithm capable of explaining itself or justifying its decision is an incredibly challenging technical feature, to the point that DARPA devoted $75 million in 2017 to research how it could be achieved (KUANG, 2017). For sure, building intelligent agents capable of explanation is a challenging task but it cannot be done in a vacuum considering only the computational problems, because trustworthiness in AI will not materialize. Recent surveys have emphasized the multidisciplinary, inclusive nature of the process of making an AI-based model interpretable. Along this process, it is of utmost importance to scrutinize and take into proper account the interests, demands and requirements of all stakeholders interacting with the system to be explained, from the designers of the system to the decision makers consuming its produced outputs and users undergoing the consequences of decisions made therein.

4.3.2. Explanation

What explainable AI means is a very complex question to answer. This research will first equate explainable to explanation. The Oxford English dictionary defines explanation as "a statement or account that makes something clear." The Cambridge Dictionary of English Language defines explanation as "the details or reasons that someone gives to make something clear or easy to understand." From these two definitions, it is clear that an explanation is a three-value predicate: someone (a communicator), explains something, to someone (the recipient) (HILTON, 1990). The success of an explanation, therefore, depends on several critical recipient factors: knowledge, assumptions, interests, and bias that the recipient has when decoding the explanation. While performing a scan of literature, I noticed that most of the research in the field of XAI focus on the communicator side – the developers of algorithms. The Montreal Declaration for a responsible development of artificial intelligence states that an explanation in the context of AI "should take the same form as the justification we would demand of a human making the same kind of decision" (UNIVERSITY OF MONTREAL, 2018, p. 12). Explanation should, therefore, have a focus on the recipient of it.

Technically, there are no standards and generally accepted definitions of explainable AI and we may never have such a generally accepted definition because explainable AI, I submit, should focus on the recipient of the explanation. Only a few papers focus on such an approach and none related to the audit profession. This section explores to whom a machine learning system might be explainable for the CPA ecosystem. While others previously identified that explainability should be considered with reference to a specific user or user group (KIRSCH, 2017), I argue that a useful framework should focus on the recipients' ecosystem that interact with, or is affected by, the machine learning system. Recipients in the CPA ecosystem have different beliefs and goals depending on their roles in relation to the machine learning system. The proposed framework will guide the analysis of what their relevant beliefs and goals might be for specifying suitable measures of explainability.

Before outlining the proposed framework, I will define the scope of it. The proposed framework is built around a machine learning system, by which I mean a system that includes one or more machine learning models, as analyzed in Section 4.2, the data used to train the model(s), any interface used to interact with the model(s), and any relevant documentation. The machine learning is monolithic or comprised of several different services owned by different entities, situated in different locations, and

trained on data from many sources. The system is situated in a machine learning ecosystem (or just ecosystem), which includes the system and the recipients that have interactions with, or are affected by, this system. An ecosystem always contains just one machine learning system and one or more agents (in the real world, ecosystems will often overlap).

Since the recipient is a human, it makes perfect sense to rely on cognitive science and philosophy to start analyzing XAI. Indeed, it is a reasonable hypothesis to assume that a human will regard an intelligent agent (machine learning system) as an intentional agent – an agent making decisions, offering a prediction or a recommendation – and, therefore, apply some components of the conceptual framework and psychological mechanisms of human behaviour explanation to them (GRAAF & MALLE, 2017). There are vast and valuable bodies of research in philosophy, psychology, and cognitive science about human behaviour's explanation – how people define, generate, select and evaluate explanations. The discussion hereafter builds on that literature.

"What is an explanation" has generated of a lot of debate in philosophy and cognitive science. Research in this field stresses the importance of causality in explanation. In *The Book of Why*, PEARL & MACKENZIE (2018) demonstrate the importance of cause and effect in explanation. JOSEPHSON & JOSEPHSON (1996) define explanation has an assignment of causal responsibility. There are many other definitions, and some are not related to cause and effect and it is out of scope of this dissertation to review and analyze all of these definitions. This study starts with JOSEPHSON & JOSEPHSON's definition and I will complement it with other important concepts from cognitive science. As an assignment of causal responsibility, an explanation can be seen as a product and a process (LOMBROZO, 2009). LOMBROZO argues that there are two processes: a cognitive process and knowledge transfer process.

According to LOMBROZO, as a product, an explanation can be defined as an answer to a why-question (LOMBROZO, 2009). Answering a why-question instead of a what-question or a how-question is more challenging since it is an open question that requires more reasoning; this is what makes XAI a complex problem because the recipients of the outcome – a prediction (risk of fraud) ask that open question – Why this factor? DENNETT (2017) argues that why-questions are ambiguous because there are two senses: How come? and What for? The former asks for a process (narrative), without an explanation of what it is for, while the latter asks for a reason, which implies some intentional thought behind the cause.

Although a thorough analysis of the subject of explanation as a cognitive process would have to cover literature spanning the entire history of Western philosophy from Aristotle onward, some important and relevant outcomes of research in cognitive psychology and philosophy must be analyzed. HARMAN (1965) argues that how people explain things is by using a cognitive process called "inference to the best explanation" (HARMAN, 1965, p.88). This concept corresponds [approximately] to what some researchers will call eliminative induction, theoretical inference, hypothetical inference and method of elimination. PIERCE (1903) labels this process abductive reasoning and was the first researcher to reject the idea that the roster of acceptable types of inference includes just the two classical types: deduction and induction. More recently, a number of experimental evaluations have reinforced that humans are using this cognitive process (see, for example, LOMBROZO, 2012; WILLIAM et al., 2013). In general, there will be several hypotheses which might explain a fact, so one must be able to reject all alternative hypotheses before one is warranted in making an inference. There is, of course, a selection problem about how a person is to judge that one hypothesis is sufficiently better than another hypothesis. Presumably, such a judgment will be based on considerations such as past knowledge and experience. Personal bias may enter the selection process as well. We can decompose the abductive reasoning process as follows:

Table 8. Pierce's Decomposition of Abductive Reasoning Process Applies to an AI Scenario

Process	Requirements
Observation of an event (e.g., a risk fraud factor – red flag).	Abduction is a trigger: the observation of something is interesting or surprising. The perception of the event hinges on the recipient's knowledge and experience.
Generation of one or more possible explanations for the recommended red flag.	The understanding of the red flag hinges on the recipient's knowledge and experience. The derivation of the explanatory rule is a creative action.
Judging the plausibility of the AI explanation.	Abduction is the search for a satisfying explanation. Of course, judgment plays a role in the selection process but the process is based on rationalist considerations of necessity and sufficiency. The process is not necessarily based on estimation of probabilities. The process can be influenced by personal bias.
Resolving the explanation.	The plausibility judgment typically (though not necessarily) results in a determination that a particular explanation is preferred.
Extending the explanation.	Abduction involves going beyond the formation of a rule to the empirical testing of the rule. The determination of a preferred explanation is always tentative, it is subject to disconfirmation by further evidence. There is an accompanying expectation that further instances will conform to the preferred explanation.

Source: Adapted for an intelligent agent scenario from HOFFMAN & KLEIN (2016)

Researchers in AI spend more time now trying to understand the abductive reasoning process and how it can explain observations such as fraud prediction or a medical diagnostic. POPLE (1973) seems to be one of the first researchers to have tried to encode the process in a suitable computational form.

An explanation can also be seen as a knowledge transfer process. "Explaining promotes learning by requiring the integration of novel information with prior beliefs and knowledge" (LOMBROZO, 2006, p. 468.). ALEVEN & KOEDINGER (2002) argue that explanations that merely identify relevant principles improve learning. O'REILLY et al. (1998) argue that explanations that do not relate novel information to prior beliefs are less effective. During that process, persuasion plays a critical role. One needs to be persuaded that the knowledge received is relevant to answer the why-question. Persuasion could be another factor impacting the assessment of an intelligent agent trustworthiness. But there are some risks in trying to persuade someone. One may be persuading about the relevance of an explanation even though it is a wrong explanation due to personal bias. We can see here that the goal of the explainer (persuading and generating trust) could be different than the goal of the recipient (understanding the decision).

How humans generate explanation provides valuable insight for XAI. An important finding in cognitive science is the concept of contrastive explanation. HILTON argues that "one does not explain events per se, but that one explains why the puzzle event occurred in the target cases but not in some counterfactual contrast case" (HILTON, 1990, p. 67). Some scientific researchers name it counterfactual case (LOMBROZO, 2012; HESSLOW, 1988). What we know from contrastive explanation theory is that people do not explain the cause for an event per se but explain the cause of an event relative to some other event that did not occur. An explanation is therefore answering a why-question in the form of why A rather than B? Explaining a contrastive question is simpler than providing a full causal attribution to a why-question (LIPTON, 1990). Open questions require more knowledge, more reasoning, more time to answer and sometimes we don't know the full spectrum of all the causes of a phenomenon or an event.

Another important finding in philosophical work is the concept of causal chain. In answering a why-question, there may be a number of sequential causes to the answer. A causal chain is a path of causes between a set of events. A cause from event M to Q would be first that O occurs, then P occurs (HILTON et al., 2005). In some situations, people don't need to understand the complete causal chain of an explanation to a why-question. While the aim of this study is not to provide a detailed survey on causality, it is important to note that there are two other important theories surrounding the causal chain: causal attribution and causal explanation. Psychological research into attribution began with the work of Fritz Heider in the early 20th century, and the theory was further advanced by Harold Kelley and Bernard Weiner. The process of trying to determine the causes of people's behaviour is known as causal attribution (HEIDER, 1958). Extracting selective causes from a causal chain is causal attribution. Causal attribution is not necessarily a[n] [causal] explanation, but people may form their own explanation by extracting relevant causes from the entire chain. For example, one may not understand the causal chain of a specific fraud but can explain the fraud by extracting some causes from the chain. This theory is important for explainable AI.

Contrastive explanation and causal chain theories together provide useful insight on explainable AI. Algorithms are developed in closed [theoretical] environments (Figure 9) and the real world is different than these pre-defined environments. The training and test data used to make the intelligent agent represent a sample of what the real world is. This implies that the causal chain of an event (a prediction) would be smaller and maybe less cognitive demanding for the recipient to understand.

Explaining a contrastive question would be less demanding for an intelligent agent rather than providing an entire causal chain and this can be an opportunity for explainable AI. Contrastive questions will also point an intelligent agent directly to what the end-users are not understanding in the model. LIM & DEY (2009) found that in the context of intelligent agents, Why not? questions were common questions asked by end-users.

"How do we generate explanation" was a field of research and interest for HEIDER (1958). HEIDER advanced the concept of social attribution, commonly known as person perception. Leveraging on HEIDER's work, MALLE & KNOBE (1997) discovered that people's intentional behaviour is often contrasted with unintentional behaviour to provide an explanation. In the same vein, KASHIMA et al. (1998) demonstrated that people use folk psychology – belief, desire and intention (BDI framework) – to generate explanation. GRAAF & MALLE (2017) assert that people will expect explanation from an intelligent agent to follow the same conceptual framework used to explain human behaviours, thus reinforcing the importance of folk psychology and, specifically, the BDI framework in explainable AI. In a knowledge-based AI system, BDI framework can be used to build the knowledge required for the computation agent. For example, desire can be built in an AI planning system or AI prediction system, such as the identification of fraud risk factors.

Information asymmetry impacts how people select explanation. In contract theory and economics, information asymmetry deals with the study of decisions in transactions where one party has more or better information than the other. This asymmetry creates an imbalance of power in transactions, which can sometimes cause the transactions to go away, a type of market failure. Information asymmetry extends to non-economic behaviour. As "private firms have better information than regulators about the actions that they would take in the absence of a regulation, the effectiveness of a regulation may be undermined" (FULLERTON & WOLFRAM, 2012, p.11). MALLE et al. (2007) argue that actors and observers offer different explanations for the same action by an actor based on what they call "actor-observer asymmetries". Due to information asymmetry, observers cannot access the intention of the actor; the intentions, therefore, must be inferred.

Some other important findings from cognitive science with respect to how we select and evaluate explanation must be highlighted. HILTON & SLUGOSKI (1986), with their abnormal condition model, state that abnormal events play a key role in causal explanation. Some scientists (LIPTON, 1990; LOMBROZO, 2010) argue that necessity and sufficiency are two important criteria to assess

the quality of an explanation. In his theory for explanatory coherence, THAGARD (1989) argued that coherence is a primary criterion for explanation. He proposed seven principles of how explanations relate to prior belief. He argued that all things being equal, simpler explanations (those that cite fewer causes – which is consistent with the causal chain theory) and more general explanation are better explanation. HILTON (1996) argues that the most likely of true cause is not always the best explanation, challenging the importance of probability as a criterion to select an explanation. MCCLURE (2002) also challenged the importance of probability as a criterion to assess the quality of an explanation. His studies found that people tend to judge the quality of an explanation based on pragmatic influence and not probability. Which means people judge an explanation based on usefulness and relevance. KULESZA et al. (2013) did interesting research on the importance of completeness and soundness as criteria to select and evaluate an explanation. Their studies concluded that completeness was more important. Interestingly they also highlighted that an explanation should not be overwhelming, which is contradictory with completeness.

Based on the above analysis, an explanation can be seen as an assignment of a causal responsibility, implying that it is a three-value predicate. An explanation is an answer to a why-question, a cognitive process and it serves as a transfer of knowledge. In this cognitive process, researches demonstrate that humans use contrastive explanation, select relevant causes (causal chain), probability is not a significant criterion, abnormal conditions matter, coherence, completeness, necessity, sufficiency simplicity, usefulness and folk psychology are important factors to consider. These findings contribute to define a framework for explainability based on recipients. One element encapsulates the aforementioned analysis: explanations are contextual. This aspect is important for explainable AI since, as discussed in Section 4.2, an intelligent agent (machine learning system) cannot contextualize as humans do, this is why it is called machine learning and not machine understanding. Machine learning thus far has not been well integrated with prior knowledge like a human.

4.3.3. Proposed Framework to Reach Explainability

As a next step, I will propose to extend the discussion on human explanation to explainable AI. If a machine learning model performs well enough and has an acceptable predictive performance, why do we not just trust the model and disregard why it made a certain decision? DOSHI-VELEZ & KIM (2017) answer this question by stating that the problem is that a single metric, such as classification accuracy, is an incomplete description of most real-world tasks. Stated differently, the need for

explainability arises from an incompleteness in problem formalization. This concept is related to the information asymmetry problem discussed previously.

We learned from the cognitive science research discussed previously that an explanation is not all-purpose and must be contextualized. One of the reasons behind this unsolved XAI problem is that explainability is a subjective concept and hard to formalize. Like human explanation, explainable AI is domain-specific and we cannot expect to come up with an all-purpose definition. Depending on the context, different types of explanations might be useful. For example, one might want to personally know the main two reasons why a mortgage was declined by the bank, but in a legal scenario, a full explanation with a list of all factors might be required (causal explanation and selection of relevant causes).

In his seminal book *The Inmates are Running the Asylum: Why High-Tech Products Drive Us Crazy and How To Restore The Sanity*, ALAN COOPER (2004) argues that a major reason why software is often poorly designed (from a user perspective) is that programmers are in charge of design decisions, rather than interaction designers. As a result, programmers design software for themselves, rather than for their target audience; a phenomenon he refers to as "the inmates running the asylum". Explainable AI risks a similar situation if the focus is not on the recipient of the explanation. But explainable AI is more likely to succeed if researchers and practitioners understand, adopt, implement, and improve models from the vast and valuable bodies of research in philosophy, psychology, and cognitive science; and if evaluation of these models is focused more on people than on technology.

I conducted a literature review by examining relevant papers on explainable AI from five major academic databases: SCOPUS, IEEExplores, ACM Digital Library, Google Scholar and Science Direct, in addition to preprints posted on arXiv. Keywords-based search was used to select relevant papers; it consists of the terms explainability, interpretability, intelligibility, transparency, understandability, comprehensibility. Based on this literature review, I developed the following conceptual framework to reach explainability:

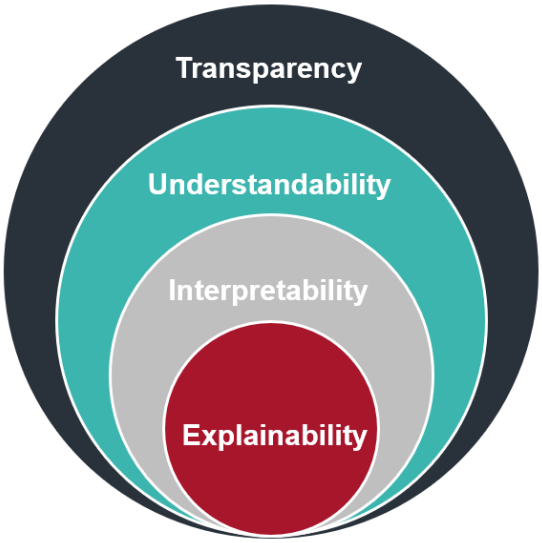

Figure 12. Framework to Reach Explainability

Source: Author's framework

- Transparency: the level to which a system provides information about its internal workings or structure, and the data it has been trained with (LIPTON, 2016). A model is considered to be transparent if, by itself, it is understandable. If a model is transparent, it can be understood. Transparency does not necessarily mean that the underlying information is easily comprehensible by everybody.
- Understandability: denotes the characteristic of a model to make a human understand its function (MONTAVON, SAMEK, & MÜLLER, 2018). If we can understand a model, it is possible to interpret the outcome.
- Interpretability: the level to which a recipient gains and can make use of the information embedded in a machine learning system.
- Explainability: I submit that if a model is transparent, understandable and interpretable, it can be explained. An explanation is a product (answering the why-question) and a process (a cognitive process and knowledge transfer process); this is why it is an active characteristic between an intelligent agent and a human. This is consistent with LIPTON's (2016) definition of explainability – the level to which a system can provide clarification for the cause of its outputs.

An explanation can be classified by the timing of its requirement: pre-model, in-model or post-model. Transparency, understandability and interpretability are three conditions required for pre-model and

in-model. Explainability is related to a post-machine learning model. These concepts are analyzed later in this section.

Explainability is not always a requirement. DOSHI-VELEZ & KIM (2017) argue that there are two kinds of situations where interpretability (and, as a result, explanations as well) are not necessary: (1) when there is no significant impact or severe consequences for incorrect results and (2) when the problem is well studied enough and validated in real applications that we trust the system's decisions, even if the system is not perfect. Those two situations are not applicable for predicting the risk of fraud due to the magnitude of the consequences if a fraud really happened.

Based on the explored literature and the professional responsibility of the auditor in Quebec, I have defined five different recipients of an explanation of a machine learning system output in the audit ecosystem in Quebec. The roles are not mutually exclusive: a single recipient could occupy any combination of roles and some combinations are more likely than others. Currently, this study assumes these roles are fulfilled by humans. However, intelligent agents may increasingly occupy some of them in future, especially if intelligent agents gain rights and start to be used by regulatory entities for examples to understand how an algorithm makes its prediction.

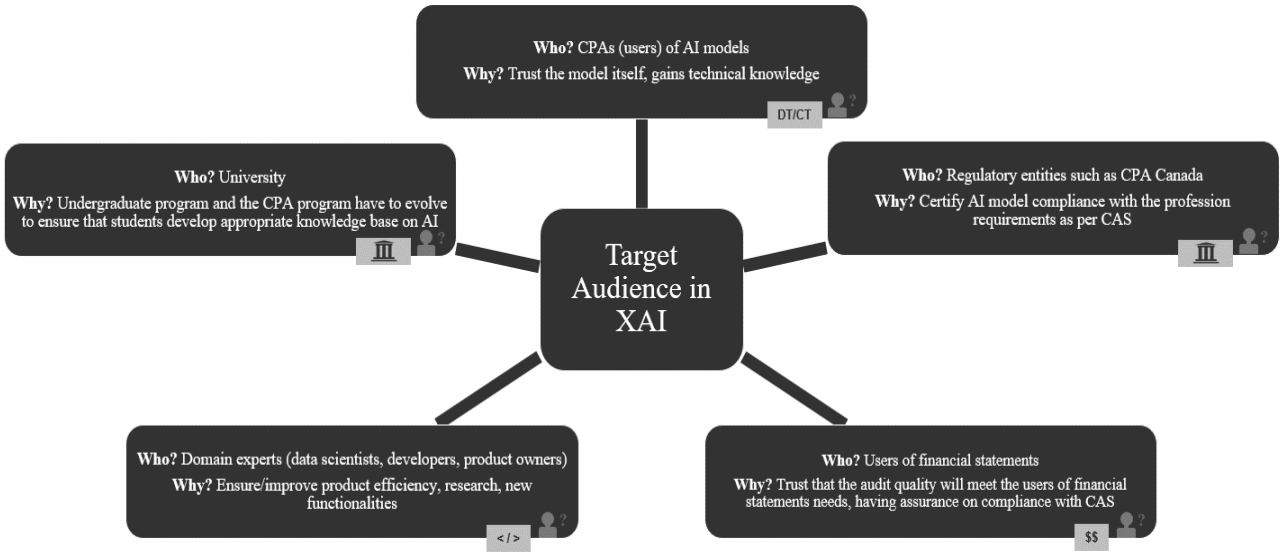

Figure 13. XAI Recipients' Framework

Source: Author's framework

The recipients of an explanation have different beliefs and goals depending on their roles in relation to the machine learning system. LEAKE (1995) demonstrated that goal-directed explanations in

abductive reasoning explicitly aim at reducing knowledge gaps (information asymmetry), specifically, to explain why an observed event is reasonable. This is consistent with LOMBROZO (2009) definition of an explanation – a knowledge transfer process and good for learning. The goal is important because the more a recipient will learn from an intelligent agent, the less ambiguity will be in his mind and the more trustworthy an intelligent agent should become over time. I will next analyze the goals of the proposed XAI recipients' framework.

Domain experts are concerned with building intelligent systems. Members of this community are in the industry (business world); some could be academics or researchers creating systems. Their primary motive in XAI is the quality of the product, i.e., system devolvement, testing, robustness, evaluation. Universities (theorists) are concerned with understanding properly AI to teach it to students in the CPA program and advancing theories. Some will also be active CPA practitioners/users. It is a group that will advance the state-of-the-art technology by bridging CPAs' needs/requirements and current AI applications. Regulator is a broad category that includes CPA Canada, CPA Quebec, lawyers, scientists, and the government (since it regulates the CPA profession in Quebec). They have many motives for XAI: accountability, verifiability of the quality of an audit, compliance with CAS, unbiased behaviour and transparency. Users are the CPAs using the intelligent system. Members of that community need explanations to help them decide whether/how to act given the outputs of the system, and/or to help justify and document their audit work. Building on LIPTON's work (2016) and my extensive literature review, goals pursued toward reaching explainability for the proposed recipients' framework can be summarized as follows:

Table 9. XAI Goals, Recipients and Related Social Science Theories

Goals	Recipients	References
Trustworthiness	All	LIPTON (2016), MURDOCH et al. (2019), SRINIVASAN et al. (2018), RIBEIRO et al. (2016), FOX et al. (2017)
Causality	All	LIPTON (2016), MURDOCH et al. (2019), SRINIVASAN et al. (2018), TICKLE et al. (1998), LOUIZOS et al. (2017),
Informativeness	All	LIPTON (2016), MURDOCH et al. (2019), SRINIVASAN et al. (2018), VELLIDO et al. (2012), RIBEIRO et al. (2016), HARBERS et al. (2010), LANGELEY et al. (2017), SAMEK et al. (2017), MARTEENS et al. (2011), TICKLE et al. (1998)
Confidence	Domain experts, regulatory entities	MURDOCH et al. (2019), THEODOROU et al. (2017), SAMEK et al. (2017)
Accessibility	Domain experts and users	SRINIVASAN et al. (2018), VELLIDO et al. (2012), RIBEIRO et al. (2016), HARBERS et al. (2010), MARTEENS et al. (2011), KRAUSE et al. (2016)
Transferability	Domain experts	LIPTON (2016), SRINIVASAN et al. (2018), VELLIDO et al. (2012), RIBEIRO et al. (2016), HARBERS et al. (2010), LANGELEY et al. (2017), THEODOROU et al. (2017), SAMEK et al. (2017), MARTEENS et al. (2011), KRAUSE et al. (2016), TICKLE et al. (1998), LOUIZOS et al. (2017)
Interactivity	Domain experts and users	SRINIVASAN et al. (2018), HARBERS et al. (2010), LANGELEY et al. (2017), KRAUSE et al. (2016)

Source: Author's compilation

Causality: although supervised learning as we have seen in Section 4.2 is only optimized directly to make associations, researchers often use them in the hope of inferring properties or generating hypotheses about the natural world. Several authors argue that explainable models might ease the task of finding relationships that, should they occur, could be tested further for a stronger causal link between the involved variables (WANG et al., 1999; RANI et al., 2006). The inference of causal relationships from observational data is a field that has been broadly studied over time (PEARL, 2009). As widely acknowledged by the community working on this topic, causality requires a wide frame of prior knowledge to prove that observed effects are causal. A machine learning model only discovers correlations among the data it learns from and, therefore, might not suffice for unveiling a cause-effect relationship. However, causation involves correlation, so an explainable machine learning model could validate the results provided by causality inference techniques or provide a first intuition of possible causal relationships within the available data.

Informativeness: machine learning models are used with the ultimate intention of supporting decision-making (HUYSMANS et al., 2011). Hence, a great deal of information is needed in order to be able to relate the user's decision to the solution given by the model, and to avoid falling into misconception pitfalls. For this purpose, explainable machine learning models should give information about the problem being tackled.

Confidence: as a generalization of robustness and stability, confidence should always be assessed on a model in which reliability is expected. The methods to maintain confidence under control are different depending on the model. Stability is a must-have when drawing interpretations from a certain model (RUPERT, 1987; BASU et al., 2018; YU et al., 2013). Trustworthy interpretations should not be produced by models that are not stable. Hence, an explainable model should contain information about the confidence of its working regime.

Accessibility: a minor subset of the reviewed contributions argues for explainability as the property that allows end users to get more involved in the process of improving and developing a certain machine learning model (CHANDER et al., 2018; MILLER et al., 2017). It seems clear that explainable models will ease the burden felt by non-technical or non-expert users when having to deal with algorithms that seem incomprehensible at first sight.

Transferability: models are always bounded by constraints that should allow for their seamless transferability. This is the main reason why a training-testing approach is used when dealing with machine learning problems (KUHN & JOHNSON, 2013; JAMES et al., 2013). Explainability is also an advocate for transferability, since it may ease the task of elucidating the boundaries that might affect a model, allowing for a better understanding and implementation. Similarly, the mere understanding of the inner relations taking place within a model facilitates the ability of a user to reuse this knowledge in another problem. There are cases in which the lack of a proper understanding of the model might drive the user toward incorrect assumptions and fatal consequences (SZEGEDY et al., 2013; CARUANA et al., 2015). Transferability should also fall between the resulting properties of an explainable model, but again, not every transferable model should be considered as explainable.

Interactivity: some specialists in the field of machine learning (HARBER et al., 2010; LANGLEY et al., 2017) include the ability of a model to be interactive with the user as one of the goals targeted by an explainable machine learning model. Once again, this goal is related to fields in which the end

users are of great importance, and their ability to tweak and interact with the models is what ensures a model is explainable.

4.3.4. Explanation Techniques

Explanation is closely linked to evaluation of an AI system. Since an explanation is an interface between an intelligent system and a stakeholder, the ecosystem in Figure 13 will benefit from explanations at two different levels: intrinsic and post-hoc. Although the following proposal is not a comprehensive list, it is a good starting point for the audit profession in Quebec and as the profession will start leveraging more on intelligent systems, the list will evolve with real life experience.

The intrinsic intelligent systems explanation can be referred to as what the software engineering world call "verification". It is about building the intelligent system right: a glass box approach which is essential because it matters greatly how the intelligent system is built. Intrinsic explanation is also called transparency (in Figure 12) as the term "glass door" implies.

Post-hoc intelligent systems explanation is often referred to as "validation": building the right system. At a very high level, this is what we can call a "reverse reengineering technique" – given the prediction records produced by the intelligent system, a decision maker can reconstruct the explanation.

At risk of simplifying my analysis, we can say that domain experts and theorists (university) communities tend to focus more on verification, the former because they want a system that is built right, and the latter because they are interested in understanding how various kinds of machine learning systems work, and what are their theoretical limits. End users and regulatory entities are more focused on validation, being more concerned with what the intelligent system does than how it is built.

There are three relevant levels of intrinsic explanation (transparency) for the target audience in Figure 13.

- Simulatability: this can be referred to as a holistic explanation. According to LIPTON (2016), a model is transparent if you can comprehend the entire model at once. This suggests that it is a simple model. In practice, it means that a human should be able to take the input data together with the parameters of the model and, in reasonable time, step through every calculation required to produce the prediction. This accords with the common claim that sparse

linear models, as produced by lasso regression (TIBSHIRANI, 1996), are more interpretable than dense linear models learned on the same inputs. Simple but not extensive rule-based systems, such as the ones analyzed in Section 4.2.2, fall in this category. Extensive rule-based (a large amount of rule) fall out of this characteristic. Under this definition, an interpretable model is one that can be easily presented to a stakeholder by means of text or visualizations (RIBEIRO et al., 2016). Endowing a decomposable model with simulatability requires that the model be self-contained enough for a human to think and reason about its whole.

- A second notion of transparency is decomposability to explain each part of a model (input, parameter, and calculation). This accords with the property of intelligibility as described by LOU et al. (2012). For example, each node in a decision tree might correspond to a plain text description. Similarly, the parameters of a linear model could be described as representing strengths of association between each feature and the label. Note that this notion of interpretability requires that inputs themselves be individually interpretable, disqualifying some models with highly engineered or anonymous features. The weights of a linear model might seem intuitive, but they can be fragile with respect to feature selection and pre-processing. For example, associations between weekend transactions and fraud could be positive or negative depending on whether the feature set includes indicators such as amounts, unusual account titles, frequency requirements, etc.
- The third one is algorithm transparency. This one applies at the level of the learning algorithm itself (LIPTON, 2016). It can be seen in different ways. It deals with the ability of the user to understand the process followed by the model to produce any given output from its input data. Put differently, a linear model is deemed transparent because its error surface can be understood and reasoned about, allowing the user to understand how the model will act in every situation it may face (JAMES et al., 2013). Contrarily, it is not possible to understand it in deep architectures as the loss landscape might be opaque (DATTA et al., 2016), since it cannot be fully observed and the solution has to be approximated through heuristic optimization (e.g., through stochastic gradient descent). The main constraint for algorithmically transparent models is that the model has to be fully explorable by means of mathematical analysis and methods.

Explainability can be achieved by a descriptive overview of how the algorithm functions, somewhat like a user's manual would explain the functioning and limitations of a car. This type of XAI allows

the end users to understand and interpret a model (Figure 12). The IEEE proposes a list of minimum information that might be included in such a user's manual (IEEE, 2019, p. 245):

- Nontechnical procedural information regarding the employment and development of a given application of autonomous and intelligence systems;
- Information regarding data involved in the development, training, and operation of the system;
- Information concerning a system's effectiveness/performance;
- Information about the formal models that the system relies on; and
- Information that serves to explain a system's general logic or specific outputs.

Although it is not in the scope of this research to cover in detail the explanation techniques (since it is a complex problem and could be the topic of a full dissertation), additional comments are warranted. Post-hoc interpretability presents a distinct approach to extracting information from learned models. Post-hoc explainability targets models that are not readily interpretable by design by resorting to diverse means to enhance their interpretability, such as text explanations, visual explanations, local explanations, explanations by example, explanations by simplification and feature relevance explanations techniques. Each of these techniques covers one of the most common ways humans explain systems and processes by themselves.

One advantage of this concept of interpretability is that we can interpret opaque models after-the-fact, without sacrificing predictive performance. GUIDOTTI and al. (2018) analyzed 54 techniques for reverse re-engineering explanation. Post-hoc approaches can be primarily classified into two groups: (i) approaches that perturb the input to create multiple input-output pairs and then fit a simple model to explain the predictions locally, or (ii) approaches based on saliency maps which assign importance scores/attribution to each input feature.

4.4. Professional Competence

This section addresses two research questions and one hypothesis. Sub-sections 4.4.1 and 4.4.2 address the 2nd Question presented in the Introduction:

- What are or could be the impacts of artificial intelligence on both the content of the curriculum to access the chartered professional accountant profession in Quebec and the reskilling requirements?

Section 4.4.3 addresses the 3rd Question and the 3rd Hypothesis presented in the Introduction:

Question

- What could be the role of the CPA Quebec ecosystem (government, professional order, universities and firms) in learning in the age of artificial intelligence?

Hypothesis

- The regional audit ecosystem is playing a key role in collective learning and developing ethical and moral regulations for the future of the audit profession in Quebec.

As accounting departments in organizations enter the digital age (an era of exponential change) the audit profession both in Quebec and globally is experiencing unique and unprecedented challenges and opportunities. The digitization of financial transactions is quickly transforming the landscape and nature of audit work. This means embracing new technologies, new auditing methods, and acquiring new competencies and skills become an imperative agenda for the audit profession in Quebec.

The audit profession in Quebec is still designed to meet the needs of the industrial age. With exponential shifts in technology, globalization, business models, geopolitics, and societal values and norms, it is time for transformation, otherwise the profession risks falling behind, losing relevance amongst users, constituencies and future talent, and being replaced by competitors. This was also echoed in the UK in the *Report of the Independent Review into the Quality and Effectiveness of Audit* (BRYDON REPORT, 2019). The audit profession in Quebec, like the rest of the world, is facing a number of important challenges:

- A swiftly evolving digital environment led by the overwhelming pervasiveness of technological change;
- The speed and scope of digitization in corporations, including the impact of new technologies used by corporations such as AI and blockchains;
- Changes in the needs of users and the move to real-time data in making decisions;
- The exponential increase in data combined with a lack of standards related to data governance and integrity; and
- The need for trust and ethics in the information age as people and organizations struggle to understand what information can be relied on.

For more than a century, the profession in Quebec has been built on its ability to synthesize vast amounts of corporate transaction information. The result was ordered, thoughtful reporting by CPA auditors that enabled evaluation of performance to date and served as a platform to consider the prospects for performance. However, the dawn of the digital age demands an evolving view:

- The perspective on what constitutes performance must broaden substantially beyond financial aspects to also consider operational metrics, Environmental, Societal and Governance (ESG) factors and other dimensions for which stakeholders in organizations want to have reporting; and
- The traditional mindset of looking back in time to report on what has already occurred must be reoriented to a real-time and forward-looking point of view. New and emerging technologies will combine with the digitization of corporate information and allow real time automated reporting. This will support much more sophisticated modelling of what will occur in the future. Hindsight has traditionally been the predominant focus of the audit profession in Quebec, with limited priority to foresight. The audit profession today is fundamentally restricted to assuring the material accuracy of historical financial information and, even with this restricted scope, is only partially meeting even that objective. Indeed, many corporate frauds remain undetected for years and, when they are identified, it is too late since their financial consequences are dramatic due to their magnitude. The recent fall of China's Luckin Coffee is an example. Luckin sold vouchers redeemable for tens of millions of cups of coffee to companies that had ties to Luckin's controlling shareholder, Charles Lu. Their purchases

boosted the company's revenue; revenue that never materialized. Overnight, the stock price of the company fell by 75% on the Nasdaq Stock Market in New York (YANG, 2020).

When it comes to where the profession in Quebec will have to play in the future, there are three primary areas:

- Mastering and shaping a data-driven economy. Every auditor must ultimately become comfortable in a world that is data-rich, data-intense and data-driven and possesses the skills requirement associated with artificial intelligence;
- Rethinking value creation by not only providing a report on historical information but also providing insights into the future of a company. Is the company at risk of going bankrupt over the next few years? What are the key risks that could compromise the business model of the company? and;
- Trust: Auditors are stewards of the public trust, therefore, the quality of the auditor work must be improved to better detect white-collar crime.

4.4.1. CPA Quebec Competency Map

To practice the audit profession in Quebec, students must meet the requirements of The CPA Competency Map.

The CPA program in Quebec is designed to meet the needs of public accounting, industry, and government by ensuring that all CPAs have a strong foundation of ethics, knowledge and skill to succeed and lead in any professional accounting role or position. The Competency Map describes the competencies for all the elements of the CPA program in Quebec. The Competency Map:

- Helps guide candidates in understanding what is expected of them when enrolled in the CPA professional education program;
- Establishes the body of competencies developed through an integrated certification process that includes education, evaluation and experience;
- Provides guidance to post-secondary educators and program developers for the further development of learning objectives for the professional education program modules; and
- Provides guidance to employers for the further development of competency objectives for practical experience.

The following figure summarizes the path to the CPA audit certification in Quebec:

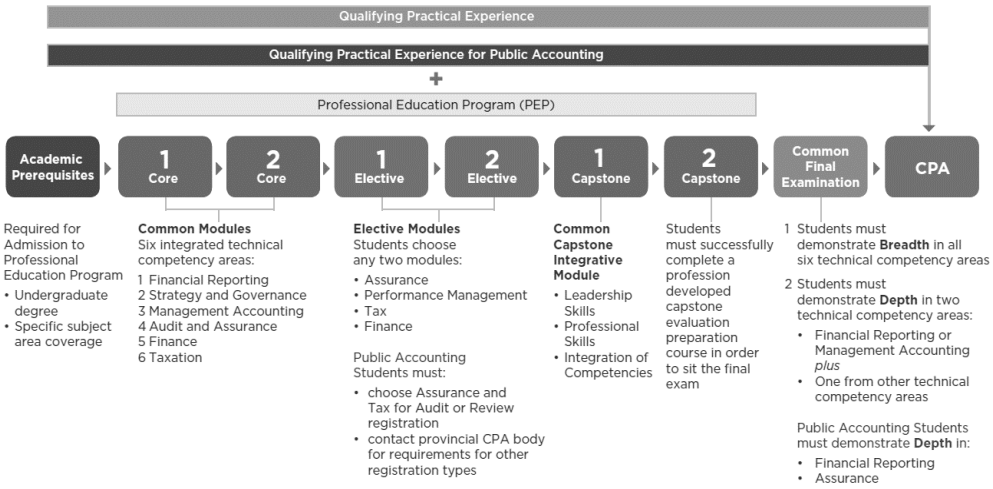

Figure 14. The Path to Certification

Source: CPA Canada.

The CPA certification program comprises the following:

- Prerequisite education: there are academic prerequisites for admission to the CPA certification program. Before entering the program, candidates must complete an undergraduate degree and cover specific subject areas. The specific subject areas may be covered during the undergraduate program, or through additional courses offered by universities, colleges, or various bridging programs;
- CPA professional education program (CPA PEP): CPA candidates must complete the CPA Professional Education Program (CPA PEP) or its equivalent, through accredited programs. It consists of a series of modules that develop professional competence. Ethics and other enabling competencies, and prerequisite subject matter in areas such as IT, are integrated throughout the program;
- Practical experience: relevant practical experience enhances the education component of the CPA program. Completion of the professional education program may run concurrently with the period of practical experience; and
- Common final examination (CFE): in addition to formative examinations throughout the program, the CPA certification program culminates in a summative final examination that evaluates candidates on the competencies defined by The Competency Map.

Figure 14 provides an overview of the technical competency area that students must possess to become an auditor. There are six technical competencies in the CPA curriculum in Quebec: financial reporting, strategy and governance, management accounting, audit and assurance, finance and taxation.

In addition to the technical competencies, students in the CPA program in Quebec must develop and acquire seven enabling competencies. The CPA enabling competencies provide the essential skills for ethical behaviour, leadership, teamwork, decision-making, problem-solving, and communication as a professional accountant.

- Acting ethically and demonstrating professional values: the CPA profession is grounded in ethics, professionalism and protection of the public interest. CPAs have a duty to their profession and to society, as well as to their individual and organizational interests. They do more than adhere to the CPA Code of Professional Conduct; CPAs' ethical behaviour exemplifies and enhances the reputation of the profession;
- Leading: CPAs recognize and promote their strategic role within an organization;
- Collaborating: CPAs are respected and trusted, enabling them to partner with individuals and teams throughout an organization;
- Managing self: central to the CPA culture is a commitment to continuous learning and professional development;
- Adding value: CPAs add value to their organizations, community and society;
- Solving problems and making decisions: CPAs draw on strong problem-solving and decision-making skills, including the ability to utilize technology and data analytics; and
- Communicating: CPAs ensure that their communications are effective when speaking, listening, presenting and writing in one of Canada's two official languages. They ensure that meaning is conveyed clearly and succinctly by attending to the needs of diverse audiences and selecting the most appropriate communication media. CPAs have the ability to tell the story of the business when presenting information.

Although the current CPA certification program in Quebec includes enabling competencies, the focus of the undergraduate studies (prerequisite education in Figure 14) in Quebec is on the six technical competencies – specialized technical knowledge. Based on the Task Formula:

$$\text{Task} = f\ (\text{data} + \text{prediction} + \text{judgment} + \text{action})$$

The CPA Competency Map in Quebec is missing two critical components: first, developing students' professional judgment and second, developing the skill requirements associated with artificial intelligence.

4.4.2. Creative Thinking

The audit profession in Quebec must rethink its value creation. Improving the ability of an auditor to better find white-collar crimes is one of the areas of value creation. Creative thinking (professional judgment) will be critical to be successful in that area. To discover a white-collar crime, you have to think like a white-collar criminal.

The current CPA certification program in Quebec focuses on specialization (technical competencies). This is not the only profession. The professed necessity of hyperspecializing forms the core of a vast majority of professions – lawyers and doctors are good examples. But there are some exceptions. As intelligent agents' contribution to find white-collar crime increases, developing the creative thinking of the students in the CPA designation program will become critical for the profession to increase its value creation.

Plenty of experts argue that anyone who wants to develop a skill or lead in their field should start early, focus intensively and rack up many hours of specialized learning and training in their field. But a closer look at research shows that this is an exception, not the rule, and intelligent agents are reviving the debate about specialization.

Psychologist ROBIN HOGARTH coined the term "kind learning environment" (HOGARTH, 2001) for an environment in which golf and chess operates. Patterns repeat over and over again and feedback is extremely accurate and usually very rapid. In chess, a piece is moved according to rules and within defined boundaries, a consequence is quickly apparent, and similar challenges occur repeatedly. This environment well suits the ten-thousand-hour rule (GLADWELL, 2011); technical training (tactics), specialized learning, early start, pattern recognition. The learning environment is kind because a learner improves simply by engaging in the activity and trying to do better.

In contrast, according to HOGARTH, wicked learning environments are those in which the correlation between outcomes and specific decisions or actions is ambiguous, deceptive, or non-existent. Two classic examples of wicked environments are the stock market and the business world. These domains

are complex environments with many variables and high volatility. HOGARTH uses an example in introducing the concept of wicked learning environments. An early 20th century physician in a New York hospital acquired a reputation for accurately diagnosing typhoid fever in its early stages. The physician believed that the appearance of the tongue was highly diagnostic. Hence, his clinical technique included palpating patients' tongues before making his pessimistic forecasts. Unfortunately, he was invariably correct since he was a more effective carrier, using only his hands, than Typhoid Mary. In most devilishly wicked environments, experience will reinforce the exact wrong lessons (the behavioural biases introduced in Section 2.2).

The current CPA certification program in Quebec is not preparing students to face wicked environments or a VUCA world – a military word for volatile, uncertain, complex and ambiguous environment. And because students in the CPA program don't know what their future will be after their undergrads study, universities should avoid positioning students in what economist BRYAN CAPLAN called "narrow vocational training" for jobs that few of them will ever have. "Three quarters of American college graduates go on to a career unrelated to their major university degree – a trend that included math and science majors – after having become competent only with the tools of a single discipline" (CAPLAN, 2018, pp. 233-35). One good tool is fairly not enough in a complex, interconnected and rapidly changing world.

JANNETTE WING, a computer science professor at Columbia University has pushed for broad computational thinking. She advocates that it becomes as fundamental as reading, even for those who will have nothing to do with computer science or programming. Computational thinking is using abstraction and decomposition when attacking large complex tasks. It is about choosing an appropriate representation for a problem (WING, 2006).

Critical thinking is important in a wicked environment. JIM FLYNN, Emeritus Professor of Political Studies at the University of Otago in Dunedin, New Zealand, is famous for his research on critical thinking. FLYNN's great disappointment is the degree to which society, and particularly high education, has responded to the broadening of the mind by pushing specialization, rather than focusing early training on conceptual, transferable knowledge. In 2008-2009, during his tenure as Visiting Fellow for the Sage Foundation in New York, he took the opportunity to pilot an index at a U.S. university to explore students' general critical thinking ability. His fears have been confirmed. Students from neuroscience to English majors were evaluated. The test gauged students' ability to

apply fundamental abstract concepts from economics, social and physical sciences, and logic to common, real-world scenarios. Only about a quarter of graduates are able to think critically outside of their discipline to any meaningful degree according to Professor Flynn's research. Economics graduates score highest on FLYNN's test; language students the worst. Overall, he concluded that students don't seem to have the key skills that would enable them to capitalize on their cognitive abilities (FLYNN, 2009). One of the arguments that may explain why economics students did well is because economics is a broad field by nature, and economic professors have been shown to apply the reasoning principles they have learned to problems outside their area (LARRICK et al., 1993).

The current Covid-19 pandemic is a good example where economists put their glasses on – economics vs. epidemiology – quantifying the trade-off. The rising economic toll of pandemic-induced shutdowns is fueling suspicions that government leaders are listening too much to epidemiologists and not enough to economists. That is understandable. Yet if economists were in charge, the policy response probably wouldn't be much different. While epidemiologists and economists study different problems, their approaches are similar, with a heavy reliance on statistics and an awareness of trade-offs. Epidemiologists know that extensive social distancing is costly, in terms of forgone wages, damage to physical and mental health, and harm to children's learning when schools are closed. Economists, similarly, know that even without lockdowns, a pandemic inflicts costs in terms of absenteeism, reduced consumption, lost lives, impaired health and uncertainty. Devising a policy response requires input from both disciplines – epidemiologists to estimate the benefits of social distancing in lives saved, and economists to calculate the costs in forgone jobs and income. So far, their work suggests the benefits exceed the costs, though that conclusion is surrounded by uncertainty. As politicians explore when and how to reopen their economies, economists and epidemiologists need to help them devise interventions that both maximize lives saved and minimize the economic cost.

As psychologist ROBIN HOGARTH pointed out about economists "…what strikes me about their discourse…. is how the terminology and reasoning process of economics work their way into almost all topics. Whether the topic is sports, politics, economic phenomena, or even academic curricula" (HOGARTH, 2001, p. 222). In FLYNN's words, the traits that earn good grades at the university do not include critical ability of any broad significance (FLYNN, 2012). The study he conducted convinced him that universities rush to develop students in a narrow specialty area, while failing to

sharpen the tools of thinking that can serve them in every area. They must be taught to think before being taught what to think about.

The more constrained and repetitive a challenge, the more likely it will be automated, while great rewards will accrue to those who can take conceptual knowledge from one problem or domain and apply it in an entirely new one. In a kind learning environment, research shows that it is possible to learn from experience. For wicked environments, it is different. KAHNEMAN (2011) studied human decision-making from the heuristics and biases model of human judgment. When KAHNEMAN probed the judgment of highly trained experts, he often found that experience had not helped at all. Even worse, experience frequently bred confidence but not skill. The wicked world demands conceptual reasoning skills that can connect new ideas and work across contexts.

In addition to being flexible, knowledge should be durable. Desirable difficulties contribute to knowledge durability. The term was coined by Dr. ROBERT BJORK over 20 years ago. It refers to conditions of learning that create challenges for learners – and even seem to slow down the rate of learning – while actually enhancing long-term retention of knowledge and skills. Leveraging on Bjork's work, NATE KORNELL, a cognitive psychologist at Williams College demonstrated that excessive hint-giving in a math classroom bolsters immediate performance but undermines progress in the long run (KORNELL & METCALFE, 2007). Desirable difficulties foster deep learning. A successful desirable difficulty is known as the "generation effect". Struggling to generate an answer on your own, even a wrong one, enhance subsequent learning. It requires the learner to intentionally sacrifice current performance for future benefit. Psychologist JANET METCALF (METCALF et al., 2012) coined the term "hypercorrection effect". The more confident a learner is of their wrong answer, the better the information sticks when they subsequently learn the right answer. Tolerating big mistakes can create the best learning opportunities.

Deep learning (learning that is flexible and durable) requires time because it has to be learned under various conditions, an approach researchers called "interleaving". Interleaving is a process where learners mix, or interleave, multiple subjects or topics while they study in order to improve their learning. Blocked practice, on the other hand, involves studying one topic very thoroughly before moving to another topic. Interleaving has been shown to be more effective than blocked practice for developing the skills of categorization and problem solving; interleaving also leads to better long-term retention and improved ability to transfer learned knowledge. This strategy forces the brain to

continually retrieve because each practice attempt is different from the last, so rote responses pulled from short-term memory won't work. Cognitive psychologists believe that interleaving improves the brain's ability to differentiate, or discriminate, between concepts and strengthens memory associations. Because interleaving involves retrieval practice, it is more difficult than blocked practice. It is important to remember that effortful studying feels worse but produces better long-term results. Interleaving tends to fool learners about their own progress. KORNELL AND BJORK's interleaving study reveals that 80 percent of students were sure they had learned better with blocked than mixed practice, whereas 80 percent performed in a manner that proved the opposite (KORNELL & BJORK, 2008). The feeling of learning, it turns out, is based on before-your-eyes progress, while deep learning is not. Deep learning is slow. The more complex the skills and knowledge, the slower the growth. Interleaving improves the ability to match the right strategy to a problem. Whether it is a fraud problem, tax problem, chemistry problem, the most successful problem solvers spend mental energy figuring out what type of problem they are facing before matching a strategy to it, rather than jumping in with memorized procedures.

When a knowledge structure is flexible, portable to new situations and can be applied by the person facing such situations, it is what NANCY DIXON, an expert in the field of organizational learning, called "far transfer" (DIXON, 2001). Far transfer occurs when the new situation is very different from that in which learning occurred. Factors that can affect transfer include (SOUSA, 2017):

- Context and degree of original learning: how well the learner acquired the knowledge;
- Similarity: commonalities between original learning and new learning, such as environment and other memory cues;
- Critical attributes: characteristics that make something unique; and
- Association: connections between multiple events, actions, bits of information, and so on; as well as the conditions and emotions connected to it by the learner.

Association is critical when you operate in a wicked environment. DEDRE GENTNER, an American cognitive and developmental psychologist and a leading researcher in the study of analogical reasoning (association), developed the structure-mapping theory of analogy and similarity (GENTNER, 1983), which has wide application. This involves the mapping of knowledge from one domain into another or from the base to the target for the purpose of guiding reasoning, to develop

conjectures and to generalize experiences into abstract schema. GENTNER also maintained that this theory of analogy can be used to model other subprocesses in analogical reasoning. Deep analogical thinking is the practice of recognizing conceptual similarities in multiple domains or scenarios that may seem to have little in common on the surface. It is an important ability to solve the dilemma that exists in a VUCA world and wicked environments. Analogical thinking takes the new and makes it familiar or takes the familiar and puts it in a new light and allows humans to reason through problems they have never seen in unfamiliar contexts.

Using analogy is not always well suited for wicked environments. As stated earlier, KAHNEMAN demonstrated that our experience-based instincts are set up for kind environments where problem and solution repeat. He reminded us in *Thinking, Fast and Slow* (KAHNEMAN, 2011) that humans have a tendency to use single analogy and this does not help battle the natural impulse to employ the inside view – stay with the familiar – and fall into the biases trap. To successfully use analogies, we must use an outside view, get out of our comfort zone and, as stated previously, not be afraid of making mistakes.

Lateral thinking is a term coined by EDWARD DE BONO in 1967 (DE BONO, 2010). Lateral thinking is a manner of solving problems using an indirect and creative approach via reasoning that is not immediately obvious. It involves reimagining information in new contexts, including the drawing together of seemingly disparate concepts or domains that can give old ideas new uses. Gunpei Yokoi, a game designer, was the first person who launched and led Nintendo's research and development (NRD) department. As the head of NRD, Yokoi had no desire to compete with electronics companies that were racing one another to invent some entirely new silver and dazzling technology. Instead, he articulated his R&D philosophy around lateral thinking and withered technology.

Lateral thinking with withered technology is the idea that new does not automatically equate to good, and that, by using technology that has matured and is market-tested, designers are able to focus on creative uses for that technology, which ultimately yield superior experiences for consumers. This philosophy is why the Game Boy used a low resolution, monochrome screen without a backlight, even though illuminated colour screens were available on other electronics of the day. By cutting these costly and energy-consuming features, NRD was able to vastly extend the battery life of the Game Boy, which served the greatest priority of playing the Game Boy. As the power-hungry Lynx and Game Gear would soon learn, your games are only good when people can play them (SHEFF, 1993).

Yokoi's design ethos continued to influence Nintendo long after his resignation in 1997 - the Nintendo Wii, for instance, prioritized the utilization of rapidly cheapening accelerometer technology ahead of the system's graphical capabilities. Motion control was paramount to the idea of the Wii, as was the console's overall affordability. By using withered technology for the Wii's processing components, Nintendo was able to produce a console that had a low barrier to entry, in terms of both price and functionality. Wii Sports didn't look that much better than a GameCube game, but that didn't matter to the 3 million people that bought a Wii during its first month in stores.

The audit profession can learn from the economist profession. Since auditors practice a profession in a complex environment (an environment that is not fully observable, stochastic, dynamic and continuous – Table 5) or what HOGARTH called a wicked environment, there is a need to better prepare the students in the CPA certification program to develop their creative thinking. As intelligent agents will contribute more and more to find risk fraud factors in structured data, auditors with better creative thinking skills that leverage the contribution of intelligent agents will provide added value by increasing their ability to detect fraud.

4.4.3. The Role of the Ecosystem and the Audit Profession of the Future

Based on the reports summarized in Table 3, the discussion about the impact of artificial intelligence on jobs concerns mainly the employment impacts. Will sufficient quantities of new jobs be created to replace disappearing ones? Which jobs are the most at risk? Are auditor's jobs going to disappear? Based on the proposed Task Formula in this dissertation, the discussion about the impact of artificial intelligence focuses on qualitative changes in working life. What types of changes will take place within tasks and modes of work for an auditor?

In both cases, policy measures and labour market institutions will influence not only the rate of computerization of a task but also the form it will take. Technology does nothing by itself, auditors do the work. Rather than what is technologically possible, the important question is what is desirable for the audit profession in Quebec. The audit profession ecosystem in Quebec will collectively impact the audit profession of tomorrow.

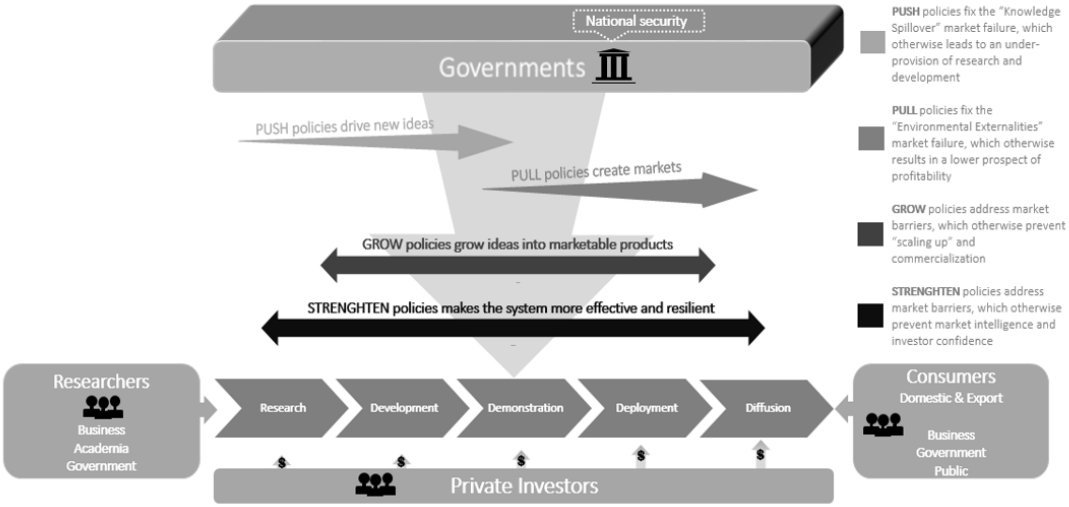

Figure 15. Audit Profession Ecosystem in Quebec

Source: MAKÓ & MALOUIN (2019)

As artificial intelligence penetrates the audit profession in Quebec, an effective audit labour market profession will be critical. Different skills will be required to perform the audit profession in Quebec. Based on the Task Formula, we can observe that as intelligent agents become better at predicting, the new skills discussed earlier should contribute to increasing the quality of an audit, and more specifically, the ability to identify fraud risk factors.

In this evolution of the audit profession in Quebec, policies with the aim of promoting and regulating intelligent agents and setting the stage for the audit profession of tomorrow will become essential. Since in Canada the audit profession is regulated at the provincial level, the government of Quebec (the State) will play a critical role to craft the future of the profession by working in collaboration with the audit profession ecosystem.

The government of Quebec can leverage push policies to do one of two things: establishing policies that incentivize private research initiatives, either through direct incentives (e.g., tax credits) or by helping audit firms capturing the economic returns from the research (e.g., through intellectual property rights), while other policies can focus on supplementing private research with public research through funding for government labs and universities. While these types of push policies focus on the early stages of innovation, they generate knowledge that carry through to later stages of the life cycle of a product or a service.

Pull policies can be used by the Quebec government to create markets. These types of policies are crucial for the audit profession. A number of countries, such as the U.K. and the U.S., are looking beyond financial statutory audits. For example, as carbon constraints are introduced in response to climate changes, and given the additional reporting requirements that are arising, auditors may be called upon to take the responsibility to audit these types of reports. These would be authorized auditors with the appropriate skill sets, which will be different than the skill sets required to audit financial statements. The contribution of artificial intelligence for this type of audit will also play a key role.

Grow policies are important to help promising innovations move from the R&D stage to the point where they are ready for large-scale market entry. This can be a long and difficult journey – one that is often called the "valley of death" for innovation. Developing intelligent agents capable of contributing to audit tasks will take time and required money. Financing this type of innovation can be a challenge for a company. Since its inception in the 1950s, AI has been falling short of this ideal. As demonstrated in Section 4.2, we are far from having intelligent agents that can do what auditors do the best – understanding, generalizing and contextualizing. This is why we call it machine learning and not machine understanding. The combination of higher risk profiles and longer scale-up timeframe chills private investment in many long tail innovations. This explains why most long tail innovations such as AI depends on a mix of public and private funds to reach the market. Over the past year, we have witnessed the importance of the Government of Quebec when it decided to invest in the biggest AI start-up in Canada: Element AI. In the fall of 2019, Element AI, a Montreal based company that builds intelligent agents for enterprises, raised CAD $200 million (USD $151 million) of funding from a host of existing and new investors, including the Government du Québec, Caisse de Dépôt et Placement du Québec, a provincial Crown Corporation, and other private investors. Among Element AI's array of cofounders is esteemed Canadian computer scientist Yoshua Bengio, who won a Turing award in 2018 for his work in deep learning.

Strengthen policies, those that support the ecosystem, magnify the impact of all other policies. Strengthening innovation must start with a bold and inclusive vision. Achieving that vision requires an equally bold and inclusive strategy – one that draws on the best existing knowledge and expertise, supported by new research in places. An effective strategy will not only articulate high-level objectives, priorities and actions, it must also dive deep and articulate potential pathways for different

sub-sectors, regions, and technology areas. It is important to identify the different challenges and opportunities that each stage of innovation faces – from R&D, to demonstration, deployment and, ultimately, to market diffusion – and how public policy can be tailored to help meet these specific needs and unleash private initiative. It's fair to expect that each stage's journey will be different.

The universities' performance guidance and incentive scheme in Quebec should be updated to create clear incentives for participating in cooperation and network research projects and commercialization of research for technologies that could improve the audit profession. According to the Ministry of Economic Affairs and employment of Finland, experience has shown that R&D funding should be increased gradually through a multiannual programme requiring commitment (MINISTRY, 2018).

As stated by the Ministry of Economic Affairs and Employment of Finland, a State should work not only for people's preparedness for change and learning but also the private sector – companies, professional firms, etc… – to renew themselves and to support the upskilling of their employees. Society's formal education system should also support continuous development of skills and lifelong learning.

In many countries such as the U.S. and the UK, private investors and stakeholders have been vocal about the failure of the audit profession to meet their expectations – what is known as the expectation gap. In April 2019, The Guardian in the U.K., in reference to Carillion and Patisserie Valerie, reported that "the auditors' failure to spot the fragility of those businesses resulted in the loss of jobs, savings, pensions, and tax revenues" (SIKKA, 2019). As demonstrated in Section 4.2, if the audit profession can leverage on intelligent agents to improve the prediction variable in the Task Formula, this will leave more time for the auditor on value added work – connecting the dots between different sources of audit evidence from the 160 sub-tasks in Appendix 2. This can be done by leveraging on an improved judgment resulting from creative thinking, as discussed previously. If intelligent agents are used appropriately by the auditors to complement human work, the quality of an audit should increase and narrow the expectation gap.

4.5. Discussion of the Findings

The study examined the potential contribution of intelligent agents on the identification of fraud risk factors. Three research questions were formulated and three hypotheses were tested.

4.5.1. Hypotheses

1. Intelligent agents are not a substitute for the audit profession in Quebec and cannot result in a massive employment loss.
2. Intelligent agents cannot assume creative cognitive tasks.
3. The regional audit ecosystem is playing a key role in collective learning and developing ethical and moral regulations for the future of the audit profession in Quebec.

All hypotheses were proven.

Leveraging on the CPA Quebec qualitative case study, nine learning agents were analyzed. These intelligent agents belong to two out of the five tribes of learning agents: symbolists (rule-based) and connectionists (machine learning algorithms).

Based on the Task Formula:

$$\text{Task} = f\ (\text{data} + \text{prediction} + \text{judgment} + \text{action})$$

This study demonstrates that these learning agents have a limited but valuable contribution to identify fraud risk factors: the prediction variable. If an auditor leverages properly on the outcome of the work of these learning agents to identify fraud risk factors, it gives them more insight and more time to leverage on their judgment. This can only contribute to increase the quality of the audit.

The intelligent agents analyzed in this dissertation cannot make a comprehensive assessment (generalization) like an auditor can do, they cannot contextualize and have no common-sense reasoning like an auditor has. As a result, intelligent agents are not a substitute of the audit profession in Quebec (as well as other countries) and cannot result in a massive employment loss.

To rely on these learning agents and ensure that an audit will comply with the Canadian Auditing Standards, the auditor must understand the limits of these learning agents and the type of biases they might be subject to.

As companies enter the digital world, the audit profession ecosystem in Quebec (and other countries as well) will play an important role to reshape the future of the profession.

4.5.2. Research Questions

1. What is the main ethical consideration CPA Quebec should analyze and understand as artificial intelligence will penetrate the audit profession in Quebec?
2. What are or could be the impacts of artificial intelligence on both the content of the curriculum to access the chartered professional accountant profession in Quebec and the reskilling requirements?
3. What could be the role of the CPA Quebec ecosystem (government, professional order, universities and firms) in learning in the age of artificial intelligence?

First Question: Professional competence and due care require auditors to comply with Canadian Auditing Standards, which require obtaining sufficient appropriate evidence as the basis for an audit opinion on the financial statements. The most significant ethical challenges created by machine learning for the audit profession in Quebec (and other countries as well) is explainability. Auditors face fundamental limits on their ability to trace inductive reasoning of complex intelligent agents and, as a result, rely blindly on their outcome which will not allow the auditors to comply with CAS. Explainable AI is important for the audit profession in Quebec (and other countries as well) since it will allow the auditors to comply with the Code of Professional Ethics. There is a large consensus on the need for machine learning to be interpretable/explainable.

Second Question: There are six technical competencies in the CPA curriculum in Quebec: financial reporting, strategy and governance, management accounting, audit and assurance, finance and taxation. The CPA enabling competencies provide the essential skills for ethical behaviour, leadership, teamwork, decision-making, problem-solving, and communication as a professional accountant. However, the CPA curriculum focuses on the technical competencies or specialization. In the undergraduate studies, there are no specific classes on the topic of creative thinking (professional judgment). Since the audit profession operates in a wicked environment the contribution of intelligent agents is limited – they cannot contextualize and generalize. Creative thinking allows humans to do it. Creative thinking can use abstraction, is portable from one situation to another, is flexible, is interleaving and more. Studies show that economists are better prepared in that regard and the audit profession may learn from them on how to improve the CPA curriculum in Quebec.

Third Question: To improve the quality of the audit work, the profession should leverage on artificial intelligence. The audit profession ecosystem can contribute to improve the quality of an audit at three levels: improving the ability of students in the CPA program to develop their creative thinking, developing the appropriate skill requirements associated with artificial intelligence for an auditor and developing intelligent agents that can contribute in the execution of specific audit tasks.

4.6. New Scientific Results

Based on the results and discussion, the new scientific results from this research are as follows:

1. The study reveals that the Task Formula provides a useful starting point to assess the contribution of intelligent agents to execute a task.
2. Another distinctive result observed from this study is the Task Complexity Framework. The framework can contribute to assess the potential contribution of intelligent agents for a specific audit task. The framework contributes to a more precise assessment than the methodologies used by the authors of the studies summarized in Table 3.
3. One of the novel results brought forth by this study is the five recipients' conceptual framework for the audit ecosystem in Quebec to reach explainability in AI. Auditors face fundamental limits on their ability to trace the inductive reasoning of complex intelligent agents. This is where explainable artificial intelligence can play a significant role because as long as high-performing models remain opaque, it will impact the ability of the audit profession in Quebec to leverage intelligent agents and comply with CAS.
4. Another distinctive result observed from this study is the need to better prepare the students in the CPA certification program to develop their creative thinking. As intelligent agents will contribute more and more to find risk fraud factors in structured data, auditors with better creative thinking that leverage on the contribution of intelligent agents will provide added value by increasing their ability to detect fraud.

5. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

The fundamental aim of this research was to assess the contribution of intelligent agents to the audit profession in Quebec. This research is exploratory and descriptive in nature and it is novel for the audit profession in Quebec since there is no research that I am aware of that encompasses the four topics analyzed in this dissertation: the audit profession in Quebec, artificial intelligence, knowledge and learning.

The dream of creating an intelligent machine – one that is as smart as or smarter than humans – is centuries old but became part of modern science with the rise of the digital computer. Since its earliest days however, artificial intelligence has been long on promise, short on delivery. General-purpose artificial intelligence with the flexibility of human intelligence is not a small achievement to reach. An important challenge starts with the meaning of intelligence. Voltaire once said "define your terms…. or we shall never understand one another." This study presented different dimensions of human intelligence. There is no one way the human brain works because the brain is not one thing. Instead, the brain has parts, and the different parts of the brain operate in different ways. Interestingly, intelligent agents don't have to work the same way as humans. There is no need for them to make the same cognitive errors that impair human thought, such as confirmation bias, or the many limitations of the human mind, such as the difficulty that human beings have in reciting the alphabet in reverse order in less than 10 seconds. Defining intelligence as a measure of skill-acquisition efficiency over a specific task, with respect to prior knowledge, experience, and local generalization ability is a good starting point to assess the contribution of intelligent agents on the audit profession in Quebec.

Predicting the scale of impacts of intelligent agents on jobs and tasks has become a cottage industry for economists and consulting firms. Depending on which model one uses, estimates range from terrifying to totally not a problem. While I respect the expertise of economists and scientists who pieced together the studies presented in this dissertation, I also respectfully disagree with their methodology and estimates. Based on the Task Formula and the Task Complexity Framework, the results indicate that intelligent agents can contribute to improve the quality of an audit but cannot replace the auditor. There are a number of reasons that can explain such an outcome. The key one is

trustworthiness. The narrow intelligent agents analyzed in this study work on what they are programmed for, but they cannot be trusted with anything that hasn't been precisely anticipated by their programmer. Intelligent agents need to be able to deal not only with specific situations for which there is an enormous amount of data, but also for tasks that are novel, and variations that have not been seen before. The phrase "barrier of meaning" perfectly captures the limitations of intelligent agents. Auditors, in some deep and essential way, understand the business situations they encounter, whereas no intelligent agents yet possess such understanding. While the state-of-the-art AI systems have nearly equaled, and in some cases surpassed, humans on certain narrowly defined tasks, the nine intelligent agents analyzed in this study lack a grasp of the rich meaning that auditors bring to bear in language, reading, perception and reasoning. This lack of understanding is revealed by the limited number of sub-tasks they can accomplish to identify fraud risk factors; their difficulties with abstracting and transferring what they have learned; and by their lack of commonsense knowledge.

5.2. Recommendations

Based on the findings of this research, an important recommendation is provided for further research. Technological innovation is changing the nature of many jobs, and the qualifications that employers seek in their workers and the audit profession in Quebec is not exempt from such situations. As intelligent agents penetrate the audit profession in Quebec, the knowledge required to practice the profession will have to evolve. In future research, it is strongly recommended that an Essential AI Audit Skills Framework be defined and developed for tomorrow's CPAs. Today, the capacity of Quebec's university ecosystem to respond to the skill evolution requirements for the audit profession is limited. The audit profession regulatory body is also struggling the understand and define the profile of tomorrow's CPAs.

5.3. Research Limitations and Future Research Directions

Limitations of this research must be recognized. First, this is an exploratory research. No such research has been conducted to assess the impact of artificial intelligence on the audit profession in Quebec. The novelty of this research is to assess the impact of intelligent agents on a specific audit task, but it carries a second limitation – only nine algorithms will be analyzed. Over time, many more algorithms can be analyzed. The third limitation is that I did not have access to the source code, neither the training or testing data, because of privacy regulation in Canada. Finally, ethics is an important

challenge in the field of artificial intelligence. There is much research on the topic, and it is subject to hundreds of PhD dissertations every year. Based on my research, I carefully selected a limited number of challenges I wanted to analyze in the field of ethics and artificial intelligence.

To prepare for tomorrow's business landscape, the profession must evolve and provide more value added. The expectation gap that exist between the profession in Quebec and the stakeholder's community must be reduced. As a result, the following research can provide valuable insights to CPA Quebec:

- Leveraging on the Task Formula, research can be undertaken to assess the readiness of audit firms to re-skill the audit workforce that have limited knowledge of intelligent agents.
- Further research can be undertaken to assess how CPA Quebec can improve the CPA Competency Map to better develop artificial intelligence technical knowledge to CPA candidates and CPA professionals.
- Assessing the contribution of natural language processing technology (NLP intelligent agents) to execute the audit task 6 (understanding the auditee's business, environment, risk, management and strategy) in Table 1. NLP intelligent agents may provide a concise and precise summary in that regard, helping auditors get a better understanding of a company.
- Further research on XAI can be undertaken and, more specifically, on post-hoc intelligent systems to assess if auditors can understand explanations provided by these systems.
- Other intelligent agents can be evaluated to assess their ability to contribute to other audit tasks in Table 1.

6. SUMMARY

This dissertation takes a step in understanding how intelligent agents can impact the audit profession in Quebec. This research should assist the audit profession ecosystem in Quebec (and other countries as well) to prepare the next generation of auditors and ensure that the CPA continuing education program for auditors in Quebec reflects the progressive impacts of intelligent agents on the audit profession.

Findings from this research can be characterized as follows: intelligent agents can contribute to some sub-tasks related to the identification of fraud risk factors. This should allow an auditor to reallocate his time to more complex sub-tasks related to the identification of fraud risk factors. The combination of human and machine should increase the quality of an audit. Intelligent agents suffer two main obstacles: they have no ability to generalize and contextualize as an auditor can do. AI biases and explainability could slow down the integration of intelligent agents by the audit profession in Quebec (and other countries). Explainability can generate an ethical challenge for the audit profession in Quebec (and other countries). As intelligent agents will contribute more to accomplish certain audit tasks, auditors will have more time to leverage on their professional judgment to address more complex audit issues. However, The CPA Quebec Competency Map focuses mainly on six technical competencies. The CPA Competency Map does not address the development of creative thinking (professional judgment) of the students in the CPA designation program. The audit profession ecosystem can contribute to improve the quality of an audit at three levels: improving the ability of students in the CPA program to develop their creative thinking, developing the appropriate skill requirements associated with artificial intelligence for an auditor and developing intelligent agents that can contribute in the execution of specific audit tasks.

7. APPENDICES

Appendix (1) References

1. ABBOTT, A. (1988): *The system of profession*. Chicago: The University of Chicago Press.
2. ACADEMY FOR LIFE (2020): Rome call for AI ethics, Retrieved 20 July 2020 from http://www.academyforlife.va/content/dam/pav/documenti%20pdf/2020/CALL%2028%20febbraio/AI%20Rome%20Call%20x%20firma_DEF_DEF_.pdf.
3. ABDUL, A., VERMMEULEN, J., WANG, D., LIM, B.Y. & KANKANHALLI, M. (2018): Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. *In Proceedings of the CHI Conference on Human Factors Computer Systems*, pp. 1-18.
4. ABDOLMOHAMMADI, M.J. (1999): A comprehensive taxonomy of audit task structure, professional rank and decision aids for behavioral research. *Behavioral Research in Accounting*, 11, pp. 51-92.
5. ACFE. (2018): Report to the nation. Retrieved on 4 February 2019 from: https://www.acfe.com/report-to-the-nations/2018/default.aspx.
6. ALASOINI, T. (2018): Work in the age of intelligent automation: Some remarks on the current automation debate. *Finish Institute of Occupational Health*, pp. 1-16.
7. AMERICAN ACCOUNTING ASSOCIATION, COMMITTEE ON BASIC AUDITING CONCEPTS (1973): A Statement of Basic Auditing Concepts. Retrieved on 7 January 2019 from: https://www.worldcat.org/title/statement-of-basic-auditing-concepts/oclc/701015.
8. ARROW, K.J. (1963): Uncertainty and the welfare economics of medical care. *American Economic Review*, 53 (3), pp. 941–973.
9. ARROW, K.J. (1972): Gifts and exchanges. *Philosophy and Public Affairs*, 1 (4), pp. 343-362.
10. ARROW, K.J. (1984): The economics of agency. *Technical Report No. 452*. A Report of the Center for Research on Organizational Efficiency, Stanford University.
11. ARROW, K.J. (1996): The economics of information: An exposition. *Empirica*, 23 (2), pp. 119-128.
12. ARTHUR, W.B. (1988): Competing technologies: An overview, pp. 590-607 in Dosi, G., Freeman, C., Nelson, R., Silverberg, G., & Soete, L. (eds), *Technical Change and Economic Theory.* London and New York: Columbia University Press and Pinter.
13. ATKINSON, D. (2018): Testimony of Robert D. Atkinson President of the information technology and innovation foundation before the Little Hoover Commission: Hearing on economic and labor force implications of artificial intelligence. Retrieved on 18 February 2019 from: https://lhc.ca.gov/sites/lhc.ca.gov/files/Reports/245/WrittenTestimony/AtkinsonJan2018.pdf.
14. ATWATER, L.E. (1988): The relative importance of situational and individual variables in predicting leader behavior. *Group and Organizations Studies*, 13 (3), pp. 290-310.
15. AUTOR, D.H., LEVY, F. & MURANNE, R.J. (2003): The skill content of recent technological change: an empirical exploration. Retrieved on 22 January 2019 from: http://economics.mit.edu/files/11574.
16. BAIN AND COMPANY. (2018): Labor 2030: the collision of demographics, automation and inequality. Retrieved on 24 January 2019 from: https://www.bain.com/insights/labor-2030-the-collision-of-demographics-automation-and-inequality/.

17. BAROCAS, S. & SELBST, A.D. (2016): Big data's disparate impact. *California Law Review*, 104, pp. 671-732.
18. BASU, S., KUMBIER, K., BROWN, J.B. & YU, B. (2018): Iterative random forests to discover predictive and stable high-order interactions. In *Proceedings of the National Academy of Sciences*, 115 (8), pp. 1943–1948. Retrieved on 5 February 2019 from: https://www.pnas.org/content/115/8/1943.
19. BAZERMAN, M.H. (1994): *Judgment in managerial decision making.* New York: Wiley.
20. BECKER, G.S. (1968): Crime and punishment: An economic approach. *Journal of Political Economy*, 76 (2), pp. 169-217.
21. BEER, M. (1968): Needs and need satisfaction among clerical workers in complex and routine job. *Personnel Psychology*, 37, pp. 209-222.
22. BLINCH, M. (2019): Justice, jobs and SNC-Lavalin. *The Globe and Mail*.
23. BOTSMAN, R. (2017): *Who can you trust: How technology brought us together and why it might drive us apart*. New York: PublicAffairs.
24. BRESNAHAN, T.F., BRYNJOLFSSON, E. & HITT, L.M. (2002): Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics*, 117 (1), pp. 339-376.
25. BROWNE, M.W. (1994): What is intelligence, and who has it? *The New York Time*. Retrieved on 7 February 2019 from: https://archive.nytimes.com/www.nytimes.com/books/00/06/04/specials/herrnstein-bell.html.
26. BRYDON, D. (2019): Assess, assure and inform: Improving audit quality and effectiveness. Retrieved on 7 May 2020 from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/852960/brydon-review-final-report.pdf.
27. BUSINESS ROUNDTABLE (2019): Business Roundtable redefines the purpose of a corporation to promote an economy that serves all Americans. Retrieved on 5 May 2020 from: https://www.businessroundtable.org/business-roundtable-redefines-the-purpose-of-a-corporation-to-promote-an-economy-that-serves-all-americans.
28. CALDERONE, K.L. (1990): The influence of gender on the frequency of pain and sedative medication administered to postoperative patients. *Sex Roles*, 23 (11/12), pp. 713-725.
29. CAMPBELL, D.J. (1984): The effect of goal-contingent payment on the performance of a complex task. *Personal Psychology*, 37 (1), pp 23-40.
30. CAMPBELL, D. J., & GINGRICH, K. F. (1986): The interactive effects of task complexity and participation on task performance: A field experiment. *Organizational Behavior and Human Decision Processes, 38* (2), pp. 162–180
31. CARUANA, R., LOU, Y., GEHRKE, J., KOCH, P., STURM, M. & ELHADAD, N. (2015): Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In: *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD pp. 1721–1730. Retrieved on 3 March 2019 from: https://dl.acm.org/doi/10.1145/2783258.2788613.
32. CATTELL, R. B. (1963): Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, pp. 1–22.
33. CHANDER, A., SRINIVASAN, R., CHELIAN, S., WANG, S. & UCHINO, K. (2018): Working with beliefs: AI transparency in the enterprise. In *Proceedings of Workshops of the ACM Conference on Intelligent User Interfaces*. Retrieved on 9 March 2019 from: http://ceur-ws.org/Vol-2068/exss14.pdf.

34. CHOLLET, F. (2019): On the measure of intelligence. Retrieved on 5 January 2020 from: https://arxiv.org/abs/1911.01547.
35. CNBC (2014): CNBC Transcript: Alibaba founder & Executive Chairman Jack Ma sits down with CNBC's David Faber today on "Squawk on the Street". Retrieved on 4 May 2019 from: https://www.cnbc.com/2014/11/11/cnbc-exclusive-cnbc-transcript-alibaba-founder-executive-chairman-jack-ma-sits-down-with-cnbcs-david-faber-today-on-squawk-on-the-street.html.
36. COASE, R.H. (1937): The nature of the firm. *Economica*, 4 (16), pp. 386-405.
37. COLEMAN, J.S. (1990): *Foundations of social theory*. Cambridge, MA: Harvard University Press.
38. COLEMAN, J.W. (1989): *The criminal elite*. New York: St. Martin's Press.
39. COOK, J. & WALL, T. (1980): New work attitude measures of trust, organizational commitment, and personal need nonfulfillment. *Journal of Occupational Psychology*, 53 (1), pp. 39-52.
40. COOPER, A. (2004): *The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity*. IND: Sams Publishing.
41. COWAN, N. (2010): The magical mystery four: How is working memory capacity limited, and why? *Current Direction in Psychological Science*, 19 (1), pp. 51–57.
42. CPA CANADA ONLINE HANDBOOK. (2020). Retrieve 4 January 2020 from: https://www.casso.ca/.
43. CRESSEY, D.R. (1973): *Other's people money*. Montclair: Patterson Smith.
44. CUMMING, L.L. (1983): Performance-evaluation systems in context of individual trust and commitment. In F.J. Landry, S. Zedrick, & J. Cleveland (eds.), *Performance measurement and theory*: pp. 89-93. Hillsdale, N.J: Earlbaum.
45. DATTA, A., SEN, S. & ZICK, Y. (2016): Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *Proceedings of IEEE Symposium on Security and Privacy*, pp. 598–617. Retrieved 9 June 2019 from: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7546525.
46. DEAL, T. & KENNEDY A. (1982): *Corporate cultures: Rite and rituals of corporate life*. Boston, MA: Addison-Wesley.
47. DE BONO, E. (2010): *Lateral thinking: Creativity step by step*. New York: HarperCollins.
48. DENNETT, D.C. (2017): *From bacteria to Bach and back: The evolution of minds*. New York: WW Norton & Company.
49. DEUTSCH, M. (1980): The effect of motivational orientation upon trust and suspicion. *Human Relations*, 13 (2), pp. 124-140.
50. DEY, A. DAVIDSON, R. & SMITH, A. (2015): Executives' "Off-the-Job" behaviors and financial reporting risk. *Journal of Financial Economics*, 117 (1), pp. 5-28.
51. DEY, A., DAVIDSON, R. & SMITH, A. (2019): Executives' legal records and the deterrent effect of corporate governance. *Contemporary Accounting Research*, pp.1-63.
52. DIXON, N.M. (2001): *Common knowledge: How companies thrive by sharing what they know*. MA: Harvard Business School Press.
53. DOMINGO, P. (2015): *The master algorithm: How the quest for the ultimate learning machine will remake our world*. New York: Basic Books.
54. DOSHI-VELEZ, F. & BEEN, K. (2017): Towards a rigorous science of interpretable machine learning. Retrieved on 15 June 2019 from: https://arxiv.org/abs/1702.08608.
55. DURKHEIM, E. (1951): *Suicide: A study in sociology*, translated by John A. Spaulding and George Simpson. New York: The Free Press.

56. DWORK, C., IMMORLICA, N., KALAI, A. T. & LEISERSON, M. (2017): Decoupled classifiers for fair and efficient machine learning. Retrieved on 25 June 2019 from: https://arxiv.org/pdf/1707.06613.pdf.
57. EDELMAN. (2019): *Trust Barometer 2019*. Retrieved on 15 June 2019 from: https://www.edelman.com/trust-barometer.
58. EGAN, M. (2018): GE pressured to fire auditor after 109 years. *CNN Business*. Retrieved on 24 May 2019 from: https://money.cnn.com/2018/04/24/investing/ge-annual-meeting-kpmg-accounting/index.html.
59. ENOS, J. L. (1962): Invention and innovation in the petroleum refining industry. Retrieved on 27 April 2019 from: https://core.ac.uk/download/pdf/6838223.pdf.
60. EUBANKS, V. (2018): *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.
61. EUROPEAN COMMISSION. (2010): Green paper: Audit policy: Lesson from the crisis. Retrieved on 14 May 2019 from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52010DC0561&from=EN.
62. EUROPEAN COMMISSION. (2018): Artificial Intelligence for Europe. Retrieved on 16 June 2019 from: https://ec.europa.eu/digital-single-market/en/artificial-intelligence.
63. EUROPEAN COMMISSION'S HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. (2018): Ethics guidelines for trustworthy AI. Retrieved on 25 June 2019 from: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.
64. FIFE, R., CHASE, S. & FINE, S. (2019): PMO pressed Justice Minister to abandon prosecution of SNC-Lavalin. *The Globe and Mail*.
65. FINANCIAL REPORTING COUNCIL. Quality audit review. (2020): Retrieved on 4 September 2020 from https://www.frc.org.uk/auditors/audit-quality-review/audit-firm-specific-reports?id=1195
66. FINE, S. (2019): A brief history of SNC-Lavalin. *The Globe and Mail*.
67. FISHBURN, P.C. (1982): *The foundations of expected utility*. New York: Springer.
68. FLYNN, J. (2009): Critical thinking index. Retrieved on 23 October 2019 from: https://www.otago.ac.nz/profiles/whatyouneedtoknow.html.
69. FORD, M. (2015): *The rise of the robot*. New York: Basic Books.
70. FREIDSON, E. (1988): *Professional power*. Chicago: The University of Chicago Press.
71. FREY, C.B. & OSBORNE, M.A. (2013): The future of employment: How susceptible are jobs to automation. Oxford Martin Program on Technology and Employment, Oxford University. Retrieved on 7 May 2019 from: https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf.
72. FULLERTON, D. & WOLFRAM, C. (2012): The design and implementation of U.S. climate policy. Chicago, Il: University of Chicago Press. Retrieved on 2 April 2019 from: https://www.nber.org/chapters/c12123.pdf.
73. FUKUYAMA, F. (1995): *Trust: The social virtues and the creation of prosperity*. New York: Free Press.
74. GABARRO, J. (1978): The development of trust, influence, and expectations. In A.G. Athos & J.J. Gabarro (Eds.), *International behavior: Communication and understanding in relationships*, pp. 290-303. Englewood Cliffs, NJ: Prentice Hall.
75. GENTNER, D. (1983): Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7 (2), pp. 155-170.

76. GIBBINS, M., WEBB, A. & SALTERIO, S. (2001): Evidence about auditor-client management negotiation concerning client's financial reporting. *Journal of Accounting*, 39 (2), pp. 535-563.
77. GIFFIN, K. (1967): The contribution of studies of source credibility to a theory of interpersonal trust in the communication department. *Psychology Bulletin*, 68 (2), pp. 104-120.
78. GLADWELL, M. (2011): *Outlier: The story of success*. New York: Back Bay Books.
79. GOOD, D. (1988): Individuals, interpersonal relations, and trust. In D.G. Gambetta (Ed.), *Trust*: pp. 131-185. New York: Basil Blackwell.
80. GRAAF, M.M. & MALLE, B.F. (2017): How people explain action (and automated intelligent systems should too). In *Proceedings of AAAI Fall Symposium on Artificial Intelligence for human-robot interaction*. Retrieved on 20 June 2019 from: https://aaai.org/ocs/index.php/FSS/FSS17/paper/view/16009.
81. GREEN, B. & CHEN, Y. (2019): Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In *Proceedings of the Conference on Fairness, Accountability and Transparency*, pp. 90–99. Retrieved on 23 January 2020 from: https://scholar.harvard.edu/files/19-fat.pdf.
82. GUIDOTTI, R., MONREALE, A., RUGGIERI, S., TURINI, F., GIANNOTTI, F. & PEDRESCHI, D. (2018): A survey of methods for explaining black box models. *ACM Computing Surveys*, 51 (5), pp. 1-93.
83. HAGE, J. (2000): Path dependencies of education systems and the division of labour within organization. In M. Maurice & A. Sorge, (eds.) *Embedding Organizations.* Amsterdam – Philadelphia: John Benjamins Publishing Company, pp. 311 -324.
84. HAMPTON, W.H., VENKATRAMAN, V. & OLSON, I.R. (2017): Dissociable front striatal white matter connectivity underlies reward and motor impulsivity. *NeuroImage*, 150 (4), pp. 336–343.
85. HANDEL, M.J. (2008): Measuring job content: Skills, technology, and management practices. Retrieved on 3 July 2019 from: https://irp.wisc.edu/publications/dps/pdfs/dp135708.pdf.
86. HARBERS, M., VAN DEN BOSCH, K., & MEYER, J.J. (2010): Design and evaluation of explainable BDI agents. *In Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, IEEE, pp. 125–132.
87. HARDIN, R. (2002): *Trust and trustworthiness*. New York: Russell Sage Foundation.
88. HEIDER, F. (1958): *The psychology of interpersonal relations*. Hillsdale, NJ: Lawrence Erlbaum.
89. HENDERSON, R. & TEMPLE-WEST, P. (2019): Group of US corporate leaders ditches shareholders-first mantra. *The Financial Times*. Retrieved on 4 January 2020 from: https://www.ft.com/content/e21a9fac-c1f5-11e9-a8e9-296ca66511c9.
90. HERLOCKER, J., KONSTAN, J. & RIEDEL, J. (2000): Explaining collaborative filtering recommendation. Retrieved on 27 June 2019 from: https://grouplens.org/site-content/uploads/explain-CSCW-20001.pdf.
91. HESSLOW, G. (1988): The problem of causal selection. In D. J. Hilton (Ed.), *Contemporary science and natural explanation: Commonsense conceptions of causality.* New York: New York University Press, pp. 11-32.
92. HIGHT-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (AI HLEG). (2019): Ethics guidelines for trustworthy artificial intelligence. Retrieved on 6 January 2020 from: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.
93. HILTON, D.J. & SLUGOSKI, B.R. (1986): Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, 93 (1), pp. 75-88.

94. HILTON, D.J. (1990): Conversational processes and causal explanations. *Psychological Bulletin*, 107 (1), pp. 65-81.
95. HILTON, D.J. (1996): Mental models and causal explanation: Judgements of probable cause and explanatory relevance. *Thinking & Reasoning*, 2 (4), pp. 273-308.
96. HILTON, D.J., MCCLURE, L.J. & SLUGOSKI, B.R. (2005): The Course of events: Counterfactuals, causal sequences and explanation. In D. R. Mandel, D. J. Hilton, P. Catellani (Eds.). *The psychology of counterfactual thinking*, UK: Routledge, University of Bristol. Chapter 9.
97. HOFFMAN, R.R. & KLEIN, G. (2016): Explaining explanation, part 1: Theoretical foundations. *Intelligent Systems*, 32 (3), pp. 68-73
98. HOGARTH, R. (2001): *Educating intuition*. Chicago: University of Chicago Press.
99. HONEGGER, M. (2018): Shedding light on black box machine learning algorithms: Development of an axiomatic framework to assess the quality of methods that explain individual predictions. Retrieved on 20 June 2020 from: https://arxiv.org/pdf/1808.05054.pdf.
100. HOOTON, E.A. (1939): *Crime and the man*. Cambridge, MA: Harvard University Press.
101. HOSMER, LT. (1995): Trust: The connecting link between organizational theory and philosophical ethics. *The Academy of Management Review*, 20 (2), pp. 379-403.
102. HOVLAND, C.I., JANIS, I.L., & KELLEY H.H. (1953): *Communication and persuasion.* New York: Yale University Press.
103. HUANG, G. B., RAMESH, M., BERG, T., & LEARNED-MILLER, E. (2007): Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *Technical Report* 07-49, University of Massachusetts, Amherst.
104. HUSTED B.W. (1989): Trust in business relations: Directions for empirical research. *Business & Profession Ethics Journal,* 8 (2), pp. 23-40.
105. HUYSMANS, J., DEJAEGER, K., MUES, C., VANTHIENEN, J. & BAESENS, B. (2011): An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. *Decision Support Systems*, 51 (1), pp. 141–154.
106. IBM. (2019): Machine learning and bias. Retrieved on 13 January 2020 from: https://developer.ibm.com/technologies/machine-learning/articles/machine-learning-and-bias/.
107. IEEE. (2019): Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems. *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*. Retrieved on 18 January 2020 from: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined.
108. IPN SIG AI. (2018): Dutch artificial intelligence manifesto. Retrieved on 19 December 2019 from: http://ii.tudelft.nl/bnvki/wp-content/uploads/2018/09/Dutch-AI-Manifesto.pdf
109. JAMES, G., WITTEN, D., HASTIE, T. & TIBSHIRANI, R. (2013): *An introduction to statistical learning.* New York: Springer.
110. JANVRIN, D., BIERSTAKER, J. & LOWE, D.J. (2008): An examination of audit information technology use and perceived importance. *Accounting Horizon*, 22 (1), pp. 1-21.
111. JOLLIFFE, I. (2011): Principal component analysis. *In International Encyclopedia of Statistical Science*. Springer: Berlin, Germany, pp. 1094–1096.
112. JONES, A.P., JAMES, L.R. & BRUNI, J.R. (1975): Perceived leadership behavior and employee confidence in the leader as moderated by job involvement. *Journal of Applied Psychology*, 60 (1), pp. 146-149.

113. JOSEPHSON, J.R. & JOSEPHSON, S.G. (1996): *Abductive inference: Computation, philosophy, technology*. Cambridge: Cambridge University Press.
114. KAHNEMAN, D. (2011): *Thinking fast and slow*. New York: Farrar, Strauss & Giroux.
115. KEE, H.K. & KNOX, R.E. (1970): Conceptual and methodological considerations in the study of trust and suspicion. *Journal of Conflict Resolution*, 14 (3), pp. 357-366
116. KIM, B., KHANNA, R. & KOYEJO, O.O. (2016): Examples are not enough, learn to criticize! Criticism for interpretability. In *Advances in Neural Information Processing Systems*; Cambridge, MA: MIT Press. pp. 2280–2288.
117. KIRSCH, A. (2017): Explain to whom? Putting the user in the center of explainable AI. In *Proceedings of the First International Workshop on Comprehensibility and Explanation in AI and ML*. Retrieved on 14 December 2019 from: https://hal.archives-ouvertes.fr/hal-01845135/document.
118. KOHLBERG, L. (1958): *The development of modes of moral thinking and choice in the years 10 to 16*. University of Chicago. Dissertation. Retrieved on 7 June 2019 from: https://www.worldcat.org/title/development-of-modes-of-moral-thinking-and-choice-in-the-years-10-to-16/oclc/4612732.
119. KOHLBERG, L. & MAYER, R. (1972): Development as the aim of education. *Harvard Educational Review*, 42 (4), pp. 449 - 496.
120. KOPLYAY, T., & MITCHELL, B. (2015): Evolution of the firm and its alliance structural forms in dynamic markets. In *Proceedings of the 2015 Industrial and Systems Engineering Research Conference*, Nashville, USA.
121. KOPLYAY, T. & MALOUIN, M. (2019): A framework for understanding illegal business activities, from competition to corruption - The case of SNC Lavalin. In *Proceedings of the American Society for Engineering Management 2019 International Annual Conference E. Schott, E-H. Ng, H. Keathley, and C. Krejci eds.*
122. KORNELL, N. & BJORK, R.A. (2008): Learning concepts and categories: Is spacing the enemy of induction? *Psychological Science*, 19 (6), pp. 585-92.
123. KUANG. C. (2017): Can AI be taught to explain itself? *The New York Times*. Retrieved 19 November 2019 from: https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html.
124. KUHN, M. & JOHNSON, K. (2013): *Applied predictive modeling*. New York: Springer.
125. KULESZA, T., STUMPF, S., BURNETT, M. YANG, S., KWAN, I., & WONG, W.K. (2013): Too much, too little, or just right? Ways explanations impact end users' mental models. In *Proceedings of Visual Languages and Human- Centric Computing,* IEEE Symposium, pp. 3-10.
126. LACAVE, C. & DIEZ, F.J. (2001): A review of explanation methods for Bayesian networks. Retrieved on 23 October 2019 from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.25.90
127. LANGER, E. (1983): *The psychology of control*. Beverly, CA: Sage Publications.
128. LANGLEY, P., MEADOWS, B., SRIDHARAN, M. & CHOI, D. (2017): Explainable agency for intelligent autonomous systems. *In Proceedings* of *AAAI Conference on Artificial Intelligence*, pp. 4762–4763.
129. LARRICK, R.P., NISBETT, R.E. & MORGAN, J.N. (1993): Who uses the cost-benefit rules of choices? *Organizational Behavior and Human Decision Processes*, 56 (3), pp. 331-347.
130. LATHAM, G. & YUKL, G. (1975): A review of research on the application of goal-setting in organizations. *Academy of Management Journal*, 18 (4), pp. 824-845.

131. LAWLER, E. (1992): *The ultimate advantage: Creating a high-involvement organization*. San Francisco: Jossey-Bass.
132. LEAKE. D.B. (1995): Abduction, experience, and goals: A model of everyday abductive explanation. *Journal of Experimental & Theoretical Artificial Intelligence*, 7 (4), pp. 407-428.
133. LECUN, Y., BENGIO, Y. & HINTON, G. (2015): Deep learning. *Nature*, 521, pp. 436-444.
134. LEGG, S. & HUTTER, M. (2007): A collection of definitions of intelligence. Retrieved on 4 April 2019 from: https://arxiv.org/abs/0706.3639.
135. LEWIS, J. & WEIGERT, A. (1985): Trust as a social reality. *Social Forces*, 63 (4), pp. 967-985.
136. LIBERMAN, J.K. (1981): *The litigious society*. New York: Basic Brooks.
137. LIBBY, R. (1995): The role of knowledge and memory in audit judgement. In *Judgement and decision making research in accounting and auditing*, edited by A.H. Ashton and R.H. Ashton, pp. 176-206. New York, NY: Cambridge Press University.
138. LIM, B.Y & DEY, A.K. (2009): Assessing demand for intelligibility in context-aware applications. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, ACM, pp. 195-204.
139. LIPTON, P. (1990): Contrastive explanation. *Royal Institute of Philosophy Supplement*, 27, pp. 247-266.
140. LIPTON, Z.C. (2016): The mythos of model interpretability. Retrieved on 23 April 2019 from: https://arxiv.org/abs/1606.03490.
141. LIPTON, Z.C. & STEINHARDT, J. (2018): Troubling trends in machine learning. Retrieved on 26 April 2019 from: https://arxiv.org/abs/1807.03341?utm_campaign=AI%20Weekly&utm_medium=email&utm_source=Revue%20newsletter.
142. LOCK, E., FREDERICK, E., LEE, C. & BOBKO, P. (1984): Effects of self-efficiency, and task performance. *Journal of Applied Psychology*, 69 (2), pp. 241-251.
143. LOMBROSO-FERRERO, G. (1972): *Criminal man according to the classification of Cesare Lombroso*. Glen Ridge, NJ: Patterson-Smith.
144. LOMBROZO, T. (2009): Explanation and categorization: How "why" informs the "what". *Cognition*, 110, pp. 248-253.
145. LOMBROZO, T. (2010): Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, 61 (4), pp. 303-332.
146. LOMBROZO, T. (2012): Explanation and abductive inference, *Oxford Handbook of Thinking and Reasoning*, pp. 260-276. Retrieved on 12 May 2019 from: https://books.google.ca/books?hl=en&lr=&id=VKnFAgAAQBAJ&oi=fnd&pg=PA260&dq=Lombrozo,+T.+2012.+Explanation+and+abductive+inference,+Oxford+Handbook+of+Thinking+and+Reasoning,+pp.+260-276&ots=Ogjc6kWtla&sig=GbCLCCNdO2bptO7X-n_iYt6vpBU#v=onepage&q&f=false.
147. LOU, Y., CARUANA, R. & GEHRKE, J. (2012): Intelligible models for classification and regression. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 150–158. Retrieved on 23 June 2019 from: https://doi.org/10.1145/2339530.2339556.
148. LUHMANN, N. (1988): Familiarity, confidence, trust: Problems and alternatives. In D.G. Gambetta (ed.), *Trust*: 90-107. New York: Blackwell.
149. MAANEN, J.V. (1979): Reclaiming qualitative research methods for organizational research: A preface. *Administrative Science Quarterly*, 24 (4), pp. 520-526.

150. MACCRIMMON, K. (1976): An overview of multiple objective decision-making. In J. Cochrane & M. Zelery (eds.), *Multiple criteria decision-making* Columbia, SC: University of South Carolina Press.
151. MAISTER, D.H. (1993): *Managing the professional service firm*. New York: Simon and Schuster.
152. MAKÓ, C.S. & MALOUIN M. (2019): The Blueprint for a dazzling future - a proper governance framework for the state investment in science: the case of Finland (Nokia) and Canada (Huwaei); Terv egy ragyogó jövőért – a tudományba eszközölt állami befektetések kormányzási kerete: Finnország (Nokia) és Kanada (Huawei) esete In: László, Gyula; Németh, Julianna; Sipos, Norbert (szerk.) Vezető és menedzser : Emlékkötet Farkas Ferenc születésének 70. évfordulójára Pécs, Magyarország : Pécsi Tudományegyetem Közgazdaságtudományi Kar Vezetés- és Szervezéstudományi Intézet, pp. 208-218.
153. MALLE, B.F. & KNOBE, J.M. (1997): The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33 (2), pp. 101-121.
154. MALLE, B.F., KNOBE, J.M. & NELSON, S.E. (2007): Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, 93 (4), pp. 491-514.
155. MARCH, J. & SIMON, H. (1958): *Organizations*. New York: Wiley.
156. MARCUS, G. (2019): Deep understanding: The next challenge for AI. *In Proceedings of NeurIPS*. Retrieved on 8 January 2020 from: https://slideslive.com/38922748/deep-understanding-the-next-challenge-for-ai.
157. MARCUS, G. & DAVIS, E. (2019): *Rebooting AI: Building artificial intelligence we can trust*. New York: Pantheon Books.
158. MARCUS, G. (2020): The next decade in AI. Retrieved on 23 July 2020 from: https://arxiv.org/ftp/arxiv/papers/2002/2002.06177.pdf.
159. MARKOPOLOS, H. (2010): *No one would listen*. New York: John Wiley & Sons.
160. MAYER, R.C., DAVIS, J.H. & SCHOORMAN, D.F. (1995): An integrative model of organizational trust. *The Academy of Management Review*, 20 (3), pp. 709-734.
161. MCCLURE, J. (2002): Goal-based explanations of actions and outcomes. *European Review of Social Psychology*, 2 (1), pp. 201-235.
162. MCGREW, K. S. (2005): The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*, p. 136–181. The Guilford Press.
163. MCKINSEY GLOBAL INSTITUTE. (2017): What the future of work will mean for jobs, skills, and wages. Retrieved on 3 March 2019 from: https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages.
164. MCLEOD, C. (2015): Trust. *The Stanford Encyclopedia of Philosophy*. Retrieved on 2 april 2019 from: https://plato.stanford.edu/archives/fall2015/entries/trust/.
165. MENDES, E. (2019): Is prosecutorial independence at risk. *The Globe and Mail*.
166. MERIAM, S.B. (2009): *Qualitative research: A guide to design and implementation*. San Francisco: Jossey-Bass.
167. METCALFE, J. & KORNELL, N. (2007): Principles of cognitive science in education. *Psychologic Bulletin and Review*, 14 (2), pp. 225-229.
168. METCALFE, J., BUTTERFIELD, B., HABECK, C. & STERN, Y. (2012): Neural correlates of people's hyperconnection of their false beliefs. *Journal of Cognitive Neuroscience*, 24 (7), pp. 1571-1583.

169. MILGROM, P. & ROBERTS, J. (1992): *Economics, organization and management*. Englewood Cliffs, NJ: Prentice Hall.
170. MILLER, B. & ATKINSON, D. (2017): Unfortunately, technology will not eliminate many jobs. Innovation Files. Retrieved on 24 March 2019 from: https://itif.org/publications/2017/08/07/unfortunately-technology-will-not-eliminate-many-jobs.
171. MISCHEL, W. & EBBESEN, E.B. & ZEISS, A.R. (1972): Cognitive and attentional mechanisms in delay and gratification. *Journal of Personality and Social Psychology*, 21 (2), pp. 204-218.
172. MONTAVON, G., LAPUSCHIN S., BINDER, A. SAMEK, W. & MÜLLER, K.R. (2017): Explaining nonlinear classification decisions with deep Taylor decomposition. Retrieved 8 July 2019 from: https://arxiv.org/abs/1512.02479.
173. MONTAVON, G., SAMEK, W. & MÜLLER, K.R. (2018): Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, pp. 1-15.
174. NELSON, R. & WINTER, S. (1982): *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.
175. NEYSHABUR, B., BHOJANAPALLI, S., MCALLESTER, D. & SREBRO, N. (2017): Exploring generalization in deep learning. Retrieved on 23 July 2019 from: https://proceedings.neurips.cc/paper/2017/file/10ce03a1ed01077e3e289f3e53c72813-Paper.pdf.
176. NICKERSON, R.S. (1998): Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2 (2), pp. 175-220.
177. NILSSON, N. J. (2010): *The quest for artificial intelligence*. New York: Cambridge University Press.
178. OECD. (2016): The risk of automation for jobs in OECD countries: A comparative analysis. Retrieved on 23 March 2019 from https://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5jlz9h56dvq7-en
179. O'NEIL, O. (2013): What we don't understand about trust. *TED Talk*. Retrieved on 10 July 2019 from: https://www.ted.com/talks/onora_o_neill_what_we_don_t_understand_about_trust?language=en.
180. PALLIER, G., WILKINSON, R., DANTHIIR, V., KLEITMAN, S., KNEZEVIC, G., STANKOV, L. & ROBERTS, R.D. (2002): The Role of individual differences in the accuracy of confidence judgments. *The Journal of General Psychology*, 129 (3), pp. 257–299.
181. PATRICK, M. & WIRTZ, M. (2019): For love and money in Mozambique: How a Credit Suisse baker helped fuel an alleged $2billion debt fraud. *The Wall Street Journal*. Retrieved on 3 January 2020 from: https://www.wsj.com/articles/for-love-and-money-in-mozambique-how-a-credit-suisse-banker-helped-fuel-an-alleged-2-billion-debt-fraud-11575227078?mod=searchresults&page=1&pos=1.
182. PEARL, J. (2009): *Causality*. New York: Cambridge University Press.
183. PEARL. J. & MACKENZIE, D. (2018): *The book of why: The new science of cause and effect*. New York: Basic Books.
184. PEREZ, C. (2010): Technological revolutions and techno-economic paradigms. *Cambridge Journal of Economics*, 34 (1), pp. 185-202.
185. POOLE D.L. & MACKWORTH, A.K. (2018): *Artificial intelligence: Foundations of computational agent*. New York: Cambridge university Press.

186. PORTUGUESE NATIONAL INITIATIVE ON DIGITAL SKILLS. (2019): AI Portugal 2030. Retrieved on 2 February 2020 from: https://www.portugal.gov.pt/download-ficheiros/ficheiro.aspx?v=5476a28b-cb00-4b20-b233-ad888206e1b3.
187. PRESIDENT DONALD J. TRUMP. (2020): Accelerating America's leadership in artificial intelligence. Retrieved on 5 August 2020 from: https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/.
188. PUTNAM, R.D. (2001): *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.
189. PwC. (2017): Will robot still our jobs? The potential impact of automation on the UK and other major economies. Retrieved on 7 May 2019 from: https://www.pwc.com/hu/hu/kiadvanyok/assets/pdf/impact_of_automation_on_jobs.pdf.
190. PwC. (2018): CEO success study. *Strategy&*. Retrieved on 6 July 2019 from: https://www.strategyand.pwc.com/gx/en/insights/ceo-success.html.
191. QUAID, J. & TAMAN, E. (2019): The SNC-Lavalin revelations show we are not a country bound by the rule of law. *The Globe and Mail*.
192. RAHMAN, K.S. & THELEN, K. (2019): The rise of the platform business model and the transformation of twenty-first-century capitalism. *Politics & Society*, 47 (2), pp. 177-204.
193. RAIN, A., LAUFER, W.S., YANG, Y., NARR, K.L., THOMPSON, P. & TOGA, A.W. (2012): Increased executive functioning, attention, and cortical thickness in white-collar criminals. *Human Brain Mapping*. Retrieved on 5 May 2019 from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3262063/.
194. RAINE, A. (2013): *The anatomy of violence*. New York: Pantheon Books.
195. RANI, P., LIU, C., SARKAR, N. & VANMAN, E. (2006): An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Analysis and Applications*, 9 (1), pp. 58–69.
196. RAPOPORT, M. (2018): Big four accounting firms' revenue rises 10.4%, strongest growth in years. *The Wall Street Journal*. Retrieved on 23 June 2019 from: https://www.wsj.com/articles/big-four-accounting-firms-revenue-rises-10-4-strongest-growth-in-years-11544713625.
197. RIBEIRO, M.T., SINGH, S. & GUESTRIN, C. (2016): Why should I trust you?: Explaining the predictions of any classifier. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 1135–1144. Retrieved on 9 July 2019 from: https://dl.acm.org/doi/10.1145/2939672.2939778.
198. RIKER, W.H. (1974): The nature of trust. In J.T. Tedeschi (ed.), *Perspectives on social power,* pp. 63-81. Chicago: Aldine.
199. REISS, A.J. & BIDERMAN, A. (1980): *Data source on white-collar law-breaking*. Washington DC: National Institute of Justice, U.S. Department of Justice. Retrieved on 20 June 2019 from: https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=72651.
200. ROTHMAN, K.J., GREENLAND, S. & LASH, T.L. (2008): *Epidemiology* (3rd edition). Philadelphia, PA: Lippincott, Williams & Wilkins.
201. ROYAL SOCIETY. (2017): Machine learning: The power and promise of computers that learn by example. Retrieved on 22 July 2019 from: https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf.
202. RUPPERT, D. (1987): *Robust statistics: The approach based on influence functions*, New York: Taylor & Francis.

203. RUSSELL, S.J. & NORVIG, P. (2015): *Artificial intelligence: A modern approach*. New York: Pearson.
204. SALZBERG, S. L. (1997): On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery*, 1 (3), pp. 317–328.
205. SAMUELSON, W. & ZECKHAUSER, R. (1988): Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1 (7), pp. 7-59.
206. SCHEIN, E. (2004): *Organisational culture and leadership*. San Francisco: Jossey-Bass.
207. SCHLEGEL, K. & WEISBURD, D. (1992): *White-collar crime reconsidered*. Boston: Northeastern University Press.
208. SCHRODER, H., DRIVER, M. & STREUFERT, S. (1967): *Human information processing*. New York: Rinehart and Winston.
209. SCIULLI, D. (2009): *Professions in civil society and the state*. Boston: Brill Academic Publishers.
210. SCOTT, D. (1980): The causal relationship between trust and the assessed value management by objectives. *Journal of Management*, Vol. 6, No. 2, pp. 157-175.
211. SELBST, A. D., BOYD, D., FRIEDLER, S. A., VENKATASUBRAMANIAN, S. & VERTESI, J. (2019): Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 59– 68. ACM. Retrieved on 6 January 2020 from: https://dl.acm.org/doi/10.1145/3287560.3287598.
212. SHAW, M. (1976): *Group dynamics: The psychology of small group behavior*. New York: McGraw-Hill.
213. SHEFF, D. (1993): *Game over: How Nintendo zapped an American industry, captured your dollars, and enslaved your children*. New York: Random House.
214. SHORTLIFFE, E.H. & BUCHANAN, B.G. (1975): A model of inexact reasoning in medicine. *Mathematical Bioscience*, 23 (3), pp. 351-379.
215. SIKKA, P. (2019): The Big 4 auditors are failing and the watch dog's report won't change that. *The Guardian*.
216. SIMON, H. A. (1957): *Models of man: Social and rational*. New York: John Wiley and Sons.
217. SIMON, H.A. (1991): *Models of my life*. New York: Basic Books.
218. SIMON, H.A. (1996): *The Science of the artificial* (3rd ed.). Cambridge, MA: MIT Press.
219. SITKIN, S.S. & ROTH, N.L. (1993): Exploring the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization Science*, 4 (3), pp. 367-392.
220. SMITH, A. 1790. *The theory of moral sentiment* (sixth edition). Retrieved on 17 June 2019 from: https://www.ibiblio.org/ml/libri/s/SmithA_MoralSentiments_p.pdf.
221. SMITH, H. (2002): Interview: Joseph Berardino. *Frontline*. Retrieved on 10 April 2019 from: https://www.pbs.org/wgbh/pages/frontline/shows/regulation/interviews/berardino.html.
222. SNC-LAVALIN ABOUT US. Retrieved on 3 February 2019 from https://www.snclavalin.com/en/about/leadership-and-governance/board-of-directors#kevin-g-lynch.
223. SOUSA, D. A. (2017): *How the brain learns*. Thousand Oaks, California: Sage Publication.
224. SPELK, S, & KINZLER, K.D. (2006): Core knowledge. *Development Science*, 10 (1), pp. 89-96.
225. STACEY, K. (2019): PwC pays $7.9m to settle improper conduct claims. *Financial Times*. Retrieved 16 January 2020 from: https://www.ft.com/content/b56d4f0c-de21-11e9-b112-9624ec9edc59#myft:my-news:page.

226. STEINMANN, D. (1976): The effect of cognitive feedback and task complexity in multiple-cue probability learning. *Organizational Behavior and Human Performance*, 15 (2), pp. 168-179.
227. STOUTHAMER-LOEBER, M. (1986): Lying as a problem behavior in children: A review. *Clinical Psychology Review* 6 (4), pp.267-289.
228. STREECK, W. (2017): *How capitalist end?* London: Verso.
229. SURESH, H. & GUTTAG, J.V. (2019): A framework to understanding unintended consequences of machine leaning. Retrieved on 20 January 2020 from: https://arxiv.org/abs/1901.10002.
230. SUTHERLAND, E.H. (1934): *Principles of Criminology*. Chicago: J.B. Lippincott.
231. SUTHERLAND, E.H. (1940): White-collar criminality. *American Sociological Review*, 5 (1), pp. 1-12.
232. SUTHERLAND, E.H. (1945): Is "white collar crime" crime? *American Sociological Review*, 10 (2), pp. 132-139.
233. SWARTOUT, W.R. (1983): Xplain: A system for creating and explaining expert consulting programs. *Artificial Intelligence*, 21 (3), pp. 285-325.
234. SZEGEDY, C., ZAREMBA, W., SUTSKEVER, I., BRUNA, J., ERHAN, D., GOODFELLOW, I. & FERGUS, R. (2013): Intriguing properties of neural networks. Retrieved on 10 September 2019: https://arxiv.org/abs/1312.6199.
235. TARDE, G. (1962): *The laws of imitation* (2nd ed.). Gloucester, MA: Peter Smith.
236. TAYLOR, R.G. (1989): The role of trust in labor-management relations. *Organization Development Journal*, 7, pp. 85-89.
237. TEGMARK, M. (2017): *Life 3.0.: Being human in the age of artificial intelligence*. New York: Alfred A. Knopf.
238. TERBORG, J. & MILLER, H. (1978): Motivation, behvior and performance: A closer examination of goal setting and monetary incentives. *Journal of Applied Psychology*, 63 (1), pp. 29-39.
239. THAGARD, P. (1989): Explanatory coherence. *Behavioral and Brain Sciences*, 12 (3), pp. 435-467.
240. THE DAILY BEAST. (2009): Madoff's lawyer issues final plea. *The Daily Beast*. Retrieved on 10 June 2019 from: https://www.thedailybeast.com/madoffs-lawyer-issues-final-plea?ref=scroll.
241. THINK 20 (2018): Future of work and education for digital age. Retrieved on 3 July 2019 from: https://www.g20-insights.org/wp-content/uploads/2018/07/TF1-1-11-Policy-Briefs_T20ARG_Towards-a-G20-Framework-For-Artificial-Intelligence-in-the-Workplace.pdf.
242. TIBSHIRANI, R. (1996): Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58 (1), pp. 267-288.
243. TOPOL, E. (2019): *Deep medicine: How artificial intelligence can make healthcare human again*. New York: Basic Books.
244. TYLER, T. R. (1990): *Why people obey law*. New Haven: Yale University.
245. UNIVERSITY OF MONTREAL. (2018): Montreal declaration for a responsible development of artificial intelligence. Retrieved on 7 June 2019 from: https://www.montrealdeclaration-responsibleai.com/the-declaration.
246. U.S. DEPARTMENT OF JUSTICE, BUREAU OF JUSTICE STATISTICS. (1981): *Dictionary of Criminal Justice Data Terminology*. Washington, DC: National Institute of

Justice, U.S. Department of Justice. Retrieved on 20 May 2019 from: https://www.ncjrs.gov/pdffiles1/Digitization/76939NCJRS.pdf.
247. UTTERBACK, J. M. & ABERNATHY, W. J. (1975): A dynamic model of process and product innovation. *Omega*, 3 (6), pp. 639–56.
248. VANDERWEELE, T. J. & HERNAN, M. A. (2012): Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs. *American Journal of Epidemiology*, 175 (12), pp. 1303– 1310.
249. VANIAN, J. (2020): Why artificial intelligence is so important in the coronavirus era. *Fortune*. Retrieved on 7 May 2020 from: http://links.info.fortune.com/e/evib?_t=5c2d888702774d17aa3d0350287b6d73&_m=b0cf11651a9e49a38b7acbd5c525eb58&_e=8Ns5q7hip9Xy7yNBS-OOr2eueIeaUo-uZ3iVxP24KkBTbK45GgGX5WoHJ1nO5bhIXVA1wMAjOkCT9V_W-VwWPWWJnlvzxRbqyFq7Hk73UH4yUkliktxYD3UKpRrX0zry.
250. VAN LENT, M. FISHER, W. & MANCUSO, M. (2004): An explainable artificial intelligence system for small-unit tactical behavior. In *Proceedings of the 16th Conference on Innovative Applications of Artificial Intelligence*, pp. 900–907.
251. VAN PRAET, N. & KILDAZE, T. (2019): SNC eye breakup in bid to halt falling market value. *The Globe and Mail*.
252. VAPNIK, V. (1999): An overview of statistical learning theory. *IEEE Transactions on Neural Network*, 10 (5), pp. 988-999.
253. VILLANI, C. (2018): AI for humanity: French National Strategy for artificial intelligence. Retrieved on 6 August 2019 from: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.
254. WALKER, N. & CATRAMBONE, R. (1992): Aggregation bias and the use of regression in evaluating models of human performance. Georgia Institute of Technology. Retrieved on 10 September 2019 from: https://smartech.gatech.edu/bitstream/handle/1853/3662/92-07.pdf.
255. WANG, H.X., FRATIGLIONI, L., FRISONI, G.B., VIITANEN, M. & WINBLAD, B. (1999): Smoking and the occurrence of Alzheimer's disease: Cross-sectional and longitudinal data in a population-based study. *American Journal of Epidemiology*, 149 (7), pp. 640–644.
256. WATSON, D. (2019): The rhetoric and reality of anthropomorphism in artificial intelligence. *Minds and Machine*, 29, pp. 417-440.
257. WEBB, M. (2019): The Impact of artificial intelligence on the labor market. Retrieved on 21 January 2020 from: https://ssrn.com/abstract=3482150.
258. WEISBURD, D., WARING, E. & CHAYET, E.J. (2001): *White-collar crime and criminal careers*. New York: Cambridge University Press.
259. WHEELWRIGHT, P. (1959): *A critical introduction to ethics*. Indianapolis: Odyssey Press.
260. WILLIAMS, J.J., LOMBROZO, T. & REHDER, B. (2013): The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology*, 142 (4), pp. 1006-1014.
261. WING, J.M. (2006): Computational thinking. *Communication of the ACM*, 49 (3), pp. 33-35. Retrieved on 12 November 2019 from: https://www.cs.cmu.edu/~15110-s13/Wing06-ct.pdf.
262. WOLF, J. (1912): *Die Volkswirtschaft der Gegenwart und Zukunft*, Leipzig: A. Deichert.
263. YANG, J. (2020): Behind the fall of China's Luckin Coffee: A network of fake buyers and fictious employees. *The Wall Street Journal*.
264. YIN, R.K. (2009): *Case study research: Design and methods*. Los Angeles: Sage Publications.
265. YORK, G., LEBLANC, D. & MCCLEARN, M. (2019): SNC received billions in support from federal export agency. *The Globe and Mail*.

266. YU, B. (2013): Stability. *International Statistical Institute/Bernoulli Society Bernoulli*, 19 (4), pp. 1484-1500.

Appendix (2) Business and Fraud Risk Factors

Governance
1. No emphasis is placed on need for integrity and ethical values 2. Directors have no financial expertise or have limited experience 3. No audit committee exists 4. Directors not independent of management 5. No strategic business plan or financial budget 6. Little attention paid by board (or audit committee) to financial reporting and internal control 7. Financial reports not provided on a timely basis 8. No comparison of actual results to budget 9. Inadequate review process for major decisions, management expense claims or large and unusual transactions with suppliers 10. Infrequent board meetings 11. Board dominated by a single person or small group 12. Inadequate policies/controls over major decisions and expenditures 13. Management roles and responsibilities not clear (no senior management job descriptions) 14. High turnover in board, management or accounting personnel 15. Allegations of fraud or non-compliance against entity, or management is not properly investigated 16. No process exists for staff and others to report suspected improprieties
Management attitudes
17. Significant deficiencies in internal control are ignored or uncorrected 18. Inadequate supervision of financial staff 19. Ineffective or improperly qualified accounting, information technology or internal auditing staff 20. History of management accepting high risks 21. Failure to monitor operation of significant controls 22. History of claims against entity or management alleging fraud 23. Unduly aggressive financial targets and expectations for operating personnel 24. High turnover of management personnel 25. Lack of mandatory vacations for personnel performing key control functions
Complex operating structure and or unusual transactions
26. Overly complex managerial lines of authority 27. Numerous or unusual legal entities 28. Overly complex organizational structure with no apparent reason 29. Significant volume of revenue not in ordinary course of business 30. Contractual arrangements exist without apparent business purpose 31. Unusual transactions at, or near, period end 32. Significant cash transactions 33. Significant audit adjustments identified 34. Correction of major errors required on a regular basis 35. Aggressive timing of revenue recognition 36. Unusual transactions, balances or accounting entries 37. Significant bank accounts or subsidiary/branch operations in tax-haven jurisdictions with no clear business justification 38. Significant related-party transactions exist. Highly complex and significant transactions, particularly toward period end

Regulatory environment
39. Legislation or regulations that will have a significant and possibly negative impact on operations have not been adequately addressed 40. The entity is prone to be involved in lawsuits or controversies 41. History of regulatory violations involving management 42. Allegations of possible fraud or compliance violations made by third parties or entity staff 43. Disregard is displayed toward regulatory or legislative authorities 44. Lack of documentation to explain impact of actual or pending events 45. Overly optimistic assessments made of re-assessments, investigations and other regulatory reviews
Accounting policies
46. Proposed changes in accounting principles are not addressed 47. Entity initiates many changes in accounting policies 48. Policies chosen are likely or known to be controversial or seem to manipulate earnings in a certain way 49. Use of unusually aggressive accounting practices that appear to maintain or increase the entity's stock price or earnings trends 50. Accounting changes ensure that lucrative personnel bonus/stock option plans are paid 51. Overly optimistic/pessimistic estimates made
Deteriorating industry conditions
52. High degree of competition or market saturation 53. Losses and/or declining profit margins 54. Declining industry with increasing business failures 55. Significant declines in customer demand 56. High vulnerability to rapidly changing technology or product obsolescence 57. Poor sales outlook 58. Intense or new competition 59. Commitments to bankers, analysts, creditors or other third parties on achieving aggressive or clearly unrealistic forecasts 60. Excessive interest in maintaining or increasing entity's stock price or earnings trend 61. Meeting bonus thresholds difficult to achieve 62. Overly optimistic projections made of earnings 63. Undue emphasis on planning reported earnings 64. Overly optimistic estimates with respect to asset impairment and/or extent of liabilities 65. Significant debt guarantees provided by management
Deteriorating financial conditions
66. Marginal ability to meet debt repayment requirements, or debt covenants that are difficult to maintain 67. Threat of imminent bankruptcy, foreclosure or hostile takeover 68. Pressures to obtain new capital 69. Negative cash flows 70. Contingent liabilities 71. New accounting or regulatory requirements that would impair performance 72. High vulnerability to changes in interest rates 73. Unusually high dependence on debt

74. Commitments to bankers, analysts, creditors or other third parties on achieving aggressive or clearly unrealistic forecasts
75. Management personally guaranteed significant debts of entity
76. Overly optimistic estimates with respect to asset impairment and/or extent of liabilities
77. Subjective judgments or uncertainties that are subject to potential significant change in the near term
78. Adverse consequences on significant pending transactions (such as a business combination or contract award) if poor financial results are reported
79. Aggressive sales or profitability incentive programs

Rapid business growth

80. Directors/management are poorly skilled, understaffed or inexperienced to deal with growth
81. Inability to attract competent personnel to run operations
82. Inability of control systems to adapt quickly enough to changing circumstances
83. Inadequate cash availability to meet growing demands for inventory, personnel and other major expenses
84. Growth is not profitable
85. Dealing with growth is distracting attention of board and senior management from addressing core business issues
86. Need for new capital
87. Public offering is in process or is anticipated
88. Rapid growth or profitability is unusual when compared to other companies in same industry
89. Little attention paid by board (or audit committee) to financial reporting and internal control
90. Poor control of tangible assets
91. Limited management knowledge, expertise or depth to administer new initiatives
92. Overly optimistic projections made of earnings
93. Undue emphasis on planning reported earnings
94. Lucrative personnel bonus or stock option plans based on performance
95. Overly optimistic or pessimistic estimates with respect to asset impairment and/or extent of liabilities

Major changes in business activities

96. Major acquisitions, divestitures or reorganizations
97. Going public or new sources of financing
98. New product launch or production facility
99. Start of operations in foreign jurisdictions
100. Change of control
101. Stated intentions by controlling shareholders to sell business
102. Public interest in entity
103. Significant new contracts
104. Adverse consequences on significant pending transactions (such as business combination or contract award) if poor financial results are reported
105. Commitments to bankers, analysts, creditors or other third parties on achieving aggressive or clearly unrealistic forecasts
106. Limited management knowledge, expertise or depth to administer new initiatives
107. Overly optimistic projections made of earnings
108. Overly optimistic or pessimistic estimates with respect to asset impairment and/or extent of liabilities
109. Significant debt guarantees provided by management
110. Material transactions or adjustments likely to occur near end of accounting periods
111. Lack of documentation of major transactions

112. Significant transactions with affiliated entities or related parties
Accounting judgments, estimates and disclosures
113. Issues regarding realization of assets, contingent liabilities or other unusual uncertainties 114. Estimates for allowance for bad debts, obsolete inventories, etc. 115. Costs allocated to inventory 116. Recoverability of equity investments 117. Impairment of long-lived assets 118. Recoverability of deferred charges 119. History of significant audit adjustments 120. Judgment required in timing of revenue recognition 121. Numerous adjusting entries through period 122. Earning projections made to bankers, creditors and other third parties 123. Stock option or performance plans tied to income 124. Contracts with payout clauses based on income 125. Desire to minimize tax liabilities 126. Excessive interest in maintaining or increasing entity's stock price or earnings trend through use of unusually aggressive accounting practices 127. Lucrative personnel bonus or stock option plans exist 128. Pursuit of inappropriate means to minimize reported earnings for tax-motivated reasons
The financial reporting process
129. No procedures or timetable in place for the period-end close 130. Persons assigned responsibility not trained 131. Roles of those involved not clearly defined 132. No cut-off requirements for inventory movements, purchases and sales 133. Period-end reconciliations not performed (e.g., bank accounts and intercompany balances) 134. No control over use of spreadsheets (formulas can get overridden) 135. No standardization of the software used 136. No supervision or review of the work performed 137. Poor intra-organizational communication 138. No benchmarks, performance measures or documentation standards
Auditor management relations
139. History of changing auditors, lawyers or other key advisors 140. Frequent disputes on accounting, auditing or reporting matters 141. Unreasonable demands/constraints in performing the audit and the issuance of the auditor's report 142. Restrictions that inappropriately limit auditor's access to people or information 143. Restrictions that limit auditor's ability to communicate effectively with TCWG (particularly the audit committee or equivalent) 144. History of receiving incomplete or misleading information 145. Overly optimistic estimates provided with respect to asset impairment and/or extent of liabilities 146. Consistent choice of aggressive accounting policies 147. Evasive answers to requests for information or access to people 148. Unreasonable requests to change audit staff members 149. Domineering management behaviour, especially attempts to influence the scope of the auditor's work 150. Provision of incomplete or misleading information

Potential for misappropriation of assets
151. Large amounts of cash on hand or processed regularly 152. Inventory that has a high value and/or high demand 153. Easily convertible assets (bearer bonds, diamonds, etc.) 154. History of asset theft 155. Poor physical safeguards over cash, investments, inventory or fixed assets 156. High number of insurance claims 157. Lack of procedures to screen job applicants for positions with access to susceptible assets 158. Poor physical or operational controls 159. Collusion with suppliers and customers 160. Financial stress of personnel or adverse relationships between entity and its personnel

Source: Own compilation based on CAS 240.

Acknowledgement

This PhD journey has been both pleasant and challenging at time. As I get closer to the finish line, I would like to express my sincere gratitude to Professor Csaba Makó who supported and challenged me for this entire journey. You have been a mentor and a role model for me the past four years. I am looking forward to collaborating with you in research projects in the future. I also want to thank Professor Tamas-Michael Koplyay for his support and collaborative research work the past four years. I am thankful for his timely interventions and advices during my PhD journey and I am looking forward to collaborating with you in research projects in the future. Finally, a special thank to a special person, Gen, who has been there the past fifteen years to encourage me and support me in my professional journey.