Ethical and Corporate Social Responsibility Implications of Artificial Intelligence Application for the Detection and Prevention of Financial Fraud

Maria-Elena Lindez-Macarro [(1)] and Felix Villalba-Romero [(2)]

(1) Faculty of Economics and Business Administration, Universidad Rey Juan Carlos, 28032 Madrid, Spain; elena.lindez@urjc.es

(2) Faculty of Economics and Business Administration, Universidad de Málaga, 29013 Málaga, Spain; fvr@uma.es

ABSTRACT

Artificial Intelligence (AI) has experienced rapid penetration in our society, playing a crucial role in everyday analysis and decision making. AI integration in the financial field stands out as one of the most advanced, especially in terms of financial fraud detection and prevention. Yet, this rapid evolution of the sector brings with it several ethical and corporate social responsibility (CRS) challenges that have not yet been properly explored. This analysis provides the essential principles for understanding the use of AI in the detection of financial fraud, assessing its effectiveness, and examining the tools used to understand the conceptual framework within which ethics and CRS are defined in the application of AI. A comprehensive review of the ethical and CRS challenges inherent in the use of AI in the detection and prevention of financial fraud is conducted, along with an analysis of the different perspectives present in the academic literature on the subject. In addition, future research directions are proposed to address these issues more comprehensively and effectively.

KEYWORDS: Artificial Intelligence; Financial Fraud; Ethical Challenges; Social Responsibility; Machine Learning; Generative Adversarial Networks; Impartiality, Transparency; Fraudulent Misinformation.

1. INTRODUCTION

As the implementation of artificial intelligence (AI) systems steadily progresses in the financial sphere, academics and researchers have turned their focus towards analysing the socio-economic and political implications of this emerging phenomenon (Gomes et al., 2020; Kamruzzaman, 2022; Kumar et al., 2023). Academic interest continues to grow given the potential of AI and the wide range of benefits AI provides such as automation and efficiency of processes (Iadanza et al., 2022), analysis of large datasets (Dutta et al., 2020), innovation and development (Mannuru et al., 2023) and process optimization (Igwegbe et al., 2023); as well as diverse application like medicine, marketing, finance or agriculture (Yang et al., 2021; Lorkowski et al., 2021; Hasan et al., 2023; Majeed et al., 2024)

However, the greater complexity of this technology has certain negative aspects. For the incorporation of AI, companies face challenges such as changing infrastructures, development and maintenance costs, and lack of specialised professionals. Nevertheless, the most important challenge lies in the ethical and corporate social responsibility (CSR), which pose unprecedented problems for society. The pervasiveness of AI raises several questions about equity, transparency, and automated decision-making, demanding rigorous scrutiny to mitigate potential negative implications for various strata of the population (Elliott et al., 2021). According to Zhao & Fariñas (2022), companies and governments must make collective efforts to achieve responsible and sustainable AI through a proactive regulatory framework backed by rigorous corporate policies and reporting (Vinuesa et. Al, 2020).

Other important issue lies on the way to integrate technological progress, such as AI implementation and cope with issues, considering the diverse business practices and cultural values of societies (Hofstede, 2001); and future AI impacts in internet business, perception and autonomy, as well as in Internet of Thing (IoT) and technologies such as blockchain (Lee et al., 2018). While in Western countries any AI implementation effort need to integrate individual rights and freedom to maximize profit, Asian countries place stronger emphasis on collective well-being and societal harmony. For example, AI implementation should emphasize on collective-welfare and institutional respect in China (Ding, 2018), respect hierarchical structures and employee welfare in Japan, or show technological prowess and social equity (Kim & Chung, 2019; Zhang et al., 2022). In this regard, a global approach needs to consider the diversity of culturally aligned CSR to be worldwide accepted (Wilson & van der Velden, 2022) or even include enough flexibility to adapt their AI strategies to different ethic and value priority, balancing innovation with ethical responsibilities (Minkkinen et al., 2022).

The integration of AI in the financial sector has focused on its application for the detection and prevention of financial fraud, leading to multiple lines of research and innovation (Ryman-Tubb et al., 2018; Li, 2022; Mill et al., 2023). The implementation of AI to combat financial fraud has led to quicker and more efficient identification of fraudulent actions, minimizing financial losses when fraud occurs (Bansal et al., 2024), as well as achieving better financial security, while providing a quick response to consumers affected by fraudulent activities and giving them more transparency and explainability in these situations (Zhou et al., 2023). Investment in security must be paramount as generative AI is also a tool used by fraudsters to their advantage, which explain the main part of the increase in fraudulent activity. As a results, main concerns stem from AI processes and how its mechanisms confront with ethics and social responsibility issues at the same time as fighting financial fraud, which have not fully been addressed by current research in financial fraud, which is the main contribution of this research.

The incorporation of AI in finance highlights its potential to transform fraud detection, enhance security, and improve transparency, but also underscores unresolved ethical challenges. This chapter explores how AI mechanisms intersect with ethics, social responsibility, and culturally diverse business practices, emphasizing the need for globally adaptable, responsible frameworks. By addressing these gaps, this research aims to contribute to a deeper understanding of AI's societal implications and its role in fostering innovation while safeguarding ethical principles.

The structure of this research consists of the following Section 2 which briefly present the methodology. Section 3 includes the theoretical framework and Section 4 the applications methodological development. Section 5 analyses the ethical and CSR challenges while section 6 presents the discussion and research agenda. Finally, section 7 contains the conclusions of the study.

2. METHODOLOGY

This research aims to set a theoretical framework that may be used as a base to apply AI to fraud detection and prevention, considering ethical and social responsibility implications. This framework offers basic definitions and requires conceptualisation, and responds to the main challenges derived from the use of AI in financial fraud detection from an ethical and social responsibility views.

Therefore, this research formulates the following research questions (RQs):

RQ1: Definition of AI, AI ethics and CSR in AI.

RQ2: How is AI applied to the detection and prevention of financial fraud?

RQ3: Is there published evidence that AI acts as an effective tool for the prevention and detection of financial fraud?

RQ4: What are the ethical and CSR challenges facing the use of AI as a tool for the detection and prevention of financial fraud?

The research questions posed align with the objectives of this research: (1) to understand what AI and how ethics and CSR are defined in this context, (2) to understand how AI is applied for fraud detection and prevention, (3) to check whether it is effective according to the existing literature, and (4) the ethical and CSR challenges that exist in the area.

To achieve the main objective of this work and answer the research questions, an extensive literature review has been conducted as well as the expert's reviews in the field.

Literature reviews play a crucial role in establishing a solid foundation for scientific research. Various methodologies have been developed to conduct effective literature reviews in different fields. Ramdhani and Ramdhani (2014) propose a verification methodology for research frameworks based on literature reviews, emphasizing analytical, conceptual, logical, and operational approaches. Chukwuere (2023) compares eleven different literature review methodologies in information systems research, highlighting their strengths and weaknesses to guide researchers in selecting appropriate methods. To enhance quantitative literature reviews, Laghrabli et al. (2015) introduce a novel framework using association rules analysis, which helps identify relationships between variables and explore new research directions. These studies collectively demonstrate the importance of selecting suitable literature review methodologies and offer innovative approaches to improve the quality and precision of research across various disciplines. A tool for semantic indexing and similarity queries can enhance literature review methods by improving efficiency, quality, and comprehensiveness (Koukal et al., 2014). As Cowell (2012), points out, literature reviews can be a rigorous research strategy if the scientific method is applied to ensure a standardized approach

As a result, the review of the literature has been integrated in a framework which outlines the main challenges in financial fraud detection and prevention.

3. THEORETICAL FRAMEWORK

In this section, we dive into a comprehensive analysis of AI concept and ethics and CRS from an AI perspective. Thus, research question 1 is answered in this section.

3.1.Artificial Intelligence (AI) Concept

As one of the fields of research and innovation that has evolved the most in recent years, AI is defined as the science that allows computers to execute tasks that require human intelligence (Finocchiaro, 2023). The European Commission, in its proposal for a European Regulation on artificial intelligence, establishes AI as: "a computer program developed with one or more of the techniques and approaches listed in Annex I and which, for a given set of defined objectives by the human being, generates results such as content, predictions, recommendations or decisions that influence the environments with which they interact" (European Parliament & Council, 2024).

AI has become an efficient and practical tool that is already part of a vast majority of activity sectors such as telecommunications, finance, and the automotive industry (Choi, 2021; Tan et al., 2022). The significant advance of AI is due to the expansion of applications and accessibility of tools, methods, and theories, facilitating their use (Cantrell & Zhang, 2018). Institutions as relevant in the field of robotics as the Stanford Robotics Center (SRC) work together with other institutions on AI projects. For example, in one of these projects, the SRC is trying to define the responsible use of AI technology together with the Human-Centered Artificial Intelligence Institute (HAI), with the aim of addressing the technical, social and economic challenges of robotics in areas such as health, education, sustainability and work (Stanford University, 2024). Computer scientists have identified three levels of AI based on the expected growth of its ability to analyse data and make predictions: narrow AI, general AI and superintelligence AI (Mersha et al., 2024). This classification is based on the level of capability and understanding that AI systems possess. AI systems designed to perform specific tasks without general awareness and understanding of the world are called weak AI and operate within a limited range of functions, while general AI systems could understand, learn and apply knowledge in a human-like manner. This includes the ability to reason, solve problems and understand complex concepts (Arrieta et al., 2019; Russell & Norvig, 1995; Searle, 1980). Nowadays, there is currently narrow AI and general AI. Most companies use narrow AI.

The evolution of AI can be traced through the advances driven by Google. At the beginning of the 21st century, machine learning was used for spell checkers, followed by Google Translate five years later. In 2012, deep learning began to be applied to speech recognition and large-scale training. In 2013, Word2Vec revolutionised natural language processing, while AtariDQN did the same with deep learning. Subsequently, neural networks and methodologies such as distillation were introduced, making it easier to work with complex models and facial recognition. Since 2016, investment in AI for security, preventive medicine, genomics and ethical principles has intensified. In recent years, advances have been almost monthly, achieving consistent milestones in natural language processing, image generation, conversational modelling, music, video and medical diagnostics (Google, 2025).

AI researcher Nick Bostrom defines the superintelligence as "an intellect far smarter than the best human brains in virtually every field, including scientific creativity, general wisdom, and social skills." (Bostrom, 2014). We are likely to see general AI appear in our lifetime, endowing machines with the ability to interact in human-like ways when working alongside humans (Baum, 2018; Brundage, 2015).

3.2.AI Ethics

Ethics is the set of moral principles that guide human interaction and decision-making, with the objective of improving the general well-being of society, protecting personal integrity, human dignity, and the rights of the most vulnerable individuals (Wirtz et al., 2018). From the AI approach, ethics refers to morality and the preference for a certain line of action in the development, implementation, and use of AI (Jobin & Ienca, 2019; Kaplan & Haenlein, 2020; Lin, 2016), as well as the protection against the inaccurate or unfair bias that algorithms may apply and the threat to privacy (Davenport et al., 2019).

Traditional ethics have been affected in business processes from automated decision-making and the comprehensibility of business models to labour relations by AI systems (Sison et al., 2023). Therefore, the ethical responsibility of public and private companies developing AI technology becomes crucially important. (Martin, 2018).

3.3.Corporate Social Responsibility in AI

Corporate Social responsibility (CSR) is a form of anticipatory reflection on the possible impacts of research on society, which involves aligning scientific research with the needs of society, the environment and democratically held values (Kumar et al., 2022; Politi, 2024). In the context of AI, CSR involves considering how AI can contribute to social good and align

with shared values in society (Cath et al., 2017), as well as mitigate potential negative consequences and maximize benefits such as equity and transparency (Helfat et al., 2023).

AI forces companies to re-evaluate their performance in terms of transparency and accountability, organisational culture, responsibility in the use of technologies and collaboration between public and private interests (Fioravante, 2024). Hence, it is necessary to align the adoption of AI with the company's overall CSR strategy, using AI to improve the company's sustainability as well as to address social and environmental issues (Ahdadou et al., 2024; Pai & Chandra, 2022).

The fast progress of artificial intelligence in recent years has developed new concepts and generated new challenges, which require a consensus identification and regulatory effort for appropriate management. A conceptual framework, developed in this section, is needed to capture the new concepts and ethical and CRS implications, in the implementation of diverse AI applications.

4. AI APPLICATION TO THE DETECTION AND PREVENTION OF FINANCIAL FRAUD

Throughout this section, the main AI trends and technologies in the field of financial fraud and their effectiveness will be developed according to existing literature. Thus, research questions 2 and 3 are answered in this section.

4.1. Transaction Pattern Analysis

Monitoring financial transactions may be crucial to detect and prevent financial crime, such as fraud, money laundering and terrorist financing. Fraudsters use AI for the creation of fake content, more persuasive phishing that is personalised to the target audience, or attacks against biometric systems such as facial, voice or fingerprint recognition systems in banking access or security systems, which can lead to huge financial damage and loss of consumer trust. Pattern analysis involves identifying and analysing modes of behaviours which may suggest illegal activity. Using the pattern analysis technique, AI can analyse large volumes of financial data, identifying patterns in transactions and anomalies that may indicate fraudulent activities (Zheng et al., 2024). When AI algorithms identify anomalous patterns, they proceed to segment financial data into different categories, facilitating the identification of atypical behaviours

within each segment (Dasari & Kaluri, 2024). Additionally, this tool can include the identification of correlations between different financial variables and the prediction of possible fraud scenarios based on the available data (Innan et al., 2023), as well as data visualization tools that simplify the detection of trends, patterns, and anomalies in financial data (Qiu & Luo, 2024).

Finally, pattern analysis-based systems offer the ability to detect fraud in real time by continuously monitoring financial transactions and triggering alerts when suspicious patterns are identified, while adapting and learning from new data to improve their accuracy in fraud detection (Maashi et al., 2023; Yang et al., 2023).

4.2.Machine Learning (ML)

Machine Learning (ML) is often defined as the capability of a machine to imitate intelligent human behaviour and thus, computer systems are able to learn and adapt by using algorithms and statistical models. This is a fundamental tool in data science that allows systems to learn and improve their performance with the experience gained from data (Ghoddusi et al., 2019). In the detection and prevention of financial fraud, ML extracts key information from data to predict and generate new insights (Zhou et al., 2023). The ML life cycle begins with Feature Engineering, which involves the selection and manipulation of data to transform it into useful features. Once the data is collected, automatic algorithms are generated capable of learning and acting autonomously in various situations, making it crucial to feed them with large volumes of data for their effectiveness (Martínez, 2016). The main strength of ML lies in its ability to continuously adapt and detect fraudulent activity in real time, marking a significant advance in cybersecurity. However, fraudsters also use it to carry out massive attacks and circumvent antivirus and protection systems.

Within ML, Supervised Learning and Unsupervised Learning are prominent approaches. The first uses annotated training data to guide the system in associating labels to training examples, making it valuable in problem classification (Sood et al., 2023). In contrast, Unsupervised Learning identifies patterns without specifying labels, and its models include Principal Component Analysis (PCA), which reduces the dimensionality of the data for better visualization and analysis (Beattie & Esmonde-White, 2021).

4.3.Natural Language Processing (NLP)

A machine learning technology that gives computers the ability to interpret, manipulate and comprehend human language is Natural Language Processing (NPL). Unlike structured

information, which can be organized in tables or matrices with neatly labelled rows and columns, unstructured information as the human language is confusing and difficult to understand.

To deal with unstructured information, computers start with segmentation, meaning take one sentence at a time. Then they divide the information into small pieces of information, called tokens, and once they have been sorted into a structure according to their meaning, NLP algorithms can work with them (Belz, 2022; Feder et al., 2022; Hamilton et al., 2024; Martinez, 2010).

The use of natural language processing (NLP) focuses on improving customer support and trend analysis, analysing fraud attempts by examining language and sentiment patterns (Faccia, 2023). Specifically, NPL detects inconsistencies or subtle clues in language that indicate whether an attempt is being made to hide information in a document (Chang et al., 2022; Faccia et al., 2023).

Cheng and Cai (2023) state that the use of NLP has led to a significant improvement in fraud detection capabilities. Along the same lines, Ebner et al. (2023) consider it a crucial technique to understand the content of messages and proactively alert people at risk, significantly contributing to the protection of consumers against financial fraud.

4.4.Neural Networks

A neural network is an AI method that teaches computers to process data and mimics the communication of neurons in the human brain using electronic circuits. Neurons receive signals, modify them, and transmit them to other neurons via axons. In this system, the perceptron acts as an individual neuron, with an input layer, hidden layers and an output layer. The signal enters the input layer, is processed by algorithms in the hidden layers, and the result is transmitted to the output layer (Guresen & Kayakutlu, 2011; Mangrulkar, 1990; Medhat, 2012; Xie et al., 2022).

The popularity of neural networks as a Deep Learning (DL) technique is due to their ability to identify and combine crucial features from unstructured data, achieving high performance without any domain knowledge (Zhu et al., 2021). In their study, Krambia-Kapardis et al. (2010) obtained an average of 90% accuracy in the fraud detection prediction model using artificial neural networks, thus demonstrating that they can be used to identify companies prone to fraud.

Likewise, Cherif et al. (2023) highlights that the use of AI such as neural networks is becoming essential for the detection of credit card fraud, while Almazroi and Ayub (2023) underlines its importance for detecting fraud in online payments. Projects such as those by Adler and Shavit (2024), aimed at understanding neural networks, focus on superposition computation, which brings significant advances in improving the applicability and efficiency of these networks in various disciplines.

4.5. Adversary Generative Networks (GANs)

Generative Adversarial Networks (GANs) are a system of two neural networks, the generator and the discriminator, which compete during training. The generator creates synthetic data similar to real data, while the discriminator attempts to distinguish between generated and real data (Cheah et al., 2023; Zhang et al., 2023).

In financial fraud detection, GANs generate synthetic samples of fraud data to address class imbalance. This interaction improves the generation of minority fraud samples, balancing the proportion between fraud and non-fraud in the data sets, which strengthens the detection of financial frauds (Chen et al., 2018; Jin et al., 2020). GANs have experienced significant development due to their versatility and usefulness, with multiple variants being proposed to improve their applicability (Strelcenia & Prakoonwit, 2023).

4.6. Social Media and Online Behaviour

Social media are online platforms that allow users to connect, interact, and share information with other people over the Internet (Lai et al., 2017). These platforms facilitate communication and interaction between individuals, groups, and organizations, thus creating virtual communities where relationships can be established, collaborate, share ideas and stay informed on various topics (Guo et al., 2021).

Social media can be used to detect financial fraud in various ways, such as through data exploration, analysis of relationships and connections, extraction of relevant data, implementation of algorithms and automated processes or the use of metadata and provenance that facilitate the organization of the information collected (Jamshidi & Hashemi, 2012; Diaz-Granados et al., 2015). Analysing behaviour patterns which derivates from the engagement with the online environment is used in fraud detection.

In this section main AI applications, namely Transaction Pattern Analysis, Machine Learning (ML), Natural Language Processing (NLP), Neural Network, Generative Adversarial Networks (GANs), the use of social media and online behaviour, have been identified and briefly introduced, as the most relevant applications for the financial fraud detection and prevention.

Nevertheless, an effective use of these applications raises ethical concerns and issues which means corporate social responsibility challenges and ethical dilemmas. Moreover, corporations and organisations need to address and manage these challenges and involve key stakeholders to face them.

5. ETHICAL AND CORPORATE SOCIAL RESPONSIBILITY CHALLENGES IN AI APPLICATIONS FOR FINANCIAL FRAUD DETECTION AND PREVENTION

AI is part of our lives, and its influence is growing, so it is crucial that people who design, develop, deploy, acquire, and use AI understand how to use AI ethics and CRS to minimize harm and optimize benefits. This section exhaustively analyses the main ethical and social responsibility challenges that AI applications generate for companies and consumers. Accordingly, this section addresses the answer to research question 4.

5.1. Privacy

Aiming to train AI models to deal with financial fraud, massive amounts of data need to be incorporated into them (Cui et al., 2022). Banking entities and digital financial corporations (Fintechs) store in their systems enormous amounts of personal data of their clients, as well as their daily financial transactions (Gai et al., 2017). This data can be classified as personal information (PI), relating to an identified or identifiable person, such as a name or postcode, and as sensitive personal information (SPI) which, if compromised, could be misused to harm or cause problems for an individual (Carmody et al., 2021; Fayoumi et al., 2022).

If a model is trained with personal or sensitive information without applying any privacy controls, there is a risk that this data may be used improperly, compromising the privacy and security of customers, and losing control over the possession of their data (Albayati et al., 2020; Cam & Kiet, 2023; Khan et al., 2023). However, another issue arises. The definition of privacy differs according to the country we are in, as well as the types of data to which this term applies (Walters et al., 2019).

As stated by Firdaus et al. (2018), it is crucial to find ways to ensure data confidentiality while using AI to effectively detect and prevent fraud. Privacy controls that can be applied during model training as model anonymisation and differential privacy or after model training, data minimisation. In the process of anonymisation, identifiable information is removed or modified from individuals' data to protect their privacy (Galbusera & Cina, 2024; Weber et al., 2024), while differential privacy ensures that the probability of any outcome of a learning algorithm does not change when modifying a single record in the training data, so that an adversary cannot infer information about a specific record (Zhu & Yu, 2019; Zhu et al., 2021). Finally, data minimisation reduces the amount of data that is collected and processed by limiting the SPI that is held and shared, helping to protect users' privacy (Razaque et al., 2022; Yaraziz et al., 2022).

5.2. Explainability

The importance of explainability lies in its epistemic and ethical benefits, as it provides insight into the extent to which the use of AI models meets the standards set by the other principles (Adams, 2023). AI tools must be able to provide clear and understandable explanations of how they reached certain conclusions or decisions (Nayak & Chandiramani, 2022; Harbinja et al., 2023), which helps to reduce errors and anticipate the strengths and weaknesses of the model, as well as to avoid unexpected behaviour in production (López-Úbeda et al., 2023).

In the pursuit of model explainability, it is important not to confuse with interpretability, as these are different ways of understanding how the model works. Interpretability is the degree to which an observer can understand the cause of a decision. It is the success rate with which humans can predict the outcome of an AI output, while explainability goes a step further and examines how the AI system arrived at an outcome (Aslam et al., 2022; Nicodeme, 2020; Tiwary et al., 2024). Improving explainability and interpretability of AI models allows companies to remain competitive in the marketplace, gain customer confidence and build trust in reliable AI (Adadi & Berrada, 2018; Firdaus et al., 2022; Pagliari et al., 2022).

5.3. Transparency

Transparency has become an essential pillar of the implementation of AI for the detection and prevention of financial fraud. Financial institutions are responsible for facilitating customers' understanding of how automated fraud detection systems work, thus avoiding

undermining trust in the system and raising concerns about privacy protection (Van Bekkum & Borgesius, 2021).

Ensuring the collection of personal and financial data through appropriate customer consent is a key challenge that financial institutions must address (Yerima & Bashar, 2022). The absence of transparency can lead to social and psychological consequences such as distrust of technology and a strong sense of powerlessness, which can have a profound impact on mental health (Nannini et al., 2024).

Transparency is about revealing information related to the data used to build AI systems, making design decisions along the way, creating models, evaluating the models and implementing the models (Johnson et al., 2022). Governance ensures that the process followed during creation and implementation follows internal policies to create reliable AI, increase public confidence and legitimise the decisions taken (De Fine Licht & De Fine Licht, 2020; Win & Beydoun, 2020).

By increasing transparency, we can improve governance. Therefore, it is important that the roles of the participants in the AI systems process and the responsibilities that each of them takes are well defined, harmonising the processes carried out (Francisco & Linnér, 2023; Sakyoud et al., 2023; Shneiderman, 2020).

5.4.Impartiality

Training plays a critical role in the development of machine learning and AI systems for financial fraud detection. This process involves the use of models and historical data sets, which can introduce inherent biases if not managed properly (Dong et al., 2022). In general, bias is a systematic error, but in the context of fairness, the problem revolves around unintended bias. Unintended bias gives some groups or individuals a systematic advantage and other groups or individuals a systematic disadvantage, perpetuating social inequalities (Chu et al., 2021; Ranard et al., 2024). In the judicial context, for example, the influence of implicit biases in judges' decisions can compromise the perception of justice and fairness in the legal system. This can lead to widespread distrust of judicial institutions, affecting the legitimacy of the system and people's willingness to participate in it (Schneider & Weber, 2024).

These biases, which may be related to demographic variables such as gender, race, or age, have the potential to bias associations between certain fraudulent behaviours and specific population characteristics (Cruz et al., 2021). Consequently, AI models can inadvertently

replicate existing social biases and generate discriminatory decisions (Fiore et al., 2019; Zheng et al., 2023).

This phenomenon highlights the critical importance of addressing algorithmic bias during the training of AI models to ensure fairness and objectivity in their application and the need for human oversight and continuous evaluation in the development of rules for algorithmic software (Simos et al., 2022).

5.5.Solidity

AI provides multiple and sophisticated tools to combat financial fraud, but each of them has weaknesses that make them vulnerable to criminal attacks (Hu et al., 2021). Analyzing the vulnerabilities and robustness of the models applied in fraud detection is imperative for companies that use them, especially financial and insurance companies due to the characteristics of the information they possess (Amerirad et al., 2023). Different techniques can be used in this process, such as the DeepFense online defence framework proposed by Rouhani et al. (2018) or the automated verification framework based on Satisfiability Modulo Theory (SMT) by Huang et al. (2017). However, the goal of different methodologies should be the same: to develop robust and resilient countermeasures for different types of likely adversary scenarios to provide reliable infrastructure in AI environments (Sengupta et al., 2020). Interacting with AI systems that fail or are perceived as unreliable can generate stress, anxiety and frustration for users. This is especially relevant in contexts where AI is used for critical tasks or where it is expected to improve quality of life. It is therefore imperative to ensure the robustness of these technologies (Hamon et al., 2024).

5.6.Responsibility and Accountability

Accountability focuses on the responsibility that people, organizations, or institutions have to inform and justify their actions before interested parties, assuming the consequences of the decisions made, being a fundamental principle in governance, management, and decision-making process (Kieslich et al., 2022; Taher et al., 2024). The complexity of AI models has made it difficult to determine the responsible authority, hindering accountability and the protection of basic rights of citizens (Horneber & Laumer, 2023; Zajko, 2023).

According to London (2019), it is essential to promote accountability and ensure that machine learning systems do not become covert tools to arbitrarily interfere with the autonomy of stakeholders. The importance of accountability has led the European Commission to incorporate it as one of the seven requirements of trustworthy AI, while highlighting auditability

and the guarantee of adequate and accessible remediation as key factors to consider in the implementation of AI (European Commission, 2019).

5.7.Impact on Decision Making

The impact of using AI on decision making in financial organizations is significant (Sabharwal et al., 2024). According to Pinheiro et al. (2023), the development of intelligent routines to support complex decision making is not always simple, presenting difficulties related to the abundance of available data sources, the number of legal regulations that must be met, and the need to incorporate transparency, auditability, standardization, and desirable reuse in information technology systems. In compliance with Giest & Klievink (2022), the integration of AI in decision-making should be encourage, as it promotes innovation in areas such as the public sector.

However, the automation generated by AI tools does not exempt the need for human supervision or the importance of a correct interpretation of the results obtained to make informed decisions, especially in critical situations where decisions based solely on algorithms could have significant consequences (De Barros et al., 2017; Jesus et al., 2021). Likewise, companies must ensure effective appeal mechanisms in case of errors made in decision-making, both because of humans and AI (Kanika et al., 2022).

5.8. Legislation and Regulatory Framework

The Council and the European Parliament have reached a provisional agreement on what is the first Artificial Intelligence (AI) Regulation, based on the European Commission's 2021 proposal to create the EU's first regulatory framework for AI (European Commission, 2019; European Commission, 2023). The key objectives of this law are to ensure that Artificial Intelligence systems introduced on the European market are safe and respect citizens' rights while stimulating investment and innovation in the field of AI in Europe (European Commission, 2023).

At a global level, politicians and competent authorities are working to review their country's financial, data protection and administrative legislation (Huh, 2022; Ridzuan et al., 2024), promoting the importance of complying with constitutional principles and human rights to ensure transparency and explainability of AI systems, which may require a strong legal framework (Kuzniacki et al., 2022; Lee et al., 2018). In 2023, AI became a key US policy issue, driven by Biden's executive order, which established risk-based industry standards and promoted transparency. In 2024, these policies are expected to be implemented through the new

Artificial Intelligence Safety Institute, while Congress evaluates proposals on transparency, deepfakes and platform accountability, however, the new US president has repealed the law against AI risks (Jiménez, 2025; MIT Technology Review, 2024).

Meanwhile, regulation of AI in China has been fragmented, with specific rules for areas such as algorithms and deepfakes. In 2023, a possible comprehensive law was announced, although implementation could take years. In the meantime, companies must register their models with the government, which limits foreign competition and reinforces state control, favouring Chinese companies but restricting competition and online expression.

A global increase in AI regulation is expected in the coming years, with Africa highlighted by the African Union's potential strategy to boost competitiveness and protect consumers. Countries such as Rwanda, Nigeria and South Africa are already moving forward with national policies. Bodies such as the UN and OECD are working on standards that could facilitate regulatory harmonisation at a global level. At the geopolitical level, differences between democratic and authoritarian approaches to AI development will become more pronounced, forcing companies to decide whether to prioritise global expansion or focus on local markets.

Assuring security from the development phase of AI technology and continuously monitoring and updating to prevent problems is essential to steer the development of AI in a positive direction (Jang, 2024). Legislative authorities face new challenges such as the legal personification of AI, responsibility for the consequences of the implementation of these tools and the continuous evolution and transformation of AI technology that makes legislation obsolete and inadequate in noticeably short periods of time (King et al., 2019; Azzutti, 2022; Novelli, 2022).

5.9. Fraudulent Misinformation

Recent advancements in AI technologies have exacerbated the challenges of misinformation and financial fraud.

Generative AI models can produce compelling but fabricated content, complicating detection efforts (Xu et al., 2023; Shoaib et al., 2023). Manipulating narrative frames into precise information can generate misinformation, requiring innovative detection approaches using large language models and deep neural networks (Wang et al., 2024).

To combat AI-generated misinformation, researchers propose examining manipulation traces at signal, perceptual, semantic, and human levels (Xu et al., 2023) and multifaceted strategies including advanced detection algorithms, cross-platform collaboration, and policy initiatives (Shoaib et al., 2023).

In the financial sector, artificial intelligence and machine learning techniques offer promising solutions for fraud prevention by identifying legitimate and fraudulent behaviors (Agarwal, 2021). However, challenges remain, such as evolving fraud patterns and the need for model interpretability.

While data science in finance offers opportunities for improved customer experiences and cutting-edge solutions, it also raises ethical concerns, potential biases, and data security issues (Zheng et al., 2023). Fraudsters are increasingly using AI-based tools such as WormGPT and FraudGPT to create misleading information and scam unsuspecting victims, requiring the evolution of fraud detection techniques (Sina, 2023). The proliferation of Artificial Intelligence Generated Content (AIGC), which uses AI to assist or replace manual content generation by generating content based on user-inputted keywords, further complicates the detection of misinformation, making traditional approaches inadequate (Xu et al., 2023).

As financial institutions adopt advanced technologies such as OpenAI to protect customers from fraud, they must also address the ethical challenges posed by generative AI in algorithmic trading and fraud detection (Zheng et al., 2023; Sina, 2023).

This section has presented most relevant challenges found in the literature derived from AI Implementation for the financial fraud detection and prevention, as well as important considerations on some of the most AI-based tools and techniques, explaining some of the difficulties to cope with the presented challenges.

The result of this research is a holistic theoretical framework for financial fraud detection and prevention, which gather a conceptualisation process including concepts, ethical and social responsibility concerns for the framing of AI applications, that eventually contribute to copying the identified challenges, as is represented in figure 1.

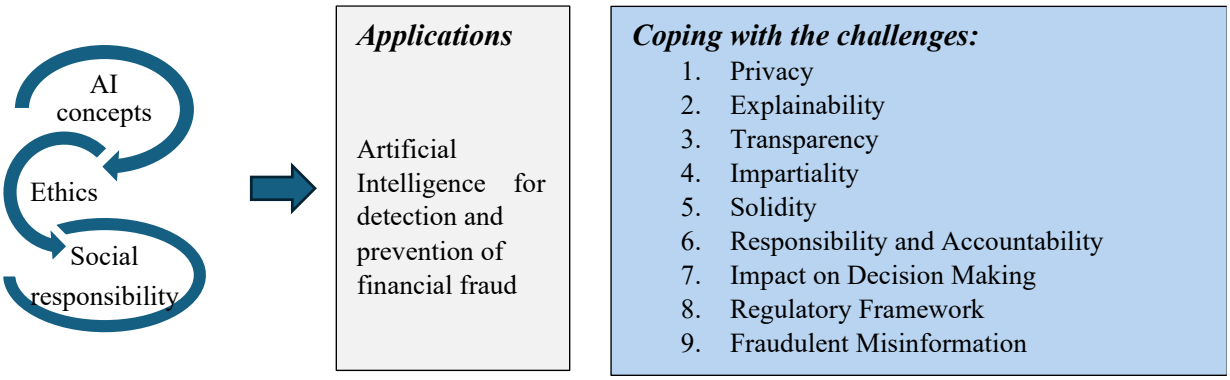

Figure 1: Financial fraud detection and prevention theoretical framework

6. DISCUSSION AND RESEARCH AGENDA

In recent years, the exponential penetration of AI has been pivotal, triggering a profound transformation across a wide range of economic sectors and catalysing the transition to the fourth industrial revolution (Mannuru et al., 2023). AI impact is particularly relevant in the financial sector and cybersecurity, becoming a protection mechanism for consumers (Kim, 2019). Companies such as and Paypal and Capital One, reports significant progress using data science techniques and machine learning models to detect and prevent financial fraud and surveys show great customers satisfaction. Similarly, credit cards companies, such as Visa and MasterCard, have used AI applications. This shows a trend to use technology-based detection services for many companies in the financial sector. Nonetheless, the rapid development of AI and the increasing reliance on AI tools bring the emergence of ethical threats and dilemmas, as well as concerns related to social responsibility (Borenstein et al., 2021; Biondi et al., 2023; Van Den Berg et al., 2024). Moraes (2024) highlights the relevance of integrating experts from multiple disciplines and sectors in the creation of regulatory frameworks for artificial intelligence (AI). This encompasses the collaboration of regulators, private sector representatives, academics and members of civil society, promoting a broader and more diverse view on the challenges and possibilities posed by AI.

Some notable episodes have highlighted the ethical challenges inherent in the implementation of AI in business, such as evidence of gender bias in Amazon's AI-based recruitment tool, the manifestation of racist and sexist comments by Microsoft's chatbot, which culminated in its retirement, as well as the deadly incidents associated with Tesla's autonomous systems (Sison et al., 2023). Other case includes discriminatory behaviour in ranking algorithm affecting the decision-making process. In fraud detection, there have also been cases where AI

presents ethical challenges, such as the SyRI case in the Netherlands, where cases of unfair persecution and the propagation of xenophobic stereotypes were reported (Züger & Asghari, 2022; Newman & Mintrom, 2023). On the positive side, it is also reported the UNICEF's Project Connect which may be a good example to overcome the solidity challenge (Züger & Asghari, 2022). Chinese companies such as Alibaba and Tecent have generated concerns on privacy data (Verma et al., 2021; Nishant et al., 2020) but also are considered like a role model in privacy policy within the country after deep changes in their policy, though they force consumers to accept it private data policies (Fu, 2019). Other companies, i.e. Softbank poses human oversight concerns in AI development and governance (Sigfrids et al., 2023).

These incidents are a tangible demonstration of the imperative to continue researching and improving AI tools, ensuring their alignment with ethical guidelines and the fundamental values of society.

Table 1 Examples of ethical concerns in case studies

Case study	**Ethical concern**	**Challenge**	**Reference**
Amazon	Evidence of gender bias	Impartiality	Sison et al., 2023 Claudy et al, 2022
Microsoft's chatbot	Racist and sexist comments	Transparency	Sison et al., 2023 Daza & Ilozumba, 2022
Deliveroo, food delivery platform	Discriminatory behaviour in reputational-ranking algorithm	Impact on Decision Making	Piccininni, 2022
Tesla's autonomous systems	Deadly incidents associated	Responsibility and accountability	Sison et al., 2023 Leben, 2023
SyRI (Dutch welfare fraud detection project)	Unfair persecution and the propagation of xenophobic stereotypes. No public interest AI approach, nor deliberative and participatory design	Misinformation Regulatory Framework	Züger Asghari, 2022 Newman & Mintrom, 2023
UNICEF's Project Connect	Public interest AI, open for validation to others, use an open-source tool	Solidity	Züger & Asghari, 2022
Alibaba (e-commerce platform)	data privacy and consumer autonomy	Privacy	Fu, 2019 Verma et al., 2021
Softbank	human oversight to ensure fairness and transparency	Fairness and transparency	Sigfrids et al., 2023
Tecent (Tech conglomerate)	data privacy, potential misuse of technology, surveillance,	Privacy Fraudulent misinformation	Fu, 2019 Nishant et al., 2020

In their work to detect financial fraud, the various AI models require high-quality financial transaction data, machine learning algorithms, the construction of predictive models based on historical data, the ability to adapt to new fraudulent techniques, and the ability to interpret results to assist analysts, requiring high investments by companies to adapt their infrastructures, train staff in new processes and recruit specialised personnel (Kapadiya et al., 2022; Mishra, 2023; Souza et al., 2024). In addition to these barriers to incorporating AI, companies may face difficulties in finding huge amounts of reliable data to feed into these systems, adaptation to legacy systems from other companies, legal restrictions that restrict the proper training of algorithms, or employee resistance to incorporating AI into operational processes.

The massive collection of personal data for the training of AI algorithms poses significant challenges in terms of privacy and protection of sensitive consumer data (Awosika et al., 2024; Baabdullah et al., 2024; Deng et al., 2023; Zaimi et al., 2023). This carries the inherent risk of potential misuse or compromise of personal information during fraud detection operations, which can lead to a profound negative impact on consumer trust (Mhlanga, 2020; Xiong et al., 2021). Furthermore, if training data contains biases, such as racial or gender discrimination, algorithms can perpetuate and amplify these biases (Lokanan, 2022; Truby, 2020), leading to unfair prosecution, inequity, and errors in decision-making. The complex nature of the Artificial Intelligence algorithms used in financial fraud detection can complicate the understanding of the decisions made, which impacts on the identification of possible errors in the process (Mytnyk et al., 2023). This complexity can result in a lack of transparency and explainability in the procedure (Boustani, 2021; Fukas et al., 2022; Kesa & Kerikmäe, 2020). Due to the ethical dilemmas present, it is imperative to institute mechanisms that clarify decision-making responsibility and ensure accountability within a sound legal framework when AI is used for the detection and prevention of financial fraud (Yang et al., 2021; Kong et al., 2024). This framework should include an appropriate compensation system for clients and consumers in case of harm (Ali et al., 2022).

To address these challenges, global governmental entities, international organisations, and technology giants have implemented concerted actions and reached consensus to build trust in AI (Zhang et al., 2021). These actions focus on data protection or the auditing of algorithms. However, the scale of AI's negative influence requires going further by regulating clear legal frameworks at the international level, encouraging ethical oversight and combating misinformation. Additionally, considerable international research and public-private

collaboration is required to promote the benefits of AI, while managing and reducing the associated risks to consumers and society at large.

7. CONCLUSIONS

This study sheds light on an area that is continually evolving, the ethics and social responsibility in using AI to prevent and detect financial fraud. Throughout this research, both theoretical and empirical knowledge have been detailed to provide a global vision of the line of research. The analysis of the literature makes clear the importance of AI to detect financial fraud and how integrated this technology is in the financial sector and the detection of financial fraud. Continuing to research AI and its ethical and social implications can foster innovation and the advancement of numerous areas of knowledge. This research leads to the following conclusions:

1. The application of AI systems for the detection of financial fraud raises a significant number of ethical issues faced by financial institutions and governmental organisations.
2. The incorporation of customers' personal and financial information by banks and FinTechs raises ethical dilemmas related to privacy and data control by consumers, eroding their privacy and exposing them to the public and to fraudsters. This scenario exposes them to anxiety and fear of vulnerability, post-violation of privacy stress, distrust and feelings of powerlessness and lack of control.
3. At the same time, the various AI models aimed at detecting financial fraud exhibit deficiencies and vulnerabilities that make them susceptible to exploitation by criminals, leading to a direct threat to the security of customers' personal data, increasing feelings of helplessness and insecurity.
4. Datasets used in training algorithms can also lead to biased decisions if they contain inherent biases related to gender, ethnicity, or age, which could result in the erroneous association of certain demographic groups with fraudulent behaviour and the perpetuation of social inequalities, negatively impacting vulnerable groups.
5. Achieving transparency in the fraud detection procedure using AI is still a work in progress, which results in a lack of clarity as to the explainability of errors and creates mistrust among consumers.

6. The use of Artificial Intelligence has catalysed a significant change in the decision-making process, supported by the capabilities provided by this technology.
7. It is essential to thoroughly understand and analyse the results derived from the application of AI, while maintaining rigorous oversight of the entire process. The question of accountability for the actions undertaken by the various entities making use of AI emerges as a fundamental ethical dilemma that still requires further academic development, given its complexity and its compelling relevance in safeguarding citizens' rights.
8. In order to secure individuals and entities against the potential threats inherent in the use of Artificial Intelligence, it is essential to establish an international legal framework that is both robust and adaptable to the constant evolution of technological innovation.
9. While AI offers considerable potential to improve the detection and prevention of financial fraud, its implementation requires a cautious and multidisciplinary approach. Collaboration between public and private sectors, investment in research to mitigate bias and improve transparency, as well as the promotion of digital literacy among consumers, are essential aspects of addressing the ethical and social challenges intrinsic in this technology.
10. The path towards an ethical and corporate socially responsible use of AI in the detection and prevention of financial fraud involves a continuous commitment to critical reflection by institutions and academics, responsible innovation and the search for a balance between effectiveness in the fight against fraud and respect for individual and collective rights.

Research on the ethical and social responsibility implications of AI in financial fraud detection and prevention is a line of research with enormous potential that has not yet been sufficiently explored. Future lines of research could analyse the impact on fairness and accuracy of fraud detection as a function of algorithmic biases, as well as examine the impact of the use of AI in fraud detection on digital financial inclusion and consumer digital financial literacy and education. The establishment of a clear legal and ethical framework for the use of AI in financial fraud is essential to protect society from the risks associated with AI as well as encouraging ethical oversight and combating misinformation. Thus, the academic community needs to focus on these topics to provide policymakers and authorities with deeper specific

recommendations according to the discipline of implementation of the technologies and to the evolution of AI.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138-52160. https://doi.org/10.1109/access.2018.2870052

Adams, J. (2023). Defending explicability as a principle for the ethics of artificial intelligence in medicine. Medicine Health Care And Philosophy, 26(4), 615-623. https://doi.org/10.1007/s11019-023-10175-7

Adler, M. & Shavit, N. (2024). On the Complexity of Neural Computation in Superposition. CSAIL Technical Reports (July 1, 2003 - present). https://hdl.handle.net/1721.1/157073

Agarwal, S. (2021). Artificial Intelligence Techniques of Fraud Prevention. In Applications of Artificial Intelligence in Business and Finance (pp. 113-132). Apple Academic Press.

Ahdadou, M., Aajly, A., & Tahrouch, M. (2024). Enhancing corporate governance through AI: a systematic literature review. Technology Analysis And Strategic Management, 1-14. https://doi.org/10.1080/09537325.2024.2326120

Albayati, H., Kim, S. K., & Rho, J. J. (2020). Accepting financial transactions using blockchain technology and cryptocurrency: A customer perspective approach. Technology In Society, 62, 101320. https://doi.org/10.1016/j.techsoc.2020.101320

Ali, A., Razak, S. A., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. Applied Sciences, 12(19), 9637. https://doi.org/10.3390/app12199637

Almazroi, A. A., & Ayub, N. (2023). Online Payment fraud detection model using machine learning techniques. IEEE Access, 11, 137188-137203. https://doi.org/10.1109/access.2023.3339226

Amerirad, B., Cattaneo, M., Kenett, R. S., & Luciano, E. (2023). Adversarial Artificial Intelligence in Insurance: From an Example to Some Potential Remedies. Risks, 11(1), 20. https://doi.org/10.3390/risks11010020

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2019). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Aslam, N., Khan, I. U., Mirza, S., AlOwayed, A., Anis, F. M., Aljuaid, R. M., & Baageel, R. (2022). Interpretable Machine Learning Models for Malicious Domains Detection Using Explainable Artificial Intelligence (XAI). Sustainability, 14(12), 7375. https://doi.org/10.3390/su14127375

Awosika, T., Shukla, R. M., & Pranggono, B. (2024). Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection. IEEE Access, 1. https://doi.org/10.1109/access.2024.3394528

Azzutti, A. (2022). AI trading and the limits of EU law enforcement in deterring market manipulation. Computer Law And Security Report/Computer Law & Security Report, 45, 105690. https://doi.org/10.1016/j.clsr.2022.105690

Baabdullah, T., Alzahrani, A., Rawat, D. B., & Liu, C. (2024). Efficiency of Federated Learning and Blockchain in Preserving Privacy and Enhancing the Performance of Credit Card Fraud Detection (CCFD) Systems. Future Internet, 16(6), 196. https://doi.org/10.3390/fi16060196

Bansal, K., Paliwal, A. C., & Singh, A. K. (2024). Analysis of the benefits of artificial intelligence and human personality study on online fraud detection. International Journal Of Law And Management./International Journal Of Law And Management. https://doi.org/10.1108/ijlma-08-2023-0198

Baum, S. (2018). Countering superintelligence misinformation. Information, 9(10), 244. https://doi.org/10.3390/info9100244

Beattie, J. R., & Esmonde-White, F. W. L. (2021). Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra. Applied Spectroscopy, 75(4), 361–375. https://doi.org/10.1177/0003702820987847

Belz, A. (2022). A Metrological Perspective on Reproducibility in NLP*. Computational Linguistics, 48(4), 1125-1135. https://doi.org/10.1162/coli_a_00448

Biondi, G., Cagnoni, S., Capobianco, R., Franzoni, V., Lisi, F. A., Milani, A., & Vallverdú, J. (2023). Editorial: Ethical design of artificial intelligence-based systems for decision making. Frontiers In Artificial Intelligence, 6. https://doi.org/10.3389/frai.2023.1250209

Borenstein, J., Grodzinsky, F. S., Howard, A., Miller, K. W., & Wolf, M. J. (2021). AI Ethics: A Long History and a Recent Burst of Attention. Computer, 54(1), 96-102. https://doi.org/10.1109/mc.2020.3034950

Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.

Boustani, N. M. (2021). Artificial intelligence impact on banks clients and employees in an Asian developing country. Journal Of Asia Business Studies, 16(2), 267-278. https://doi.org/10.1108/jabs-09-2020-0376

Brundage, M. (2015). Taking superintelligence seriously. Futures, 72, 32-35. https://doi.org/10.1016/j.futures.2015.07.009

Cam, N. T., & Kiet, V. T. (2023). FlwrBC: Incentive Mechanism Design for Federated Learning by Using Blockchain. IEEE Access, 11, 107855-107866. https://doi.org/10.1109/access.2023.3320045

Cantrell, B., & Zhang, Z. (2018). A third intelligence. Landscape Architecture Frontiers, 6(2), 42. https://doi.org/10.15302/j-laf-20180205

Carmody, J., Shringarpure, S., & Van de Venter, G. (2021). AI and privacy concerns: a smart meter case study. Journal Of Information Communication And Ethics In Society, 19(4), 492-505. https://doi.org/10.1108/jices-04-2021-0042

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2017). Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. Science And Engineering Ethics. https://doi.org/10.1007/s11948-017-9901-7

Chang, J., Yen, N. Y., & Hung, J. C. (2022). Design of a NLP-empowered finance fraud awareness model: the anti-fraud chatbot for fraud detection and fraud classification as an instance. Journal Of Ambient Intelligence & Humanized Computing/Journal Of Ambient Intelligence And Humanized Computing, 13(10), 4663-4679. https://doi.org/10.1007/s12652-021-03512-2

Cheah, P. C. Y., Yang, Y., & Lee, B. G. (2023). Enhancing Financial Fraud Detection through Addressing Class Imbalance Using Hybrid SMOTE-GAN Techniques. International Journal Of Financial Studies, 11(3), 110. https://doi.org/10.3390/ijfs11030110

Chen, J., Shen, Y., & Ali, R. (2018). Credit Card Fraud Detection Using Sparse Autoencoder and Generative Adversarial Network. 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). https://doi.org/10.1109/iemcon.2018.8614815

Cheng, C., & Cai, W. (2023). Double-weight LDA extracting keywords for financial fraud detection system. Multimedia Tools And Applications. https://doi.org/10.1007/s11042-023-17334-1

Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., & Imine, A. (2023). Credit card fraud detection in the era of disruptive technologies: A systematic review. Journal Of King Saud University. Computer And Information Sciences/Mağalaẗ Ğam'aẗ Al-malīk Saud : Ùlm Al-ḥasib Wa Al-ma'lumat, 35(1), 145-174. https://doi.org/10.1016/j.jksuci.2022.11.008

Choi, Á. (2021). España ante la Revolución Industrial 4.0: mercado laboral y formación. Araucaria, 47, 479-505. https://doi.org/10.12795/araucaria.2021.i47.21

Chukwuere, J. E. (2023). Exploring Literature Review Methodologies in Information Systems Research: A Comparative Study. Education & Learning in Developing Nations (ELDN), 1(2), 38-46.

Claudy, M. C., Aquino, K., & Graso, M. (2022). Artificial intelligence can't be charmed: The effects of impartiality on laypeople's algorithmic preferences. Frontiers in Psychology, 13, 898027.

Comisión Europea (CE) (2021). Proposal for a Regulation of the European Parliament and of the Council on Harmonised Rules in the field of Artificial Intelligence (Artificial Intelligence Act) and amending certain legislative acts of the Union. Brussels, 21.04.2021. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52021PC0206

Cowell, J. M. (2012). Literature reviews as a research strategy. The Journal of School Nursing, 28(5), 326-327.

Cruz, A.F., Saleiro, P., Belem, C., Soares, C., & Bizarro, P. (2021). Promoting Fairness through Hyperparameter Optimization. 2021 IEEE International Conference On Data Mining (ICDM). https://doi.org/10.1109/icdm51629.2021.00119

Cui, X., Chang, Y., Yang, C., Cong, Z., Wang, B., & Leng, Y. (2022). Development and Trends in Artificial Intelligence in Critical Care Medicine: A Bibliometric Analysis of Related Research over the Period of 2010–2021. Journal Of Personalized Medicine, 13(1), 50. https://doi.org/10.3390/jpm13010050

Dasari, S., & Kaluri, R. (2024). An Effective Classification of DDoS Attacks in a Distributed Network by Adopting Hierarchical Machine Learning and Hyperparameters Optimization Techniques. IEEE Access, 12, 10834-10845. https://doi.org/10.1109/access.2024.3352281

Davenport, T. H., Guha, A., Grewal, D., & Breßgott, T. (2019). How artificial intelligence will change the future of marketing. Journal Of The Academy Of Marketing Science, 48(1), 24-42. https://doi.org/10.1007/s11747-019-00696-0

Daza, M. T., & Ilozumba, U. J. (2022). A survey of AI ethics in business literature: Maps and trends between 2000 and 2021. Frontiers in Psychology, 13, 1042661.

De Barros, R. S. M., De Lima Cabral, D. R., Gonçalves, P. M., & De Carvalho Santos, S. G. T. (2017). RDDM: Reactive drift detection method. Expert Systems With Applications, 90, 344-355. https://doi.org/10.1016/j.eswa.2017.08.023

De Fine Licht, K., & De Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. AI & Society, 35(4), 917-926. https://doi.org/10.1007/s00146-020-00960-w

Deng, W., Liang, G., Yu, C., Yao, K., Wang, C., & Zhang, X. (2023). An Early Warning Model of Telecommunication Network Fraud Based on User Portrait. Computers, Materials & Continua/Computers, Materials & Continua (Print), 75(1), 1561-1576. https://doi.org/10.32604/cmc.2023.035016

Diaz-Granados, M., Diaz-Montes, J., & Parashar, M. (2015). Investigating insurance fraud using social media. Proceedings 2015 IEEE International Conference on Big Data, pp.1344-1349. https://doi.org/10.1109/bigdata.2015.7363893

Ding, J. (2018). Deciphering China's AI Dream: The Context, Components, Capabilities, and Consequences of China's Strategy to Lead the World in AI. Future of Humanity Institute, University of Oxford. https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf

Dong, Y., Liu, N., Jalaian, B., & Li, J. (2022). EDITS: Modeling and Mitigating Data Bias for Graph Neural Networks. Proceedings Of The ACM Web Conference 2022. https://doi.org/10.1145/3485447.3512173

Dutta, A., Deb, T., & Pathak, S. (2020). Automated Data Harmonization (ADH) using Artificial Intelligence (AI). OPSEARCH/Opsearch, 58(2), 257-275. https://doi.org/10.1007/s12597-020-00467-4

Ebner, N., Pehlivanoglu, D. & Shoenfelt, A (2023). Financial Fraud and Deception in Aging. Advances In Geriatric Medicine And Research. https://doi.org/10.20900/agmr20230007

Elliott, K., Price, R., Shaw, P., Spiliotopoulos, T., Ng, M., Coopamootoo, K., & Van Moorsel, A. (2021). Towards an Equitable Digital Society: Artificial Intelligence (AI) and Corporate Digital Responsibility (CDR). Society, 58(3), 179-188. https://doi.org/10.1007/s12115-021-00594-8

European Commission (2019) Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

European Commission. (2023). The Commission welcomes the political agreement on the Artificial Intelligence Act [Press release] https://ec.europa.eu/commission/presscorner/api/files/document/print/es/ip_23_6473/IP_23_6473_ES.pdf

European Parliament & Council. (2024). Regulation laying down harmonised rules on artificial intelligence (PE-CONS 24/24). (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

Faccia, A. (2023). National Payment Switches and the Power of Cognitive Computing against Fintech Fraud. Big Data And Cognitive Computing, 7(2), 76. https://doi.org/10.3390/bdcc7020076

Faccia, A., McDonald, J. A. K., & George, B. (2023). NLP Sentiment Analysis and Accounting Transparency: A New Era of Financial Record Keeping. Computers, 13(1), 5. https://doi.org/10.3390/computers13010005

Fayoumi, M. A., Odeh, A., Keshta, I., Aboshgifa, A., AlHajahjeh, T., & Abdulraheem, R. (2022). Email phishing detection based on naïve Bayes, Random Forests, and SVM classifications: A comparative study. 2022 IEEE 12th Annual Computing And Communication Workshop And Conference (CCWC). https://doi.org/10.1109/ccwc54503.2022.9720757

Feder, A., Keith, K. A., Manzoor, E., Pryzant, R., Sridhar, D., Wood-Doughty, Z., Eisenstein, J., Grimmer, J., Reichart, R., Roberts, M. E., Stewart, B. M., Veitch, V., & Yang, D. (2022). Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond. Transactions Of The Association For Computational Linguistics, 10, 1138-1158. https://doi.org/10.1162/tacl_a_00511

Finocchiaro, G. D. (2023). The regulation of artificial intelligence. AI & Society. https://doi.org/10.1007/s00146-023-01650-z

Fioravante, R. (2024). Beyond the Business Case for Responsible Artificial Intelligence: Strategic CSR in Light of Digital Washing and the Moral Human Argument. Sustainability, 16(3), 1232. https://doi.org/10.3390/su16031232

Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information Sciences, 479, 448-455. https://doi.org/10.1016/j.ins.2017.12.030

Firdaus, A., Anuar, N. B., Karim, A., & Razak, M. F. A. (2018). Discovering optimal features using static analysis and a genetic search based method for Android malware detection. Frontiers Of Information Technology

& Electronic Engineering/Frontiers Of Informaion Technology & Electronic Engineering, 19(6), 712-736. https://doi.org/10.1631/fitee.1601491

Firdaus, R., Xue, Y., Gang, L., & Ali, M. S. E. (2022). Artificial Intelligence and Human Psychology in Online Transaction Fraud. Frontiers In Psychology, 13. https://doi.org/10.3389/fpsyg.2022.947234

Francisco, M., & Linnér, B. (2023). AI and the governance of sustainable development. An idea analysis of the European Union, the United Nations, and the World Economic Forum. Environmental Science & Policy, 150, 103590. https://doi.org/10.1016/j.envsci.2023.103590

Fu, T. (2019). China's personal information protection in a data-driven economy: A privacy policy study of Alibaba, Baidu and Tencent. Global Media and Communication, 15(2), 195-213.

Fukas, P., Rebstadt, J., Menzel, L., & Thomas, O. (2022). Towards Explainable Artificial Intelligence in Financial Fraud Detection: Using Shapley Additive Explanations to Explore Feature Importance. En Lecture notes in computer science (pp. 109-126). https://doi.org/10.1007/978-3-031-07472-1_7

Gai, K., Qiu, M., Sun, X., & Zhao, H. (2017). Security and Privacy Issues: A Survey on FinTech. En Lecture notes in computer science (pp. 236-247). https://doi.org/10.1007/978-3-319-52015-5_24

Galbusera, F., & Cina, A. (2024). Image annotation and curation in radiology: an overview for machine learning practitioners. European Radiology Experimental, 8(1). https://doi.org/10.1186/s41747-023-00408-y

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. Energy Economics, 81, 709–727. https://doi.org/10.1016/j.eneco.2019.05.006

Giest, S. N., & Klievink, B. (2022). More than a digital system: how AI is changing the role of bureaucrats in different organizational contexts. Public Management Review, 26(2), 379-398. https://doi.org/10.1080/14719037.2022.2095001

Gomes, M. G., Da Silva, V. H. C., Pinto, L. F. R., Centoamore, P., Digiesi, S., Facchini, F., & De Oliveira Neto, G. C. (2020). Economic, Environmental and Social Gains of the Implementation of Artificial Intelligence at Dam Operations toward Industry 4.0 Principles. Sustainability, 12(9), 3604. https://doi.org/10.3390/su12093604Google (2025). Making AI helpful for everyone. Google IA. https://ai.google/

Guo, Z., Cho, J., Chen, I., Sengupta, S., Hong, M., & Mitra, T. (2021). Online Social Deception and Its Countermeasures: A Survey. IEEE Access, 9, 1770-1806. https://doi.org/10.1109/access.2020.3047337

Guresen, E., & Kayakutlu, G. (2011). Definition of artificial neural networks with comparison to other networks. Procedia Computer Science, 3, 426-433. https://doi.org/10.1016/j.procs.2010.12.071

Hamilton, S. A., Ambrosy, A. P., Parikh, R. V., Tan, T. C., Fitzpatrick, J. K., Avula, H. R., Sandhu, A. T., Ku, I. A., Go, A. S., Sax, D., & Bhatt, A. S. (2024). Applying natural language processing to identify emergency department and observation encounters for worsening heart failure. ESC Heart Failure. https://doi.org/10.1002/ehf2.14829

Hamon, R., Junklewitz, H., Garrido, J. S., & Sanchez, I. (2024). Three Challenges to Secure AI Systems in the Context of AI Regulations. IEEE Access, 12, 61022-61035. https://doi.org/10.1109/access.2024.3391021

Harbinja, E., Edwards, L., & McVey, M. (2023). Governing ghostbots. Computer Law And Security Report/Computer Law & Security Report, 48, 105791. https://doi.org/10.1016/j.clsr.2023.105791

Hasan, Z., Vaz, D., Athota, V. S., Désiré, S. S. M., & Pereira, V. (2023). Can artificial intelligence (AI) manage behavioural biases among financial planners? Journal Of Global Information Management, 31(2), 1-18. https://doi.org/10.4018/jgim.321728

Helfat, C. E., Kaul, A., Ketchen, D. J., Barney, J. B., Chatain, O., & Singh, H. (2023). Renewing the resource-based view: New contexts, new concepts, and new methods. Strategic Management Journal, 44(6), 1357-1390. https://doi.org/10.1002/smj.3500

Hofstede, G. (2001). Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations (2nd ed.). SAGE Publications

Horneber, D., & Laumer, S. (2023). Algorithmic accountability. Business & Information Systems Engineering, 65(6), 723-730. https://doi.org/10.1007/s12599-023-00817-8

Hu, Y., Kuang, W., Qin, Z., Li, K., Zhang, J., Gao, Y., Li, W., & Li, K. (2021). Artificial Intelligence Security: Threats and Countermeasures. ACM Computing Surveys, 55(1), 1-36. https://doi.org/10.1145/3487890

Huang, X., Kwiatkowska, M., Wang, S., & Wu, M. (2017). Safety Verification of Deep Neural Networks. En Lecture notes in computer science (pp. 3-29). https://doi.org/10.1007/978-3-319-63387-9_1

Huh, Y. K. (2022). Legal Implications of Financial Supervision with Artificial Intelligence. The Korean Journal of Securities Law, 23(1), 221-250. https://www.webofscience.com/wos/alldb/full-record/KJD:ART002839443

Iadanza, E., Benincasa, G., & Ventisette, I. (2022). Automatic Classification of Hospital Settings through Artificial Intelligence. Electronics, 11(11), 1697. https://doi.org/10.3390/electronics11111697

Igwegbe, C. A., Obi, C. C., Ohale, P. E., Ahmadi, S., Onukwuli, O. D., Nwabanne, J. T., & Białowiec, A. (2023). Modelling and optimisation of electrocoagulation/flocculation recovery of effluent from land-based aquaculture by artificial intelligence (AI) approaches. Environmental Science And Pollution Research International, 30(27), 70897-70917. https://doi.org/10.1007/s11356-023-27387-2

Innan, N., Khan, M. A., & Беннаи, M. (2023). Financial fraud detection: A comparative study of quantum machine learning models. International Journal Of Quantum Information, 22(02). https://doi.org/10.1142/s0219749923500442

Jamshidi, S., & Hashemi, M. R. (2012). An efficient data enrichment scheme for fraud detection using social network analysis. 2012 SIXTH INTERNATIONAL SYMPOSIUM ON TELECOMMUNICATIONS (IST) , pp.1082-1087. https://doi.org/10.1109/istel.2012.6483147

Jang, S. (2024). Recent Trends and Implications for Artificial Intelligence Regulation. Science, Technology and Law, 15 (1), 153-174. Doi: 10.34267/cbstl.2024.15.1.153.

Jesus, S., Belém, C., Balayan, V., Bento, J., Saleiro, P., Bizarro, P., & Gama, J. (2021). How can I choose an explainer? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp.805-815. https://doi.org/10.1145/3442188.3445941

Jiménez, M. (2025, 21 enero). Trump barre la era Biden con una avalancha de decretos contra la inmigración, la agenda verde y la diversidad. El País. https://elpais.com/internacional/2025-01-20/trump-barre-la-era-biden-con-una-avalancha-de-decretos-contra-la-inmigracion-la-agenda-verde-y-la-diversidad.html

Jin, Q., Lin, R., & Yang, F. (2020). E-WACGAN: Enhanced Generative Model of Signaling Data Based on WGAN-GP and ACGAN. IEEE Systems Journal, 14(3), 3289-3300. https://doi.org/10.1109/jsyst.2019.2935457

Jobin, A., & Ienca, M. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2

Johnson, M., Albizri, A., Harfouche, A., & Fosso-Wamba, S. (2022). Integrating human knowledge into artificial intelligence for complex and ill-structured problems: Informed artificial intelligence. International Journal Of Information Management, 64, 102479. https://doi.org/10.1016/j.ijinfomgt.2022.102479

Kamruzzaman, M. M. (2022). Impact of Social Media on Geopolitics and Economic Growth: Mitigating the Risks by Developing Artificial Intelligence and Cognitive Computing Tools. Computational Intelligence And Neuroscience, 2022, 1-12. https://doi.org/10.1155/2022/7988894

Kanika, Singla, J., Bashir, A. K., Nam, Y., Hasan, N. U., & Tariq, U. (2022). Handling Class Imbalance in Online Transaction Fraud Detection. Computers, Materials & Continua/Computers, Materials & Continua (Print), 70(2), 2861-2877. https://doi.org/10.32604/cmc.2022.019990

Kapadiya, K., Patel, U., Gupta, R., Alshehri, M. D., Tanwar, S., Sharma, G., & Bokoro, P. N. (2022). Blockchain and AI-Empowered Healthcare Insurance Fraud Detection: an Analysis, Architecture, and Future Prospects. IEEE Access, 10, 79606-79627. https://doi.org/10.1109/access.2022.3194569

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. Business Horizons, 63(1), 37-50. https://doi.org/10.1016/j.bushor.2019.09.003

Kesa, A., & Kerikmäe, T. (2020). Artificial Intelligence and the GDPR: Inevitable Nemeses? TalTech Journal Of European Studies/TalTech Journal Of European Studies., 10(3), 68-90. https://doi.org/10.1515/bjes-2020-0022

Khan, H. U., Malik, M. Z., Nazir, S., & Khan, F. (2023). Utilizing Bio Metric System for Enhancing Cyber Security in Banking Sector: A Systematic Analysis. IEEE Access, 11, 80181-80198. https://doi.org/10.1109/access.2023.3298824

Kim, Doo Jin (2019). Digital Economy and FinTech in the view of Consumer Protection. The Journal of Comparative Private Law, 26, (3), 285-330. https://www.webofscience.com/wos/alldb/full-record/KJD:ART002501742

Kim, S., & Chung, S. (2019). Explaining organizational responsiveness to emerging regulatory pressure: The case of illegal overwork in South Korea. Human Relations, 72(9), 1436-1461.

King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2019). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. Science And Engineering Ethics, 26(1), 89-120. https://doi.org/10.1007/s11948-018-00081-0

Kong, M., Li, R., Wang, J., Li, X., Jin, S., Xie, W., Hou, M., & Cao, C. (2024). CFTNet: a robust credit card fraud detection model enhanced by counterfactual data augmentation. Neural Computing & Applications, 36(15), 8607-8623. https://doi.org/10.1007/s00521-024-09546-9

Koukal, A., Gleue, C., & Breitner, M. (2014). Enhancing Literature Review Methods-Evaluation of a Literature Search Approach based on Latent Semantic Indexing.

Krambia-Kapardis, M., Christodoulou, C., & Agathocleous, M. (2010). Neural networks: the panacea in fraud detection? Managerial Auditing Journal, 25(7), 659–678. https://doi.org/10.1108/02686901011061342

Kumar, A., Bhattacharyya, S. S., & Krishnamoorthy, B. (2023). Automation-augmentation paradox in organizational artificial intelligence technology deployment capabilities; an empirical investigation for achieving simultaneous economic and social benefits. Journal Of Enterprise Information Management, 36(6), 1556-1582. https://doi.org/10.1108/jeim-09-2022-0307

Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S. K. (2022). Past, present, and future of sustainable finance: insights from big data analytics through machine learning of scholarly research. Annals Of Operation Research/Annals Of Operations Research. https://doi.org/10.1007/s10479-021-04410-8

Kuzniacki, B., Almada, M., Tyliński, K., & Górski, Ł. (2022). Requirements for Tax XAI Under Constitutional Principles and Human Rights. En Lecture notes in computer science (pp. 221-238). https://doi.org/10.1007/978-3-031-15565-9_14

Laghrabli, S., Benabbou, L., & Berrado, A. (2015, October). A new methodology for literature review analysis using association rules mining. In 2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA) (pp. 1-6). IEEE.

Lai, C. Y., Li, Y., & Lin, L. F. (2017). A social referral appraising mechanism for the e-marketplace. Information & Management, 54(3), 269-280. https://doi.org/10.1016/j.im.2016.07.001

Leben, D. (2023). Explainable AI as evidence of fair decisions. Frontiers in Psychology, 14, 1069426.

Lee, K. F., Triolo, P., & Kania, E. (2018). AI Superpowers: China, Silicon Valley, and the New World Order. Houghton Mifflin Harcourt.

Lee, K. Y., Kwon, H. Y., & Lim, J. I. (2018). Legal Consideration on the Use of Artificial Intelligence Technology and Self-regulation in Financial Sector: Focused on Robo-Advisors. En Lecture notes in computer science (pp. 323-335). https://doi.org/10.1007/978-3-319-93563-8_27

Li, J. (2022). E-Commerce Fraud Detection Model by Computer Artificial Intelligence Data Mining. Computational Intelligence And Neuroscience, 2022, 1-9. https://doi.org/10.1155/2022/8783783

Lin, P. (2016). Why ethics matters for autonomous cars. In L. A. P. Lin, K. Abney, & G. A. Bekey (Eds.), Robot ethics: The ethical and social implications of robotics (pp. 69-85). MIT Press.

Lokanan, M. (2022). The determinants of investment fraud: A machine learning and artificial intelligence approach. Frontiers In Big Data, 5. https://doi.org/10.3389/fdata.2022.961039

London, A. J. (2019). Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. The Hastings Center Report, 49(1), 15-21. https://doi.org/10.1002/hast.973

López-Úbeda, P., Martín-Noguerol, T., & Luna, A. (2023). Radiology, explicability and AI: closing the gap. European Radiology, 33(12), 9466-9468. https://doi.org/10.1007/s00330-023-09902-8

Lorkowski, J., Kolaszyńska, O., & Pokorski, M. (2021). Artificial Intelligence and Precision Medicine: A Perspective. En Advances in experimental medicine and biology (pp. 1-11). https://doi.org/10.1007/5584_2021_652

Maashi, M., Alabduallah, B. I., & Kouki, F. (2023b). Sustainable Financial Fraud Detection Using Garra Rufa Fish Optimization Algorithm with Ensemble Deep Learning. Sustainability, 15(18), 13301. https://doi.org/10.3390/su151813301

Majeed, Y., Fu, L., & He, L. (2024). Editorial: Artificial intelligence-of-things (AIoT) in precision agriculture. Frontiers In Plant Science, 15. https://doi.org/10.3389/fpls.2024.1369791

Mangrulkar, S. (1990). Artificial neural systems. ISA Transactions, 29(1), 5-7. https://doi.org/10.1016/0019-0578(90)90024-f

Mannuru, N. R., Shahriar, S., Teel, Z. A., Wang, T., Lund, B. D., Tijani, S., Pohboon, C. O., Agbaji, D., Alhassan, J., Galley, J., Kousari, R., Ogbadu-Oladapo, L., Saurav, S. K., Srivastava, A., Tummuru, S. P., Uppala, S., & Vaidya, P. (2023). Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development. Information Development. https://doi.org/10.1177/02666669231200628

Martin, K. (2018). Ethical Implications and Accountability of Algorithms. Journal Of Business Ethics, 160(4), 835-850. https://doi.org/10.1007/s10551-018-3921-3

Talaei, J., Yang, A., Takishova, T., & Masialeti, M. (2024). How Does Cost Leadership Strategy Suppress the Performance Benefits of Explainability of AI Applications in Organizations? Journal Of Global Information Management, 32(1), 1-23. https://doi.org/10.4018/jgim.354062

Martinez, A. R. (2010). Natural language processing. Wiley Interdisciplinary Reviews Computational Statistics, 2(3), 352-357. https://doi.org/10.1002/wics.76

Martínez, R. (2016). Fraud Detection Using Deep Learning. https://cybercamp.es/cybercamp2016/sites/default/files/contenidos/material/cybercamp2016-fraud_detection_using_deep_learning-ruben_martinez.pdf

Medhat, M. (2012). Artificial intelligence methods applied for quantitative analysis of natural radioactive sources. Annals Of Nuclear Energy, 45, 73-79. https://doi.org/10.1016/j.anucene.2012.02.013

Mersha, M., Lam, K., Wood, J., AlShami, A., & Kalita, J. (2024). Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. Neurocomputing, 128111. https://doi.org/10.1016/j.neucom.2024.128111

Mhlanga, D. (2020). Industry 4.0 in Finance: The Impact of Artificial Intelligence (AI) on Digital Financial Inclusion. International Journal Of Financial Studies, 8(3), 45. https://doi.org/10.3390/ijfs8030045

Mill, E. R., Garn, W., Ryman-Tubb, N. F., & Turner, C. (2023). Opportunities in real time fraud detection: an explainable artificial intelligence (XAI) Research Agenda. International Journal of Advanced Computer Science and Applications, 14(5), 1172-1186. https://www.webofscience.com/wos/alldb/full-record/WOS:001015077500001

Minkkinen, M., Niukkanen, A., & Mäntymäki, M. (2022). What about investors? ESG analyses as tools for ethics-based AI auditing. AI & SOCIETY. https://doi.org/10.1007/s00146-022-01415-0.

Mishra, S. (2023). Exploring the Impact of AI-Based Cyber Security Financial Sector Management. Applied Sciences, 13(10), 5875. https://doi.org/10.3390/app13105875

MIT Technology Review (2024). Vuelta al mundo por las regulaciones de la IA en 2024. Opinno. https://www.technologyreview.es/s/16069/vuelta-al-mundo-por-las-regulaciones-de-la-ia-en-2024

Moraes, T. (2024). Ethical AI Regulatory Sandboxes: Insights from cyberspace regulation and Internet governance. 2nd International Symposium On Trustworthy Autonomous Systems, 10, 1-10. https://doi.org/10.1145/3686038.3686049

Mytnyk, B., Tkachyk, O., Shakhovska, N., Fedushko, S., & Syerov, Y. (2023). Application of Artificial Intelligence for Fraudulent Banking Operations Recognition. Big Data And Cognitive Computing, 7(2), 93. https://doi.org/10.3390/bdcc7020093

Nannini, L., Alonso-Moral, J. M., Catala, A., Lama, M., & Barro, S. (2024). Operationalizing Explainable Artificial Intelligence in the European Union Regulatory Ecosystem. IEEE Intelligent Systems, 39(4), 37-48. https://doi.org/10.1109/mis.2024.3383155

Nayak, S., & Chandiramani, J. (2022). A crisis that changed the banking scenario in India: exploring the role of ethics in business. Asian Journal Of Business Ethics, 11(S1), 7-32. https://doi.org/10.1007/s13520-022-00151-4

Newman, J., & Mintrom, M. (2023). Mapping the discourse on evidence-based policy, artificial intelligence, and the ethical practice of policy analysis. Journal Of European Public Policy, 30(9), 1839-1859. https://doi.org/10.1080/13501763.2023.2193223

Nicodeme, C. (2020). Build confidence and acceptance of AI-based decision support systems - Explainable and liable AI. 13th IEEE International Conference On Human System Interaction (IEEE HSI). https://doi.org/10.1109/hsi49210.2020.9142668

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. International Journal of Information Management, 53, 102104. https://doi.org/10.1016/j.ijinfomgt.2020.102104

Novelli, C. (2022). Legal personhood for the integration of AI systems in the social context: a study hypothesis. AI & Society, 38(4), 1347-1359. https://doi.org/10.1007/s00146-021-01384-w

Pagliari, M., Chambon, V., & Berberian, B. (2022). What is new with Artificial Intelligence? Human–agent interactions through the lens of social agency. Frontiers In Psychology, 13. https://doi.org/10.3389/fpsyg.2022.954444

Pai, V., & Chandra, S. (2022). Exploring Factors Influencing Organizational Adoption of Artificial Intelligence (AI) in Corporate Social Responsibility (CSR) Initiatives. Pacific Asia Journal Of The Association For Information Systems, 14, 82-115. https://doi.org/10.17705/1pais.14504

Piccininni, M. (2022). Counterfactual fairness: The case study of a food delivery platform's reputational-ranking algorithm. Frontiers in Psychology, 13, 1015100.

Pinheiro, A. F., Santos, W. B., & De Lima Neto, F. B. (2023). Intelligent Framework to Support Technology and Business Specialists in the Public Sector. IEEE Access, 11, 15655-15679. https://doi.org/10.1109/access.2023.3243195

Politi, V. (2024). Who ought to look towards the horizon? A qualitative study on the collective social responsibility of scientific research. European Journal For Philosophy Of Science, 14(2). https://doi.org/10.1007/s13194-024-00580-x

Qiu, S., & Luo, Y. (2024). How to detect and forecast corporate fraud by media reports? An approach using machine learning and qualitative comparative analysis. Journal Of Forecasting, 43(1), 58-80. https://doi.org/10.1002/for.3022

Ramdhani, M. A., & Ramdhani, A. (2014). Verification of research logical framework based on literature review. International Journal of Basic and Applied Science, 3(2), 1-9.

Ranard, B. L., Park, S., Jia, Y., Zhang, Y., Alwan, F., Celi, L. A., & Lusczek, E. R. (2024). Minimizing bias when using artificial intelligence in critical care medicine. Journal Of Critical Care, 82, 154796. https://doi.org/10.1016/j.jcrc.2024.154796

Razaque, A., Alotaibi, B., Alotaibi, M., Hussain, S., Alotaibi, A., & Jotsov, V. (2022). Clickbait detection using deep recurrent neural network. Applied Sciences, 12(1), 504. https://doi.org/10.3390/app12010504

Ridzuan, N. N., Masri, M., Anshari, M., Fitriyani, N. L., & Syafrudin, M. (2024). AI in the Financial Sector: The Line between Innovation, Regulation and Ethical Responsibility. Information, 15(8), 432. https://doi.org/10.3390/info15080432

Rouhani, B. D., Samragh, M., Javaheripi, M., Javidi, T., & Koushanfar, F. (2018). DeepFense. 2018 IEEE/ACM INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN (ICCAD) DIGEST OF TECHNICAL PAPERS. https://doi.org/10.1145/3240765.3240791

Russell, S. J., & Norvig, P. (1995). Artificial intelligence: a modern approach. Choice Reviews Online, 33(03), 33-1577. https://doi.org/10.5860/choice.33-1577

Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. Engineering Applications Of Artificial Intelligence, 76, 130-157. https://doi.org/10.1016/j.engappai.2018.07.008

Sabharwal, R., Miah, S. J., Wamba, S. F., & Cook, P. (2024). Extending application of explainable artificial intelligence for managers in financial organizations. Annals Of Operation Research/Annals Of Operations Research. https://doi.org/10.1007/s10479-024-05825-9

Sakyoud, Z., Aaroud, A., & Akodadi, K. (2023). Optimization of purchasing business process in Moroccan public universities based on COBIT and artificial intelligence techniques. Kybernetes, 53(5), 1607-1635. https://doi.org/10.1108/k-02-2022-0167

Schneider, D., & Weber, K. (2024). AI for decision support: What are possible futures, social impacts, regulatory options, ethical conundrums and agency constellations? TATuP - Zeitschrift Für Technikfolgenabschätzung In Theorie Und Praxis, 33(1), 8-54. https://doi.org/10.14512/tatup.33.1.08

Searle, J. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3 (3), 417-424. doi:10.1017/S0140525X00005756

Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., Ravi, V., & Peters, A. (2020). A review of deep learning with special emphasis on architectures, applications and recent trends. Knowledge-based Systems, 194, 105596. https://doi.org/10.1016/j.knosys.2020.105596

Shneiderman, B. (2020). Bridging the Gap Between Ethics and Practice. ACM Transactions On Interactive Intelligent Systems, 10(4), 1-31. https://doi.org/10.1145/3419764

Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023). Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models. In 2023 International Conference on Computer and Applications (ICCA) (pp. 1-7). IEEE

Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023). Human-centricity in AI governance: A systemic approach. Frontiers in Artificial Intelligence, 6, 976887. https://doi.org/10.3389/frai.2023.976887

Simos, T. E., Katsikis, V. N., & Mourtas, S. D. (2022). A multi-input with multi-function activated weights and structure determination neuronet for classification problems and applications in firm fraud and loan approval. Applied Soft Computing, 127, 109351. https://doi.org/10.1016/j.asoc.2022.109351

Sina, A. (2023). Open AI and its Impact on Fraud Detection in Financial Industry. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 263-281. https://doi.org/10.60087/jklst.vol2.n3.p281

Sison, A., Ferrero, I., Ruiz, P. G., & Kim, T. W. (2023). Editorial: Artificial intelligence (AI) ethics in business. Frontiers In Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1258721

Sood, P., Sharma, C., Nijjer, S., & Sakhuja, S. (2023). Review the role of artificial intelligence in detecting and preventing financial fraud using natural language processing. International Journal Of System Assurance Engineering And Management, 14(6), 2120-2135. https://doi.org/10.1007/s13198-023-02043-7

Souza, M. A., Gouveia, H. T. V., Ferreira, A. A., De Lima Neta, R. M., Neto, O. N., Da Silva Lira, M. M., Torres, G. L., & De Aquino, R. R. B. (2024). Detection of Non-Technical Losses on a Smart Distribution Grid Based on Artificial Intelligence Models. Energies, 17(7), 1729. https://doi.org/10.3390/en17071729

Stanford University (2024). Stanford HAI and Stanford Robotics Center Launch New Partnership (2024, 29 octubre). Stanford HAI. https://hai.stanford.edu/news/stanford-hai-and-stanford-robotics-center-launch-new-partnership

Strelcenia, E., & Prakoonwit, S. (2023). A Survey on GAN Techniques for Data Augmentation to Address the Imbalanced Data Issues in Credit Card Fraud Detection. Machine Learning And Knowledge Extraction, 5(1), 304-329. https://doi.org/10.3390/make5010019

Taher, S. S., Ameen, S. Y., & Ahmed, J. A. (2024). Advanced Fraud Detection in Blockchain Transactions: An Ensemble Learning and Explainable AI Approach. Engineering, Technology And Applied Science Research/Engineering, Technology And Applied Science Research, 14(1), 12822-12830. https://doi.org/10.48084/etasr.6641

Tan, Z., Liu, B., & Wu, A. (2022). Artificial Intelligence and Feature Identification Based Global Perception of Power Consumer: Definition, Structure, and Applications. Frontiers In Energy Research, 10. https://doi.org/10.3389/fenrg.2022.907027

Tiwary, N., Noah, S. A. M., Fauzi, F., & Yee, T. S. (2024). Max Explainability Score–A quantitative metric for explainability evaluation in knowledge graph-based recommendations. Computers & Electrical Engineering, 116, 109190. https://doi.org/10.1016/j.compeleceng.2024.109190

Truby, J. (2020). Governing Artificial Intelligence to benefit the UN Sustainable Development Goals. Sustainable Development, 28(4), 946-959. https://doi.org/10.1002/sd.2048

Van Bekkum, M., & Borgesius, F. Z. (2021). Digital welfare fraud detection and the Dutch SyRI judgment. European Journal Of Social Security, 23(4), 323-340. https://doi.org/10.1177/13882627211031257

Van Den Berg, M., Gerlings, J., & Kim, J. (2024). Empirical Research on Ensuring Ethical AI in Fraud Detection of Insurance Claims: A Field Study of Dutch Insurers. En Communications in computer and information science (pp. 106-114). https://doi.org/10.1007/978-3-031-50485-3_9

Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. International Journal of Information Management Data Insights, 1(1), 100002. https://doi.org/10.1016/j.jjimei.2020.100002

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, 11(1), 233. https://doi.org/10.1038/s41467-019-14108-y

Walters, R., Trakman, L., & Zeller, B. (2019). Jurisdictional [Comparative] Differences. En DATA PROTECTION LAW: A COMPARATIVE ANALYSIS OF ASIA-PACIFIC AND EUROPEAN APPROACHES (pp. 265-290). https://doi.org/10.1007/978-981-13-8110-2_11

Wang, G., Frederick, R., Duan, J., Wong, W., Rupar, V., Li, W., & Bai, Q. (2024). Detecting misinformation through Framing Theory: the Frame Element-based Model. arXiv preprint arXiv:2402.15525.

Weber, M. T., Schaaf, J., Storf, H., Wagner, T. O., Berger, A., & Noll, R. (2024). Editing Physicians' Responses Using GPT-4 for Academic Research. Studies In Health Technology And Informatics. https://doi.org/10.3233/shti240019

Wilson, C., & van der Velden, M. (2022). Sustainable AI: An integrated model to guide public sector decision-making. Technology in Society, 68, 101926. https://doi.org/10.1016/j.techsoc.2022.101926

Win, K. T., & Beydoun, G. (2020). Preface: Towards the Next Generation of Information Systems: Enhancing Traceability and Transparency. AJIS. Australasian Journal Of Information Systems/AJIS. Australian Journal Of Information Systems/Australian Journal Of Information Systems, 24. https://doi.org/10.3127/ajis.v24i0.2823

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. Journal Of Service Management, 29(5), 907-931. https://doi.org/10.1108/josm-04-2018-0119

Xie, J., Dmour, A. A., & Lakys, Y. (2022). Application of Nonlinear Fractional Differential Equations in Computer Artificial Intelligence Algorithms. Applied Mathematics And Nonlinear Sciences, 8(1), 1145-1154. https://doi.org/10.2478/amns.2022.2.0101

Xiong, T., Ma, Z., Li, Z., & Dai, J. (2021). The analysis of influence mechanism for internet financial fraud identification and user behavior based on machine learning approaches. International Journal Of Systems Assurance Engineering And Management, 13(S3), 996-1007. https://doi.org/10.1007/s13198-021-01181-0

Xu, D., Fan, S., & Kankanhalli, M. (2023). Combating misinformation in the era of generative AI models. In Proceedings of the 31st ACM International Conference on Multimedia (pp. 9291-9298)

Yang, J., Tang, Z., Guan, Z., Hua, W., Wei, M., Wang, C., & Gu, C. (2021). Automatic Feature Engineering-Based Optimization Method for Car Loan Fraud Detection. Discrete Dynamics In Nature And Society, 2021, 1-10. https://doi.org/10.1155/2021/6077540

Yang, X., Li, H., Ni, L., & Li, T. (2021). Application of Artificial Intelligence in Precision Marketing. Journal Of Organizational And End User Computing, 33(4), 209-219. https://doi.org/10.4018/joeuc.20210701.oa10

Yang, X., Zhang, C., Sun, Y., Pang, K., Li, J., Wa, S., & Lv, C. (2023). FinChain-BERT: A High-Accuracy Automatic Fraud Detection Model Based on NLP Methods for Financial Scenarios. Information, 14(9), 499. https://doi.org/10.3390/info14090499

Yaraziz, M. S., Jalili, A., Gheisari, M., & Liu, Y. (2022). Recent trends towards privacy-preservation in Internet of Things, its challenges and future directions. IET Circuits Devices & Systems, 17(2), 53-61. https://doi.org/10.1049/cds2.12138

Yerima, S. Y., & Bashar, A. (2022). Semi-supervised novelty detection with one class SVM for SMS spam detection. 2022 29TH INTERNATIONAL CONFERENCE ON SYSTEMS, SIGNALS AND IMAGE PROCESSING (IWSSIP). https://doi.org/10.1109/iwssip55020.2022.9854496

Zaimi, R., Hafidi, M., & Lamia, M. (2023). A deep learning approach to detect phishing websites using CNN for privacy protection. Intelligent Decision Technologies, 17(3), 713-728. https://doi.org/10.3233/idt-220307

Zajko, M. (2023). Automated government benefits and welfare surveillance. Surveillance & society, 21(3), 246-258. https://www.webofscience.com/wos/alldb/full-record/WOS:001078822300002

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J. C., Sellitto, M., Shoham, Y., Clark, J., & Perrault, R. (2022). The AI Index 2022 Annual Report. AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University

Zhang, T., Qin, Y., & Li, Q. (2021). Trusted Artificial Intelligence: Technique Requirements and Best Practices. 2021 INTERNATIONAL CONFERENCE ON CYBERWORLDS (CW 2021) , pp.303-306. https://doi.org/10.1109/cw52790.2021.00058

Zhang, W., Zhang, S., Chen, J., & Huang, H. (2023). CCFD-GAN: Credit Card Fraud Detection Based on Generative Adversarial Networks Enhanced by Penalty Mechanism. 2023 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, IJCNN. https://doi.org/10.1109/ijcnn54540.2023.10191810

Zhao, J., & Fariñas, B. G. (2022). Artificial Intelligence and Sustainable Decisions. European Business Organization Law Review, 24(1), 1-39. https://doi.org/10.1007/s40804-022-00262-2

Zheng, X., Gildea, E., Chai, S., Zhang, T., & Wang, S. (2023). Data Science in Finance: Challenges and Opportunities. AI, 5(1), 55-71. https://doi.org/10.3390/ai5010004

Zheng, X., Gildea, E., Chai, S., Zhang, T., & Wang, S. (2023). Data science in finance: Challenges and opportunities. AI, 5(1), 55-71.

Zheng, X., Li, J., Lu, M., & Wang, F. (2024). New Paradigm for Economic and Financial Research With Generative AI: Impact and Perspective. IEEE Transactions On Computational Social Systems, 1-11. https://doi.org/10.1109/tcss.2023.3334306

Zhou, Y., Li, H., Xiao, Z., & Qiu, J. (2023). A user-centered explainable artificial intelligence approach for financial fraud detection. Finance Research Letters, 58, 104309. https://doi.org/10.1016/j.frl.2023.104309

Zhu, T., & Yu, P. S. (2019). Applying Differential Privacy Mechanism in Artificial Intelligence. 39th IEEE International Conference On Distributed Computing Systems (ICDCS). https://doi.org/10.1109/icdcs.2019.00159

Zhu, T., Ye, D., Wang, W., Zhou, W., & Yu, P. (2021). More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence. IEEE Transactions On Knowledge And Data Engineering, 1. https://doi.org/10.1109/tkde.2020.3014246

Zhu, X., Ao, X., Qin, Z., Chang, Y., Liu, Y., He, Q., & Li, J. (2021). Intelligent financial fraud detection practices in post-pandemic era. The Innovation, 2(4), 100176. https://doi.org/10.1016/j.xinn.2021.100176

Züger, T., & Asghari, H. (2022). AI for the public. How public interest theory shifts the discourse on AI. AI & Society, 38(2), 815-828. https://doi.org/10.1007/s00146-022-01480-5