

PUBLISHED RESEARCH ARTICLES

On the Transposition of FAIR Data Principles to Financial Services: An Adapted FAAIR Guideline

Emer Owens¹, Barry Sheehan¹, Martin Mullins¹, Martin Cunneen¹, Ben O'Connell²

¹ University of Limerick, ² Unaffiliated

Keywords: research reproducibility, fair data principles, financial services, artificial intelligence, systematic review

<https://doi.org/10.52399/001c.138720>

Accounting, Finance & Governance Review

Vol. 34, 2025

Adopting a critical view on the governance and policymaking of artificial intelligence (AI), commercial entities in the financial services industry retain research knowledge and power, leading to policymaking challenges. This research critically analyses the FAIR (Findable, Accessible, Interoperable, Reusable) discourse on financial services research. Research reproducibility is assessed in peer-reviewed research articles that apply AI methods in finance, taxation, and insurance fields published between 2000 and 2023. A dataset of 126 articles (66 finance, 40 insurance, 20 tax) is analysed, measuring each article's application of the FAIR Data Principles. Findings highlight that 50% of analysed articles comply with the FAIR Data Principles, indicating a reproducibility crisis in AI-applied financial services research.

1. Introduction

The financial services industry is positioned to benefit substantially from artificial intelligence (AI) applications. The ever-growing interest of regulators, consumers and financial service providers in AI applications has created an influx of research discussing the benefits, threats, and opportunities of AI's use; however, the trustworthiness, bias and auditability of AI models remain as barriers to implementation (Shawat et al., 2023). Effective research and governance of AI applications are necessary to ensure the continued successful application of AI in the financial services industry. The Trustworthy AI framework (European Commission, 2019) highlights the complex technical and governance requirements relating to the safe use of AI in business, with research in these fields of prime importance in its successful and trustworthy implementation. The FAIR (Findable, Accessible, Interoperable, Reusable) Data Principles aim to support effective, transparent and reproducible research, which is of prime importance in emerging and/or novel research fields, such as AI applications in financial services.

The FAIR Data Principles aid the quest for system-wide research transparency and accessibility; however, this paper highlights a lack of adherence to the FAIR Principles in research on AI applications in financial services. Through an analysis of AI-applied literature published between 2000 and 2023 in finance, taxation and insurance, we find the pervasive issue of

opaque and irreproducible research in this space. Findings note that a key reason for this is the ineffectiveness of the FAIR Principles' applicability to the complex personal and sensitive nature of data employed in this field.

This research paper considers the efficacy of the FAIR Data Principles for the governance of research transparency and accessibility in AI-applied financial services research. The Dutch Tech Centre for the Life Sciences developed FAIR to circumvent similar reproducibility issues regarding research results (Wilkinson et al., 2016) and it has been extended to further research fields, including nanotechnology (Furxhi et al., 2021; Jeliaskova et al., 2021), medicine (Meyer et al., 2021), disaster management (Mazimwe et al., 2021) and research software (Lamprecht et al., 2020). To support the financial services industry in accessing FAIR's benefits, we propose adapting FAIR by making it more context-specific to the unique industry requirements that exist in this domain. The additional component concerns the introduction of anonymised data (FAAIR) to address the specific operational risks of data collected and employed in financial services. There are significant challenges concerning open data sources in the financial services industry, including sensitive personal information and consumers' trust in their financial services provider. A lack of research data and results' availability hinders future research expansion, with "blocked" data sources excluding international development organisations (IDOs) and non-governmental organisations (NGOs) from key decision-making (Rotolo et al., 2022).

Our research issues a call to action in the financial services research domain to develop suitable data principles to protect the integrity of our discipline and shift the power of research knowledge to regulators. This paper extends the FAIR Data Principles to include an Anonymised data guideline, creating the FAAIR Data Principles. The proposed FAAIR Data framework allows financial services research to be transparent and reproducible by researchers and societal bodies.

The importance of AI applications in financial services is noted in the EU with the introduction of the AI Act (European Parliament, 2024), which identifies financial services provision as an area which "deserves special consideration" due to its potential impact on an individual's standard of living (Smuha, 2025). This aligns with the complex nature of financial services data used for AI applications requiring effective and timely governance, with academic research providing the foundation for many evidence-based policymaking decisions (Head, 2016). However, research funding is often localised within commercial entities, with the research data used and associated results then protected by the organisation.

We draw on Foucault's power-knowledge theory to explain the problems this research protection creates (Foucault, 2020). Certainly, we highlight the influence financial services organisations can wield over regulators through their ability to control flows of knowledge and information. Foucault's thinking around the power-knowledge nexus supports the intrinsic link between power and knowledge. Thus, we argue that organisations with access

to research funding, data, and research results hold power in the financial services sphere (Adolf & Stehr, 2016). Regulators are hindered in their ability to maintain control of AI's application in the financial services industry, as they lack access to siloed research outputs on the subject. Additionally, our research supplements capture theory literature by highlighting the role played by knowledge control in policymaking dynamics (King & Hayes, 2018).

This systematic review critically assesses the reviewed articles based on defined criteria (FAIR) for reproducible datasets, highlighting myriad irreproducible studies in the financial services literature set. Research surrounding the use of FAIR Data in medicine has paved the way for reviews that analyse an article's adherence to the Principles. Meyer et al. (2021) reviewed 152 articles surrounding veterinary epidemiological research published between 2017 and 2020, finding that none of the datasets met the complete requirements set by the FAIR Principles. Similarly, Bahlo and Dahlhaus (2021) reviewed livestock farming datasets in Australia, where low levels of adherence to certain aspects of the FAIR Principles were found.

The literature assessing adherence to the FAIR Principles extends into further inter- and transdisciplinary fields. Mazimwe et al. (2021) is one such paper providing a review of disaster management articles to highlight the need for reusing knowledge discovered in previous disaster management phases. Similar concerns are cited in the following research areas: research software (Katz et al., 2021; Lamprecht et al., 2020), bioscience (Smith & Sandbrink, 2022; Suver et al., 2020), computational science (Wolf et al., 2021), nanotechnology (Furxhi et al., 2021), and environmental science (Harjes et al., 2020). To rectify the reproducibility issue in financial services research, an adapted FAIR Guideline to account for anonymised personal sensitive data for research purposes is proposed to aid the expanding cross-disciplinary network of AI-based financial services research. Additionally, a collation of relevant academic and professional literature on AI applications in financial services will be helpful for regulators', financial services practitioners', IDOs', and NGOs' informed decision-making surrounding ethical considerations of AI regulation in response to social injustices, namely financial exclusion.

This paper presents the first systematic review and critical analysis of AI-applied finance, taxation and insurance articles' adherence to the FAIR Principles. This research is significant as it highlights the current state of AI applications in the financial services industry and the lack of transparency and reproducibility in the literature. This lack of transparency acts as a roadblock to the further application and development of AI while ignoring the research field's importance in civil society's decision-making. A strong regulatory regime is necessary to aid the further deployment of AI's use in financial services. The development of an adapted FAAIR framework will promote evidence-based policymaking by increasing research data and results sharing between industry, academia, and policymakers (Mulgan, 2005).

1.1. Related Literature

This paper has four main components: the FAIR Data Principles, AI applications in finance, AI applications in taxation, and AI applications in insurance. This section of the paper analyses the existing literature related to each of these components, thus highlighting the gap in the literature surrounding the absence of FAIR Data evaluation within AI-based financial services articles and the ensuing issues for policymakers. Transparency in AI applications is twofold – there are “outward” forms of transparency, which are most often discussed in the literature, and “functional” transparency, which is the subject of discussion in this paper (Walmsley, 2021). Functional transparency includes, yet is not limited to, the datasets used in both the training of AI models and the understandability of the AI models’ results, having contextualised such results with the knowledge of the inner workings of the models. The use of quality datasets for applied AI research is paramount, as the AI models rely heavily on their data inputs for training and testing purposes. Without sound data management and data quality checks in applied AI financial services research, the literature set may suffer the consequences of biased or inaccurate research results (Aldoseri et al., 2023). Additionally, for successful replication and validation of applied AI financial services research, the completeness, integrity, relevance and timeliness of analysed datasets should be verifiable by other researchers (Batista & Monard, 2018; H. Chen et al., 2012; Halevy et al., 2016; Karkouch et al., 2016).

Open data refers to the provision of research data publicly, which plays a crucial role in transparency ambitions for governments and research bodies (Splitter et al., 2023). The decision to decide what data eventually becomes available is often left to data providers. However, a call for open science necessitates open data repositories. Open science is the process of involving the public in science (Splitter et al., 2023). This facilitates inclusive engagement by the public, in turn empowering citizens in evidence-based decision-making (Janssen et al., 2012; Yu & Robinson, 2011; Zhu et al., 2019). Underpinning open science are three foundational pillars: accessibility, transparency, and inclusivity (Beck et al., 2020; Maedche et al., 2024). The FAIR Principles promote open data as a precursor to open science by focusing on data sharing, ensuring access and the reusability capacity of research data (Jati et al., 2022).

FAIR Data Principles

Appropriate data management is essential for research validation throughout each stage of a research project and is a step researchers understand they ought to engage in but rarely want to (Kowolenko & Vouk, 2018). Reasons abound for the necessity of reproducible research findings, including the opportunity to validate research results and encourage the reusability of data and research methodologies to further contribute to a given research space (Peng & Hicks, 2021). Issues with data management arise as metadata capturing is not implemented in everyday scientific research

practices. This is partly owing to a lack of data management training and the general understanding that data management occurs at the end of a project following data publication (Furxhi et al., 2021; Papadiamantis et al., 2020).

The establishment of the FAIR Data Principles in 2014 (published in 2016¹) heralded the beginning of inter- and cross-disciplinary research transparency by establishing a minimal set of rules for research data stewardship in the life sciences domain (Boeckhout et al., 2018). Structured on four main pillars – Findable, Accessible, Interoperable, and Reusable – these Principles present an inter-sectional attempt to ameliorate the access, use and reuse of digital resources by humans and machines (Wilkinson et al., 2016). These Principles intend to give a minimal set of community-agreed guidelines to follow, such that humans and machines can carry out the four facets (FAIR) on the used dataset relatively easily (Dunning et al., 2017). Wise et al. (2019) emphasise the importance of research data which complies with the FAIR guidelines: Findable, Accessible, Interoperable and Reusable data enable digital transformation and societal progress (Romanos et al., 2019). A significant note on the Principles is that they precede the implementation choices of (AI) models and, therefore, only provide a set of recommendations for data researchers and publishers to assist in evaluating the data being used (Wilkinson et al., 2016). Although a step in the right direction for research transparency, there is a lack of actions in implementing FAIR Data practices in many research areas, with the risk of incompatible implementations likely to cause future issues for researchers (Jacobsen et al., 2020).

Introducing more prevalent use of FAIR Data has been seen most commonly amid medical literature in the hopes of solving the issue of data that are not interpretable by those who are not an expert on the dataset (Kalendralis et al., 2021). Wise et al. (2019) point towards the current lack of technology and standards in the biopharmaceutical industry to support the implementation of the FAIR Principles. The paper suggests that the industry must move from an inward focus to a more outward-looking one that fully supports data-sharing to progress quickly. Similarly, Alharbi et al. (2021) highlight the need to build a technical infrastructure within the biopharmaceutical industry that supports large-scale FAIR implementation. In contrast, Holub et al. (2018) proposed an extended version of the FAIR Principles for use within medical research, suggesting that the current guidelines are not practical enough for that industry.

Both governments and researchers are exploring alternative methods of data governance to improve research transfer and data sharing in industries particularly impacted by AI, including finance, insurance, and medicine (He et al., 2022). The European Union has provided substantial investment in data-sharing spaces, aiming to develop the FAIR Principles into standardised criteria and indicating which facets of the Principles can become official

¹ Consulted and first published under FORCE11 in 2016.

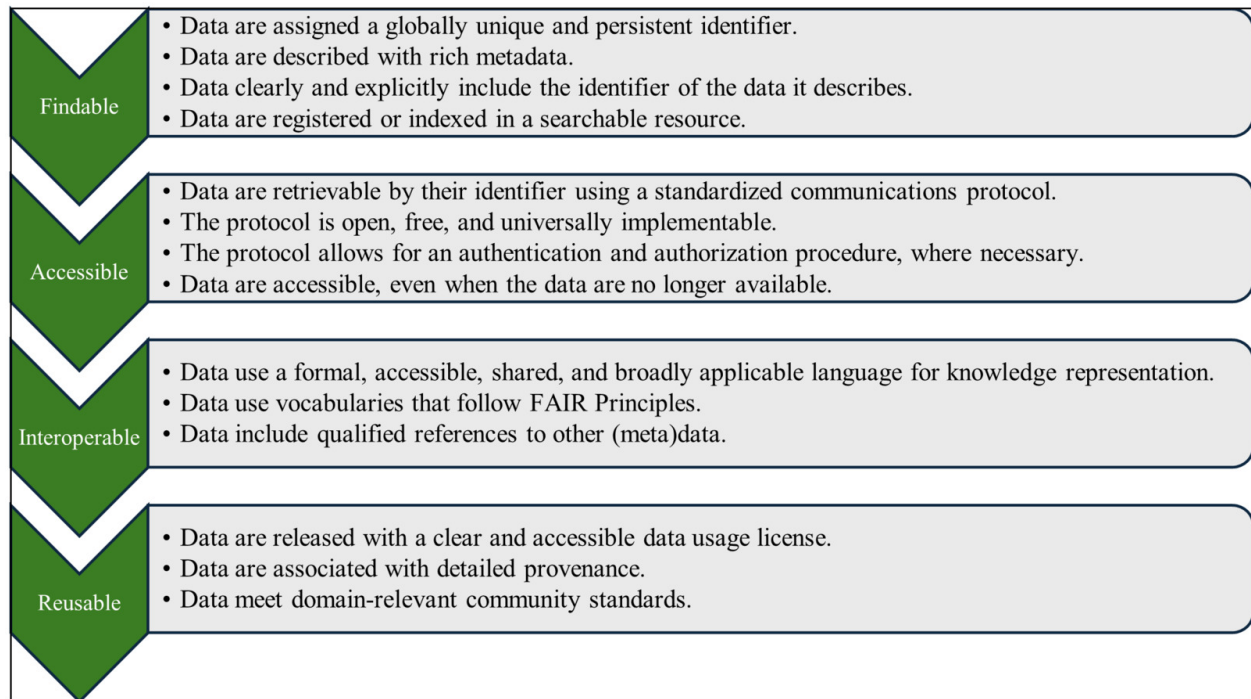


Figure 1. Wilkinson et al. (2016)'s Official FAIR Data Framework

^a "Data" refers to any case where the guidelines should be applied to metadata and data – additional source: David et al. (2018).

research policy (Groth & Dumontier, 2020). Furxhi et al. (2021) provide important work on the "FAIRification" process of research datasets and methods in nanotechnology, while Rashid et al. (2020) provide a practical approach for proper annotation of standardised and machine-readable datasets. Gajbe et al. (2021) review several data management plans to assist researchers in choosing appropriate structures to aid their data management process. There appears to be a shift from viewing datasets as a byproduct of a research project to a value-enhancing asset whose value increases the more it is used (Leonelli, 2016; Wilkinson et al., 2016). This shift necessitates further classification of "FAIRified" datasets in AI research and identifying methods to promote the implementation of the FAIR Principles across a wide range of research disciplines.

Defining the four facets of the FAIR Data Principles concisely using plain language is quite difficult. It is important to note that data are information that can be operated on or analysed, whereas metadata describes the relevant information about the data. The first publication of the FAIR Principles by Wilkinson et al. (2016) provides complex, official explanations of each of the four pillars, breaking them down further into a number of actionable guidelines ([Figure 1](#)).

The first pillar, "Findable", is the most essential step in the process, as data openness and findability enhance their value to the academic community and decision-makers. Vesteghem et al. (2020) define Findable data as that which can be found online through indexing in search engines. While this definition is concise and straightforward, this systematic review analysis will

refer to Findable data as that which are described, identified, and indexed in a searchable resource clearly and unequivocally (adaptation of Boeckhout et al. (2018)’s definition).

The second pillar, “Accessible”, relates to the need to make the used databases more accessible to the public. Vesteghem et al. (2020) claim that the data must be retrievable through a standard procedure or an approval process. Conversely, Wilkinson et al. (2016) note that metadata must always remain accessible. This issue regarding the lack of persistent metadata (when the data are inaccessible) is an essential but overlooked part of the FAIR Principles. To construct this more plainly, Wise et al. (2019) break it down into three major components: access protocol, access authorisation, and metadata longevity. So, the data must have a distinguished access procedure that should allow for authentication, if necessary, with metadata that always remains accessible.

The third pillar, “Interoperable”, is premised on increasing data integration at every level to allow the data to be used and understood by computer systems. Boeckhout et al. (2018) provide a simple but concise version of the definition where it is claimed that the data and metadata must be conceptualised, expressed and structured using common published standards. However, one concept that is not explicitly stated in these definitions yet assumed to be inherently present is the importance of the data being machine-accessible (Tanhua et al., 2019). This machine accessibility is essential for a data publisher, as datasets can slip through automated searches if this accessibility is not present.

The final pillar, “Reusable”, aims to optimise the ability of the data to be reused by others. If data are to be deemed Reusable, then they must already be deemed Findable, Accessible and Interoperable. We combine definitions given by Boeckhout et al. (2018) and Rodríguez-Iglesias et al. (2016) and state that Reusable data should have their characteristics described in detail according to domain-relevant community standards and with a clearly defined mechanism for accessing provenance and license information.

AI Applications in Finance

Ahmadi (2024) provides an all-encompassing overview of how AI is changing the current financial landscape, citing fraud detection and compliance, banking chatbots, and algorithmic trading as the most successful implementations of AI in financial services (Mahalakshmi et al., 2022). Similarly, Tekaya et al. (2020) offer an overview of AI applications in other areas of finance, namely credit risk management, the banking sector, and insurance. Several other articles provide more precise research on a specific use of AI, such as Choi and Lee (2018)’s paper, which reviewed the newest financial fraud-detection technique to assess its viability. Similarly, Gómez Martínez et al. (2019)’s paper describes an algorithmic trading system whose decisions are based solely on investors’ moods through the use of big data. Singh and Jain (2018) focus on using AI engines within financial institutions, describing how certain banks use an AI engine that mines data, creating consumer profiles that can then be matched to wealth management products

they would most likely desire. Similarly, Alexandra and Sinaga (2021) conducted data mining on a dataset of client subscriptions in Portuguese banks to predict whether the banks' clients would subscribe to a term deposit. The research found that the Random Forest model (RF) was the most effective classification method, and the K-medoids model was the most effective clustering method in terms of prediction accuracy.

The use of AI in the financial services industry has already proven successful (Fisher et al., 2016; Senave et al., 2023; Xing et al., 2018), with the current wave of AI growth (especially relating to generative AI and large-language models (LLMs)) impacting the industry even further. AI models hold the key to greater financial intelligence, which presents the financial services industry with the opportunity to reduce inefficiencies and costs (Douglas, 2024; Fares et al., 2023). The increased use of AI presents many commercial benefits. However, there are risks associated with the opaqueness of such AI models (Agu et al., 2024). Many of the aforementioned AI models lack transparency, which explains how they reached their conclusions. With public trust in banks and financial institutions remaining low since the global financial crisis, deploying an AI system that accepts or denies loan applications without an accessible reason could only further dissipate this trust (Jadhav & Khang, 2025).

AI Applications in Taxation

The use of AI in taxation is a complicated process, as there is a balancing act present between the rationalisation and democratisation needed to implement a successful AI system and striving for equality using tax policies and legislation while achieving maximum possible tax compliance (Mökander & Schroeder, 2024). Z. Huang (2018) contributes to the overall literature with an overview of current AI developments that are remoulding the taxation sector, discussing tools that are primarily easing the issue of complex data processing within the sector for pertinent issues, such as tax evasion. Adamov (2019) further adds to this discussion, focusing on the importance of AI in tax fraud detection, with AI research and models signalling a unique opportunity to enhance tax compliance.

Raikov (2021) presents a potential AI system based on neural networks and cognitive modelling that predicts tax evasion events, research that enhances tax practitioners' and regulators' capacity to circumvent such occurrences. Interestingly, the author also notes the relationships created in the digital environment (digital transactions, e-commerce and cryptocurrencies), which make such evasion events difficult to regulate (Rahman et al., 2024). The paper showcases how deep learning applications can be leveraged to predict evasion using complex decision-making systems involving text analysis (Raikov, 2021). However, the authors do not consider the issues of trust, transparency, and fairness due to the lack of interpretability or explainability of these deep neural network models, issues to which regulators must pay close attention. Predictive models such as these have gained popularity in the literature, with Nembe et al. (2024) further

highlighting the benefits of AI for tax-evasion mitigation and tax non-compliance identification. Additionally, Baghdasaryan et al. (2022) developed a tax-fraud classification model using numerous machine learning (ML) tools in a similar fashion to Pérez López et al. (2019), who used a neural networks-based model for tax-fraud detection. Baghdasaryan et al. (2022) developed a useful ML tool for tax-fraud prediction, even with missing significant data features, such as historical audits or previous instances of fraud. Alongside this, Pérez López et al. (2019) present a classification analysis of an individual's propensity to commit income tax fraud, applying the Multilayer Perceptron network analysis. The model was found to be ~85% effective at classifying potentially fraudulent taxpayers, therefore offering governments a method of bolstering their tax-compliance capabilities.

Looking toward policies, Zheng et al. (2022) comment on the form AI-driven taxation policies would take in comparison to the current baseline policies, noting that the policies utilising reinforcement learning are significantly different and more efficient than the baseline policies in reducing inequality. Mökander and Schroeder (2024) also highlight the social and ethical tensions in using AI in tax policy automation and optimisation. As AI implementation requires rationalisation and democratisation of the tax policymaking process, ethical policymaking and decision-making autonomy may be hindered (Mökander & Schroeder, 2024).

All of these AI methods require vast amounts of sophisticated and personal data to function effectively. However, taxpayer data are highly sensitive and must be protected from unauthorised access (Ezeife, 2021). Aziz and Dowling (2019) comment on the problem of sharing personal financial data, which poses challenges to those publishing their research to propel the use of AI within finance or taxation. This paper highlights how this problem is much more prevalent within the taxation literature, where papers often fail to adhere to the FAIR Principles simply because the data they use cannot be disclosed or even described. (Tax) accountants are tasked with deploying the ethical use of AI in financial services, with building their knowledge of ethical AI practices seen as a proactive approach to AI risk management (Schweitzer, 2024). Aziz and Dowling (2019) particularly highlight the role of the General Data Protection Regulation (GDPR) (EU, 2016) in this risk-management process, which enforces regulations regarding the use of personal data. A potential solution to this problem is presented later in the paper in the form of data anonymisation.

AI Applications in Insurance

Many articles highlight the importance and advantages of AI applications in the insurance industry, with a further influx of AI applications in the industry expected in the coming years (Owens et al., 2022; Paruchuri, 2020; Riikinen et al., 2018; Umamaheswari & Janakiraman, 2014). Fraud detection (Sithic & Balasubramanian, 2013; Verma et al., 2017) and claims reserving (Baudry & Robert, 2019; Blier-Wong et al., 2021; Lopez & Milhaud, 2021; Wüthrich, 2018) are two areas which offer the most

significant potential for AI implementation. AI provides long-term profitability growth potential to insurance companies primarily through enhanced risk assessment (Grize et al., 2020). Sarkar (2020) argues that the insurance industry holds the potential for algorithmic capabilities to enhance each stage of the insurance value chain (IVC). By highlighting AI's offerings at each stage of the IVC, the research prompted further studies from Walsh and Taylor (2020) and Eling et al. (2021) to determine precise AI opportunities available to the insurance industry. Walsh and Taylor (2020) highlight AI models' ability to mimic or augment human capabilities with Natural Language Processing (NLP), Internet of Things (IoT), and computer vision. Eling et al. (2021) analyse AI's impact at each stage along the IVC and specifically highlight AI's potential to enhance risk assessment (loss prediction and prevention) and revenue streams for insurance practitioners.

Applied research in the AI for insurance literature set highlights the specific instances where AI can be used along the IVC. Ensemble models are most popular in insurance use cases, with neural networks and clustering models also proving popular (Owens et al., 2022). Fang et al. (2016) used big data to develop a new profitability method for insurers using historical customer data, where they found that the RF model outperformed other forecasting methods (linear regression and Support Vector Machine (SVM)). Shapiro (2007) documents the extent to which fuzzy logic (FL) has been applied to insurance practices, which prompted Baser and Apaydin (2010)'s research on claims reserving using hybrid fuzzy least squares regression and Khuong and Tuan (2016)'s creation of a neuro-fuzzy inference system for insurance forecasting. Decision trees (DT) are often used in insurance claims prediction (Quan & Valdez, 2018), with further research citing the low predictive power of DTs. Therefore, Henckaerts et al. (2021) boosted DTs' intrinsic interpretability to create a robust insurance pricing model. As AI is impacting each stage of the IVC, an analysis of its specific use cases along the IVC will help regulators assess and, ultimately, govern AI-applied insurance industry operations. Human bias and dataset bias, leading to potentially discriminatory insurance practices, are pertinent concerns for insurance policymakers.

The bias inherent to black-box AI systems threatens trust within the insurance industry, with this bias primarily driven by either humans' input or algorithmic bias (H. Koster, 2020; Ntoutsis et al., 2020). There is potential for these models' impediments to compound and extenuate bias in their decision-making processes with unfair outcomes possible within the insurance industry (Confalonieri et al., 2021; O. Koster et al., 2021). Bias in AI models could potentially lead to discriminatory behaviour in the AI system caused by the model's tendency to use sensitive information resulting in unfair decisions (Barocas & Selbst, 2016). There is substantial research conducted on the determination of responsible AI, with O. Koster et al. (2021) providing a framework to create a responsible AI system, and Arrieta et al. (2020) outlining degrees of fairness to be implemented in an AI

system to reduce discriminatory issues. Transparency and accountability of AI research in the insurance industry are therefore of paramount importance to ensure a successful AI-infused insurance industry for many years to come (EIOPA, 2021). Additionally, taxation and the provision of insurance are agents of governmentality and, ultimately, state power (Pellandini-Simányi, 2020). Therefore, refining an open and accessible research set on AI's impact on these industries is necessary.

2. Methodology

The selection of suitable databases as the sources for the systematic search was based on Gusenbauer and Haddaway (2020)'s paper, which reviewed over 20 databases and tested their ability to perform quality systematic reviews while also identifying which databases are best suited to each research area. The systematic search was conducted on EBSCOhost (EconLit and Business Source Complete), ACM Digital Library, Scopus, Web of Science, and IEEE Xplore. These databases were selected due to their content primarily focusing on economics and business, along with a computer science-based database to access research papers with AI applications. This group of academic search systems are well-suited to systematic reviews because they meet all necessary performance requirements and can be used as either principal or supplementary search systems (Gusenbauer & Haddaway, 2020).

To form a consistent, systematic search, this paper uses a combination of keywords to grasp the full breadth of relevant papers ([Table 1](#)). The breakdown of “taxation” terms is influenced by Redondo et al. (2018). Similarly, interchangeable “finance” terms are adapted from Akomea-Frimpong et al. (2022)'s and X. Xu et al. (2018)'s systematic reviews. Interchangeable search terms for “insurance” are derived from Owens et al. (2022) on the proliferation of Explainable Artificial Intelligence (XAI) in insurance. Eling et al. (2021) provide a table of terms that can be used interchangeably with three key AI terms (Artificial Intelligence, Smart Devices and Artificial Neural Network). The broad set of search terms encapsulates a comprehensive list of articles that demonstrate the breadth of AI applications within financial services.

The search criteria for each of the aforementioned databases are as follows²:

- *Time period*: Articles³ published between 1st January 2000–31st December 2023 are included.
- *Relevancy*: The presence of keywords ([Table 1](#)) in the abstract is necessary for the article's inclusion. Additionally, the article must be relevant to assessing AI applications in financial

² Adapted from Owens et al. (2022).

³ Throughout this review “articles” refers to both academic articles and conference papers.

Table 1. Keywords and Derived Search Terms used in the Systematic Search for AI in Financial Services Literature

Keyword	Search terms
Taxation	Taxation, Financial Policy, Taxes, Tax, Taxing, Fiscal Policy OR
Finance	Finance, Financing, Loan, Investment, Securities, Bond, Credit, Portfolio Management OR
Insurance	Insurance, Insurer AND
Artificial Intelligence	Artificial Intelligence, AI, Machine Learning, Deep Learning, Big Data OR
Artificial Neural Network	Artificial Neural Network, Swarm Intelligence, Support Vector Machine, Computational Intelligence

services (encompassing finance, taxation and insurance for the current study); and only articles which utilise a dataset for their study are included (i.e., no systematic reviews or discussion articles).

- *Singularity*: Duplicate articles found across the various databases are excluded.
- *Accessibility*: Only peer-reviewed articles that are accessible through the aforementioned databases and are accessible in full text are included (i.e., extended abstracts are not included).
- *Language*: Only articles published in English are included.

The initial screening process involved the analysis of 1,672 articles (following duplicate removal) based on their title, abstract and relevancy to the analysis. After filtering the literature using the above criteria, 102 articles were included for FAIR assessment. A backward search involving the evaluation of the 102 relevant articles' bibliographies for additional articles relevant to the topic was conducted, identifying a further 24 relevant articles. This entailed searching through the citations of the remaining literature, ensuring the review would gather up any related articles that may have been missed in the initial search process. A final sample of 126 articles (including backward search results) was established. The PRISMA flow diagram ([Figure 2](#)) depicts the systematic review process, displaying the number of articles excluded at each review stage.

2.1. Literature Extraction Process

The evaluation of the full-text articles within the three literature sets followed a two-step process:

1. Recognition: The applied AI method was recognised, along with the employed dataset.

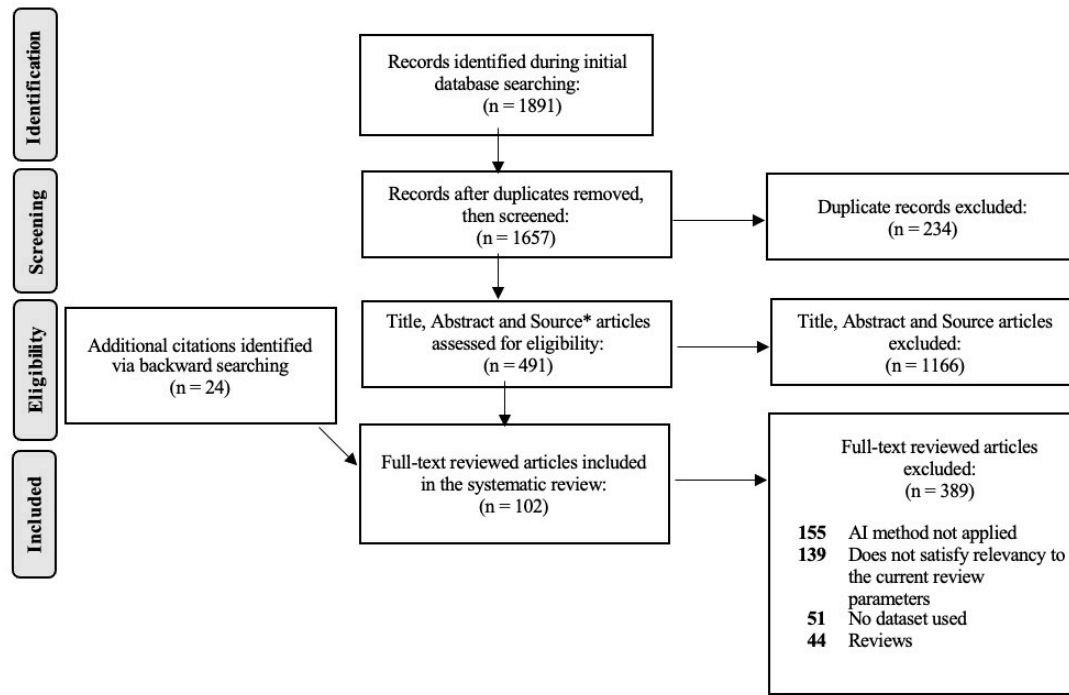


Figure 2. PRISMA Flow Diagram

Note. The PRISMA Flow Diagram is a recognised standard for systematic review literature search processes. “Source” refers to the article inclusion criteria for this systematic review: journal articles and conference papers/proceedings are included.

Table 2. Article Analysis Process

Recognition	Article + Dataset
	↓
FAIR Data Analysis	Neural Network Regression Data Mining Clustering Bayesian Network Rule-Based Discriminant Analysis Ensemble Support Vector Machine Reinforcement Learning Decision Tree
	↓
	Findable → Accessible → Interoperable → Reusable

Note. The process includes the recognition phase, whereby articles which meet the inclusion criteria for the present systematic review are identified. Articles including the above keywords are additionally assessed for their inclusion of AI applications’ processes. The second phase of the process includes assessing these articles’ adherence to the FAIR Data Principles.

2. FAIR Data Analysis: Each article was assessed per the FAIR Data Principles.

The authors conducted a manual assessment of the articles to determine their adherence to the FAIR Data Principles. Using the FAIR Principles framework ([Figure 1](#)), the authors assessed each of the FAIR steps according to their specific features (i.e., *Findable* – Is the unique identification number visible to the authors? Are the data adequately described with rich metadata? Are the data indexed in a searchable resource/database?). These features are adapted from Wilkinson et al. (2016)’s official FAIR framework.

The listed AI methods are adapted from Payrovnaziri et al. (2020) and Owens et al. (2022) systematic review papers. An outline of the process is described in [Table 2](#) above.

2.2. Research Limitations

Industry reports are excluded to ensure the validity and reproducibility of the systematic review. Access to complete methodological processes is pertinent to the systematic review, as the used datasets and their corresponding description are often observed in this section. Industry reports often omit this section in their publication.

3. Results – Reproducibility Evaluation

Like many other research fields, AI and ML face a reproducibility crisis (Gibney, 2022). We similarly observed a lack of reproducible studies during this systematic review of articles. A summary of the FAIR Data Principles is provided in [Figure 1](#), which informs the reproducibility criteria used in the current systematic review. Following the review of 126 articles, 35 articles (29%) were identified as fully satisfying the criteria of open and reproducible datasets, with a further 26 articles (21%) partially satisfying the criteria. [Table 3](#) acknowledges those articles which satisfy the FAIR Data Principles. This low number of reproducible studies highlights the lack of adherence to open data and open science objectives amongst the financial services research community, which could impede research, development, and innovation in the applied AI domain.

The adherence to the FAIR Principles in the financial services research domain highlights a stark need for further efforts in reproducibility-enhancing practices by researchers. Finance articles have the highest level of FAIR Principles adherence (67%), primarily due to the use of synthetic datasets for analysis. Insurance (44%) and tax (5%) typically do not refer to the datasets used for analysis, or the datasets analysed present personal information such that they cannot be shared publicly with the article's publication. Several articles that did not provide clear direction or insights on the dataset used were concerned with prediction tasks. Such datasets were either anonymised or did not highlight the reasoning for the lack of dataset identification. We note that several of the reviewed studies utilised a simulated dataset, as real-life datasets can sometimes be too short term to test the validity of newly proposed models⁴.

The analysis further finds that data repositories (including Kaggle and UCI Machine Learning Repository) show strong adherence to the FAIR Data Principles, compared with stand-alone published datasets, e.g., Lending Club Dataset and PPD Credit Dataset. These repositories offer a structured place for researchers to deposit their datasets, allowing easier and safer access and downloading. Normore and Tebo (2011) point towards the importance of long-term planning for data preservation and storage, along with standard data methods and policies. Data repositories that are frequently used in

⁴ Including, yet not exclusive to, the following reviewed articles: Baudry and Robert (2019), Gan (2013), Gan and Huang (2017), and Kowshalya and Nandhini (2018).

Table 3. Reproducible Research in the Systematic Articles Segmented by Sector

Reference	AI method	Findable	Accessible	Interoperable	Reusable
Finance					
(Abakarim et al., 2018)	Neural Network	Yes	Yes	Yes	Yes
(Abe et al., 2010)	Reinforcement Learning	No	Yes	Yes	No
		No	No	No	No
(Ala'raj et al., 2022)	Neural Network	No	No	No	No
		Yes	Yes	Yes	Yes
(Alarfaj et al., 2022)	Neural Network	Yes	Yes	Yes	Yes
(Alasbahi & Zheng, 2022)	Neural Network	Yes	Yes	Yes	Yes
		Yes	Yes	Yes	Yes
		Yes	Yes	Yes	Yes
		No	No	No	No
(Antonakis & Sfakianakis, 2009)	Bayesian Network	No	No	No	No
		No	Yes	Yes	No
(Bestami Yuksel et al., 2020)	Ensemble	Yes	Yes	Yes	Yes
(Bianchi et al., 2021)	Neural Network	Yes	Yes	Yes	No
		No	No	No	No
(Chang et al., 2021)	Neural Network	Yes	Yes	Yes	Yes
(Q. Chen et al., 2020)	Ensemble	Yes	Yes	Yes	Yes
(X. Chen et al., 2019)	Neural Network	Yes	Yes	Yes	Yes
(Chuang & Huang, 2011)	Neural Network	Yes	Yes	Yes	Yes
(Deng et al., 2019)	Support Vector Machine	Yes	Yes	Yes	Yes
		Yes	Yes	Yes	Yes
(Esenogho et al., 2022)	Ensemble	Yes	Yes	Yes	Yes
(Fan et al., 2018)	Support Vector Machine	No	Yes	Yes	No
(Gao et al., 2021)	Data Mining	Yes	Yes	Yes	Yes
(Gu et al., 2020b)	Neural Network	Yes	No	No	No
(Guo et al., 2019)	Ensemble	No	Yes	Yes	No
		Yes	Yes	Yes	Yes
		No	Yes	Yes	No
(Gweon et al., 2020)	Ensemble	Yes	Yes	Yes	Yes
(Harris, 2015)	Support Vector Machine	Yes	Yes	Yes	Yes
(D. Huang et al., 2021)	Data Mining	Yes	Yes	Yes	Yes
(S.-H. Huang et al., 2021)	Reinforcement Learning	Yes	No	No	No
(Ileberi et al., 2021)	Data Mining	Yes	Yes	Yes	Yes
(Jin et al., 2021)	Ensemble	Yes	Yes	Yes	Yes
		No	No	No	No
		Yes	Yes	Yes	Yes
(Junyu, 2020)	Data Mining	Yes	Yes	Yes	Yes
(Kalid et al., 2020)	Ensemble	Yes	Yes	Yes	Yes
		Yes	Yes	Yes	Yes
(Lahmiri & Bekiros, 2019)	Neural Network	Yes	Yes	Yes	Yes
(Z. Li et al., 2021)	Reinforcement Learning	Yes	Yes	Yes	Yes
		Yes	No	No	No
(Lin et al., 2019)	Data Mining	No	Yes	Yes	No
		Yes	Yes	Yes	Yes
		No	No	No	No

Reference	AI method	Findable	Accessible	Interoperable	Reusable
(Liu et al., 2021)	Data Mining	Yes	Yes	Yes	Yes
		No	No	No	No
		No	No	No	No
		No	No	No	No
(Prusti et al., 2021)	Data Mining	Yes	Yes	Yes	Yes
(Randhawa et al., 2018)	Ensemble	Yes	Yes	Yes	Yes
(Ravi et al., 2017)	Fuzzy Logic	Yes	Yes	Yes	Yes
(Soui et al., 2020)	Data Mining	No	No	No	No
		Yes	Yes	Yes	Yes
(Tripathi et al., 2019)	Ensemble	No	Yes	Yes	No
		No	Yes	Yes	No
		Yes	Yes	Yes	Yes
		Yes	Yes	Yes	Yes
(Tsai, 2008)	Data Mining	No	Yes	Yes	No
		Yes	Yes	Yes	Yes
		No	No	No	No
		No	Yes	Yes	No
(Turiel & Aste, 2020)	Data Mining	Yes	Yes	Yes	Yes
(Wang et al., 2019)	Data Mining	Yes	Yes	Yes	Yes
(S. Wei et al., 2019)	Ensemble	No	Yes	Yes	No
		No	Yes	Yes	No
		No	No	No	No
(C. Xu, 2021; R. Xu et al., 2021)	Ensemble	No	Yes	Yes	No
(Ye, 2021)	Data Mining	Yes	No	No	No
		No	No	No	No
(Yuan, 2022)	Ensemble	Yes	Yes	Yes	Yes
(Zhang & Lu, 2021)	Ensemble	Yes	Yes	Yes	Yes
(Zou, 2021)	Data Mining	Yes	Yes	Yes	Yes
Insurance					
(Alshamsi, 2014)	Ensemble	Yes	Yes	Yes	Yes
(Benedek & László, 2019)	Data Mining	Yes	No	No	Yes
(Bian et al., 2018)	Ensemble	No	No	No	Yes
(Deprez et al., 2017)	Ensemble	Yes	Yes	Yes	Yes
(Herland et al., 2018)	Ensemble	Yes	Yes	Yes	Yes
(Johnson & Khoshgoftaar, 2019)	Neural Network	Yes	Yes	Yes	Yes
(Khodairy & Abosamra, 2021)	Neural Network	Yes	Yes	Yes	Yes
(Lee et al., 2020)	Regression	Yes	Yes	Yes	Yes
(Neumann et al., 2019)	Decision Tree	Yes	No	No	No
(Schelldorfer & Wuthrich, 2019)	Ensemble	Yes	Yes	Yes	Yes
(Shah & Guez, 2009)	Neural Network	Yes	Yes	Yes	Yes
(Smyth & Jørgensen, 2002)	Regression	Yes	Yes	Yes	Yes
(Viaene et al., 2004)	Ensemble	Yes	No	No	No
(Viaene et al., 2002)	Regression	Yes	No	No	No
(C. Wei & Dan, 2019)	Support Vector Machine	Yes	Yes	Yes	Yes
(W. Xu et al., 2011)	Ensemble	Yes	Yes	Yes	Yes
Tax					

Reference	AI method	Findable	Accessible	Interoperable	Reusable
(Höglund, 2017)	Discriminant Analysis	Yes	Yes	No	No

Note. The datasets included in the systematic review may be open and accessible for further research. However, as this is not explicitly stated in many of the articles, the result returns “No”. Of the 61 articles displaying any level of adherence to the FAIR criteria, there are 90 different instances of dataset use. Only 35 of the 90 instances of dataset usage (of which some are repeated) fully satisfy the FAIR criteria. Note that some articles utilised 2+ datasets in their analysis, all of which are referred to below.

articles that fully satisfy the FAIR Principles offer long-term data storage and preservation, with the ability to update the dataset if required. Repositories such as Kaggle display Creative Commons licensing information, offering clear usage rights to potential data users. This is one of the main benefits of these repositories, as it clarifies any researcher’s ambiguity about reusing a dataset. A reoccurring theme in datasets that were not published on respected repositories such as Kaggle or the UCI Machine Learning Repository was that the dataset failed under just a single facet of the FAIR Principles. Some of the reoccurring themes include but are not exclusive to (1) inadequate metadata, (2) lack of persistent identifiers, (3) accessibility issues, and (4) reusability issues.

3.1. Inadequate Metadata

A common inadequacy of several datasets was the lack of rich and extensive metadata. Metadata is data or descriptive information about the dataset, including its context, quality, and characteristics (Guptill, 1999). Rich metadata ensures that the data can be used for other purposes along with its intended initial use. However, datasets that lacked sufficient metadata were not only exclusive to datasets outside of respected repositories. One such dataset is the UCI Machine Learning Repository – Australian Credit Dataset, which appears numerous times across the literature. The data concerns Australian credit card applicants and is not dissimilar to the UCI Machine Learning German Credit Dataset (Hoffman, 1994). However, the Australian Credit Dataset lacks the rich metadata needed for the data to be considered flexible and useable in different contexts. The issue lies in data confidentiality, which forces the data providers to assign meaningless values to the attributes. The UCI Machine Learning – Japanese Credit Screening Dataset (Sano, 1992) provides even less metadata, lacking even basic details such as the number of attributes in the data or if there are missing value points – essential pieces of information. Data confidentiality is not mentioned as an issue in this instance, meaning a dataset such as this can smoothly transition into a FAIR dataset with increased metadata input.

3.2. Lack of Persistent Identifiers

Several datasets were not shared due to confidentiality reasons. However, the data, along with their corresponding metadata, should still be assigned a persistent identifier so that they are easily Findable. Didimo et al. (2020) apply ML algorithms in a tax-risk assessment manner using taxpayer data provided by the Italian Revenue Agency. These data were not shared due to

confidentiality reasons, nor were they assigned a persistent identifier, making the informative research irreproducible. Similarly, L.-H. Li et al. (2017) use credit records of small loan applications provided privately by a Brazilian commercial bank. Again, the dataset is not contained within a persistent identifier.

3.3. Accessibility Issues

The issue of open and closed datasets is an important topic of debate, and the FAIR Principles offer flexibility around the issue. The Principles allow for a dataset to be private once its metadata is publicly available and accessible. Often, datasets did not provide open-source metadata, meaning they did not fully satisfy the accessibility requirements. One such instance is the dataset analysed by Gu et al. (2020a). The article provides sufficient metadata to distinguish the dataset from others, but its persistent identifier exists on the Wharton Research Data Services website. An institutional login is required to see both the data and metadata, meaning the dataset is not fully accessible. This is a persistent issue across datasets that exist on the Wharton Research Data Services website. Siloed data such as the abovementioned datasets create an access-to-knowledge imbalance amongst researchers and commercial institutions, whereby only researchers with the means to pay for datasets behind paywalls can further contribute to the literature set using those data. Additionally, research published using inaccessible datasets is often irreproducible, thereby hindering the validity of the research outputs. It is those researchers, research groups or commercial bodies who have the means to either pay to access datasets or collect primary data from individuals themselves who hold the power to further the literature set via research publications (Foucault, 2020).

In a similar fashion, hyperlinks to datasets can become obsolete over time. In this scenario, there should still be accessible metadata at the very least to describe these now obsolete data, which was not the case in most instances. For example, Liu et al. (2021) are open with their data, providing links to two datasets used in the research. However, the link to one of these datasets has since become obsolete, with no persistent metadata present, meaning a failure under the second Accessibility principle.

3.4. Reusability Issues

An issue that is briefly discussed above is the lack of clear usage rights attached to a dataset. Often, datasets that are not published on academic data repositories do not clearly state their usage rights, meaning a researcher hoping to reuse the data may simply avoid using them to prevent copyright infringement. An example of one such dataset is that which is used in Bianchi et al. (2021)'s paper focused on the predictability of bond returns. This dataset (Cynthia Wu, 2021) fully satisfies all aspects of the FAIR Principles but fails to provide any obvious licensing information that would be clearly

acknowledged in any automated searches. Providing a simple usage license to these data would transform this unique dataset into a distinctly valuable and reusable asset.

4. Discussion

4.1. Research Accessibility Crisis and Regulatory Challenges

Research progress hinges on the sharing of information through two generic processes: exchange and collaboration (Nahapiet & Ghoshal, 1998). As outlined in the previous results section, there is a significant lack of reproducible datasets present in AI and financial services research, with 65 articles analysed in this paper not adhering to the FAIR Principles in any instance. Highlighted in Section 3 are the most commonly reoccurring reasons for research datasets not being published: 1) inadequate metadata, 2) lack of persistent identifiers, 3) accessibility issues and 4) reusability issues. Firstly, the lack of emphasis on AI studies' reproducibility hinders both significant regulatory progress and research enhancement within the industry. If researchers used publicly available datasets which satisfy the outlined FAIR Data Principles criteria, their studies' validity could be ascertained. In addition to the validation of research findings, future studies could expand on research methodologies in tandem with advances in technology adoption. The lack of available datasets significantly restricts the furthering of regulatory, academic, and professional knowledge on the ubiquity of AI applications in a wide range of financial services tasks, limiting researchers' scope for research development. The exchange and collaboration of sensitive datasets between commercial and academic entities for research purposes may hamper cross-industry innovation for AI use in financial services (Nahapiet & Ghoshal, 1998). Additionally, if the academic community cannot gain access to relevant commercial datasets due to their sensitive nature, the conversation between policymakers and academia will falter, leading to ill-informed policy decisions on AI applications in financial services (Head, 2016).

We cite the issue of lacking dataset availability in financial services research as a reproducibility crisis, which contrasts with Jensen et al. (2023)'s findings, which negates a "replication crisis" in financial economics. Their findings conclude that much of the literature sample can be replicated⁵, which the authors initially expected to be similar to the findings of the current paper. However, this contrast in reproducibility potential in three closely related fields of finance research (vs. financial economics research) shows that reproducibility is possible in financial services research datasets, albeit with additional obstacles in the case of the financial services industry, such as sensitive and confidential personal data on consumers, as outlined above. Without harnessing the potential for research reproducibility through

⁵ Note that Jensen et al. (2023) did not evaluate their data sample using the FAIR Data Principles guidelines but rather a unique Bayesian model of factor replication.

accessible datasets, financial services research is at risk of significantly lagging behind its financial economics colleagues. Without transparent research practices and accessible data, the advancement of AI-applied financial services research could additionally fall behind commercial financial institutions' practices, rendering researchers' work out of date and irrelevant once published.

Although a call for publicly available datasets in various research communities is welcomed (Christensen & Miguel, 2018; Vetter et al., 2016), the financial services industry faces additional challenges regarding dataset access. Data collected by insurers and banks, in particular, are usually of a highly sensitive nature, with insurance and credit proposal forms asking for specific information about proposed insureds in many cases. Researchers have attempted to overcome this issue by utilising synthetic datasets in their analysis due to heavily restricted access to financial services companies' datasets. Such issues are only heightened by recent and ongoing regulations in the industry that protect proposed insureds' and credit customers' right to privacy and the safe-keeping of sensitive data shared with insurance companies and banks, such as GDPR regulations (EU, 2016). Therefore, only financial services entities (insurance companies, banks, commercial finance firms) regularly hold a sufficient repository of data on multitudes of data subjects. In some instances, however, academic researchers can access anonymised datasets when needed for analysis. This practice is facilitated by commercial entities charging a fee in exchange for anonymised datasets. This "paywall" maintains the integrity of datasets' use to some degree by ensuring that proper anonymisation of datasets is completed, as commercial entities are paid for this service by researchers via the dataset-access fee.

Although a practical and regulatory sound process in relation to data privacy, the practice of holding financial services subjects' anonymised data behind paywalls is damaging to regulatory and academic research progress (Hagger, 2022). As outlined in the above review of financial services literature, research reproducibility levels are low as often researchers cannot share datasets they have used in their analysis, even if they have paid a fee for such data. In essence, researchers can pay a fee to utilise datasets in their research. However, this fee does not make them an "owner" of such anonymised datasets. Therefore, scholarly research results published in financial services domains, although peer-reviewed and published in academic journals, often lack validation and research reproduction capabilities. The continuation of this practice in tandem with the FAIR Principles' popularity growth seems counterintuitive. However, in some research fields, such as financial services, the protection of data subjects' privacy remains a priority. It is for this reason that an adapted FAAIR framework, including the process of anonymisation, is necessary for the future of financial services research. Regulatory advancements and the associated risk management of AI applications in financial services will benefit from extensive and verified research. These research results will be based upon analyses of relevant

datasets (with or without sensitive information included), not just those datasets which the researchers can access while operating within the bounds of stringent anonymisation regulations.

The anonymisation of datasets is possible, with such actions already undertaken in genomic research (Corpas et al., 2018) and government-level policy decisions (Ritchie, 2014). The creation of an open-access data anonymisation toolkit for financial services research would benefit the research community by enabling reproducibility and validation of results, which will, in turn, allow for a successful and thorough implementation of AI models in financial services practices.

Finally, from a broader societal perspective, policymakers, IDOs, and NGOs routinely struggle to gain sufficient insights to inform evidence-based decision-making for social progress. While government bodies and regulators increasingly rely on external organisations' research data to inform their policymaking, there exists a distinct lack of open and accessible research data in the financial services industry (Mehrban et al., 2020). Financial services regulations employ a subsidiarity approach to regulation, whereby industry players can govern themselves internally once complying with specific industry-wide regulations. By allowing commercial entities to maintain control of their internal operations, policymakers lack access to key research data created and protected by commercial entities (McKinlay et al., 2012). Siloed and inaccessible data are also a key obstacle to IDOs' and NGOs' ability to partake in decision-making on behalf of their stakeholders. One such issue is financial exclusion, as consumer groups and NGOs working on behalf of minority groups require a baseline knowledge of current research trends to take part in societal conversations (Grohmann et al., 2018). The United Nations' Sustainable Development Goals (SDGs) highlight the importance of financial inclusion for economic growth (Ozili, 2018), with financial literacy and knowledge significantly improving society's credit access (Kara et al., 2021). The power of knowledge remains with those commercial entities with access to large sums of research funding, hindering social progress in regulatory spaces and also in the protection of disadvantaged members of society. Again, likening the current situation of research data accessibility in financial services research to Foucault's (2020) power-knowledge theory, the commercial industry, as holders of sensitive personal data on consumers, also holds power. These valuable data are often siloed by commercial entities who have gathered data from their customers. Without a standardised anonymisation framework to allow the transfer of these data for academic analysis, policymakers do not see the full picture relating to the potential for AI applications in financial services. Without the knowledge of the full potential of such applications, there may be regulatory gaps related to AI applications in financial services.

4.2. The Transposability of FAIR Principles to Financial Services Research and FAAIR Adaptation

Following the above analysis, which finds a significant lack of reproducible research on AI applications across finance, tax, and insurance, one must consider the validity of the FAIR Principles in financial services research. Although a step in the right direction for research data transparency and openness, the FAIR Principles as they currently stand do not encapsulate the true meaning of the acronym in English, i.e., the advancement of fair, equitable and inclusive data and research (Sterner & Elliott, 2023). Even if datasets satisfy the FAIR Principles, they could perpetuate existing social injustices borne from sampling biases or inadequate representation of vulnerable groups (Leonelli et al., 2021). Legally, the opacity of black-box AI systems hinders regulatory bodies from determining whether data are processed fairly (Carabantes, 2020). However, evolving XAI methods are enhancing the potential for AI systems' regulation under the European GDPR (EU, 2016), with Hagras (2018) suggesting how FAIR Data can facilitate an easier move to more explainable AI methods. As financial services data are comprised of customers' sensitive and personal information, the anonymity of the data subject is paramount. The GDPR protects the data subject's right to anonymity during data processing stages and ensures that sensitive information is not shared with third parties with whom the data subject has not given consent to view such sensitive personal information (EU, 2016). The FAIR Principles assess the availability of published research data. Considering that financial service research predominantly uses anonymised sensitive information, of which the researcher is usually not the owner, the publication of datasets is not always possible. Therefore, the FAIR Principles are rarely fully satisfied by financial services research. This paper proposes an enhancement of the FAIR Principles to encapsulate the special case of the financial services industry, similar to the extension of the FAIR Principles for medical research proposed by Holub et al. (2018) and Landi et al. (2020) and the CARE Principles⁶ which address the importance of Indigenous peoples in making decisions about themselves, their land and waters (Mushi et al., 2020).

Through including an additional facet, "Anonymisation", the FAAIR Principles (Findable, Accessible, Anonymised, Interoperable, Reusable) will allow for financial services research to be fully accessible and usable to those research and civil society bodies who need data to enact sustainable change in society. Anonymised data encompasses GDPR-compliant data that are not easily traceable back to the original by reducing the precision of data attribute values (with a view to removing them entirely) (El Emam et al., 2015). Though attribute disclosure cannot be entirely avoided (Sun et al.,

⁶ CARE principles = Collective, Authority, Responsibility, Ethics.

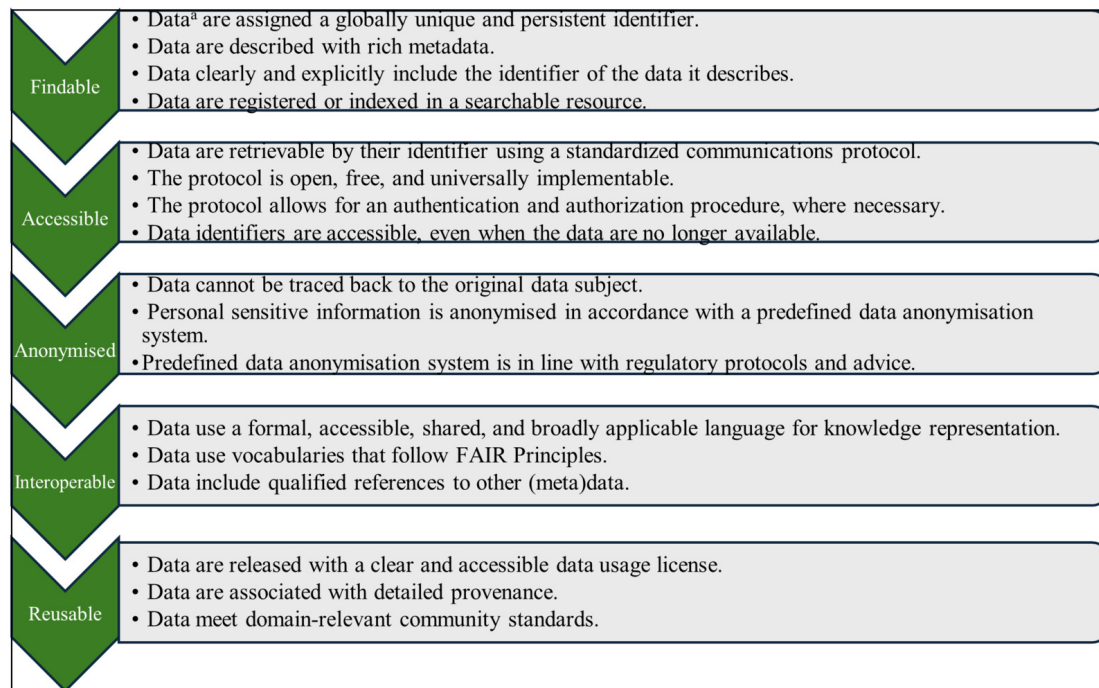


Figure 3. FAAIR Principles Extension of Wilkinson et al. (2016)'s Official FAIR Data Principles Components' Framework

Note. The addition of Anonymised data will allow for further research transparency and reproducibility potential in the financial services domain.

^a "Data" refers to any case where the principles should be applied to metadata and data – additional source: David et al. (2018).

2011), a set of techniques to minimise risks of privacy breaches is at the heart of Anonymised data in the proposed FAAIR Principles. The ensuing ability of IDOs, NGOs, and civil society to access financial research data will facilitate open science initiatives and citizen empowerment in the face of social inequalities, namely financial exclusion (Janssen et al., 2012; Zhu et al., 2019).

Commercial practices will also benefit from the practice of FAAIR research findings through the adaptation of academic research for commercial use. An example of such AI-driven commercial innovations could include the adoption of novel insurance products proposed by the academic community and their translation into successful insurance product offerings. The Pay-As-You-Move (PAYM) multi-modal insurance product framework proposed by Owens et al. (2024) provides an extension and enhancement of the successful Pay-As-You-Drive (PAYD) insurance product, harnessing AI-driven analytics and smart city technologies in its development. Financial services product innovations are produced and driven by both academia and industry, and the sharing of data and insights between both groups allows for continued successful innovations. Through providing a FAAIR framework for financial services research, open science can become a reality by promoting transparency and inclusivity in financial services product development, policymaking, and decision-making (Beck et al., 2020).

5. Conclusion

Research transparency and reproducibility are key concerns of many researchers and journal publication houses, with analysed datasets often requested upon acceptance of an article for publication. Although this practice is a great effort in providing the academic research community with the means to validate and reuse data for future research projects, it is not always possible in financial services research. An analysis of the applications of AI in the financial services literature set showcases a lack of available data used during research analysis. Often, the FAIR Principles framework is used to assess the level of transparency and accessibility of research projects. However, we argue that such Principles are not altogether applicable to the financial services research domain. Often sensitive and personal data are collected, collated and owned by commercial entities who, for regulatory reasons, cannot provide researchers with such sensitive personal data in the majority of circumstances. By incorporating an additional facet (Anonymised data) in the FAIR Principles, these groups can reduce social injustices and support financial inclusion. The financial services industry is a high-value socio-economic example of where the benefits of FAIR can be increased through the implementation of informed strategies to support it. The FAAIR Principles will enhance the development of explainable and responsible AI to reduce social injustices, in particular financial exclusion.

An open-source solution could provide benefits to research reproducibility. However, the difficulty remains that any strategy requires a dynamic and sustainable solution. In this way, FAAIR is an example of such progressive adaptation. If anonymised, datasets gathered by commercial entities and researchers may be accessed by (other) researchers, encouraging further analysis by the wider academic community. IDOs and NGOs representing the public (including minority groups) in the financial services domain will also enjoy adequate access to research data to support informed decision-making, a need which will be satisfied by data published under the FAAIR Data framework. The value of the FAAIR framework to financial services research lies in the potential to analyse datasets containing sensitive personal data which may have been previously protected under GDPR rules. FAAIR allows for continued adherence to data privacy regulations while also aiding advancements in financial services research. Additionally, researchers are encouraged to practice transparent publication practices by publishing datasets (where possible, considering regulation requirements) alongside their peer-reviewed work. This practice of ensuring studies are reproducible and verifiable will further boost research in AI-applied financial services research.

Published: June 16, 2025 IST.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc/4.0> and legal code at <https://creativecommons.org/licenses/by-nc/4.0/legalcode> for more information.

References

- Abakarim, Y., Lahby, M., & Attioui, A. (2018). An efficient real time model for credit card fraud detection based on deep learning. *Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications*, 1–7. <https://doi.org/10.1145/3289402.3289530>
- Abe, N., Melville, P., Pendus, C., Reddy, C. K., Jensen, D. L., Thomas, V. P., Bennett, J. J., Anderson, G. F., Cooley, B. R., & Kowalczyk, M. (2010). Optimizing debt collections using constrained reinforcement learning. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 75–84. <https://doi.org/10.1145/1835804.1835817>
- Adamov, A. Z. (2019). Machine learning and advanced analytics in tax fraud detection. *IEEE 13th International Conference on Application of Information and Communication Technologies (AICT)*, 1–5. <https://doi.org/10.1109/AICT47866.2019.8981758>
- Adolf, M., & Stehr, N. (2016). *Knowledge: Is knowledge power?* Taylor & Francis. <https://doi.org/10.4324/9781315543093>
- Agu, E. E., Abhulimen, A. O., Obiki-Osafiele, A. N., Osundare, O. S., Adeniran, I. A., & Efunniyi, C. P. (2024). Discussing ethical considerations and solutions for ensuring fairness in AI-driven financial services. *International Journal of Frontier Research in Science*, 3, 001–009.
- Ahmadi, S. (2024). A comprehensive study on integration of big data and AI in financial industry and its effect on present and future opportunities. *International Journal of Current Science Research and Review*, 7, 66–74. <https://doi.org/10.47191/ijcsrr/V7-i1-07>
- Akomea-Frimpong, I., Kukah, A. S., Jin, X., Osei-Kyei, R., & Pariafsai, F. (2022). Green finance for green buildings: A systematic review and conceptual foundation. *Journal of Cleaner Production*, 131869. <https://doi.org/10.1016/j.jclepro.2022.131869>
- Ala'raj, M., Abbod, M. F., Majdalawieh, M., & Jum'a, L. (2022). A deep learning model for behavioural credit scoring in banks. *Neural Computing and Applications*, 1–28. <https://doi.org/10.1007/s00521-021-06695-z>
- Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *IEEE Access*, 10, 39700–39715. <https://doi.org/10.1109/ACCESS.2022.3166891>
- Alasbahi, R., & Zheng, X. (2022). An online transfer learning framework with extreme learning machine for automated credit scoring. *IEEE Access*, 10, 46697–46716. <https://doi.org/10.1109/ACCESS.2022.3171569>
- Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-thinking data strategy and integration for artificial intelligence: Concepts, opportunities, and challenges. *Applied Sciences*, 13, 7082. <https://doi.org/10.3390/app13127082>
- Alexandra, J., & Sinaga, K. P. (2021). Machine learning approaches for marketing campaign in Portuguese banks. *3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, 1–6. <https://doi.org/10.1109/ICORIS52787.2021.9649623>
- Alharbi, E., Skeva, R., Juty, N., Jay, C., & Goble, C. (2021). Exploring the current practices, costs and benefits of FAIR implementation in pharmaceutical research and development: A qualitative interview study. *Data Intelligence*, 3, 507–527. https://doi.org/10.1162/dint_a_00109
- Alshamsi, A. S. (2014). Predicting car insurance policies using random forest. *10th International Conference on Innovations in Information Technology (IIT)*, 128–132. <https://doi.org/10.1109/INNOVATIONS.2014.6987575>

- Antonakis, A., & Sfakianakis, M. (2009). Assessing naive Bayes as a method for screening credit applicants. *Journal of Applied Statistics*, 36, 537–545. <https://doi.org/10.1080/02664760802554263>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). 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>
- Aziz, S., & Dowling, M. (2019). *Machine learning and AI for risk management*. Springer International Publishing. https://doi.org/10.1007/978-3-030-02330-0_3
- Baghdasaryan, V., Davtyan, H., Sarikyan, A., & Navasardyan, Z. (2022). Improving tax audit efficiency using machine learning: The role of taxpayer's network data in fraud detection. *Applied Artificial Intelligence*, 1–23. <https://doi.org/10.1080/08839514.2021.2012002>
- Bahlo, C., & Dahlhaus, P. (2021). Livestock data – Is it there and is it FAIR? A systematic review of livestock farming datasets in Australia. *Computers and Electronics in Agriculture*, 188, 106365. <https://doi.org/10.1016/j.compag.2021.106365>
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104, 671. <https://doi.org/10.2139/ssrn.2477899>
- Baser, F., & Apaydin, A. (2010). Calculating insurance claim reserves with hybrid fuzzy least squares regression analysis. *Gazi University Journal of Science*, 23, 163–170.
- Batista, G., & Monard, M. (2018). Data quality in machine learning: A study in the context of imbalanced data. *Neurocomputing*, 275, 1665–1679.
- Baudry, M., & Robert, C. Y. (2019). A machine learning approach for individual claims reserving in insurance. *Applied Stochastic Models in Business and Industry*, 35, 1127–1155. <https://doi.org/10.1002/asmb.2455>
- Beck, M. W., O'Hara, C., Lowndes, J. S. S., Mazor, R. D., Theroux, S., Gillett, D. J., Lane, B., & Gearheart, G. (2020). The importance of open science for biological assessment of aquatic environments. *PeerJ*, 8, e9539. <https://doi.org/10.7717/peerj.9539>
- Benedek, B., & László, E. (2019). Identifying key fraud indicators in the automobile insurance industry using SQL Server Analysis Services. *Studia Universitatis Babes-Bolyai*, 64, 53–71. <https://doi.org/10.2478/subboec-2019-0009>
- Bestami Yuksel, B., Bahtiyar, S., & Yilmazer, A. (2020). Credit card fraud detection with NCA dimensionality reduction. *13th International Conference on Security of Information and Networks*, 1–7.
- Bian, Y., Yang, C., Zhao, J. L., & Liang, L. (2018). Good drivers pay less: A study of usage-based vehicle insurance models. *Transportation Research Part A: Policy and Practice*, 107, 20–34. <https://doi.org/10.1016/j.tra.2017.10.018>
- Bianchi, D., Büchner, M., & Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34, 1046–1089. <https://doi.org/10.1093/rfs/hhaa062>
- Blier-Wong, C., Cossette, H., Lamontagne, L., & Marceau, E. (2021). Machine Learning in P&C Insurance: A review for pricing and reserving. *Risks*, 9, 4.
- Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? *European Journal of Human Genetics*, 26, 931–936. <https://doi.org/10.1038/s41431-018-0160-0>
- Carabantes, M. (2020). Black-box artificial intelligence: An epistemological and critical analysis. *AI & Society*, 35, 309–317. <https://doi.org/10.1007/s00146-019-00888-w>

- Chang, V., Man, X., Xu, Q., & Hsu, C. H. (2021). Pairs trading on different portfolios based on machine learning. *Expert Systems*, 38, e12649. <https://doi.org/10.1111/exsy.12649>
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 1165–1188. <https://doi.org/10.2307/41703503>
- Chen, Q., Zhang, W., & Lou, Y. (2020). Forecasting stock prices using a hybrid deep learning model integrating attention mechanism, multi-layer perceptron, and bidirectional long-short term memory neural network. *IEEE Access*, 8, 117365–117376. <https://doi.org/10.1109/ACCESS.2020.3004284>
- Chen, X., Liu, Z., Zhong, M., Liu, X., & Song, P. (2019). A deep learning approach using DeepGBM for credit assessment. *Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence*, 774–779. <https://doi.org/10.1145/3366194.3366333>
- Choi, D., & Lee, K. (2018). An artificial intelligence approach to financial fraud detection under IoT environment: A survey and implementation. *Security and Communication Networks*. <https://doi.org/10.1155/2018/5483472>
- Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56, 920–980. <https://doi.org/10.1257/jel.20171350>
- Chuang, C. L., & Huang, S. T. (2011). A hybrid neural network approach for credit scoring. *Expert Systems*, 28, 185–196. <https://doi.org/10.1111/j.1468-0394.2010.00565.x>
- Confalonieri, R., Coba, L., Wagner, B., & Besold, T. R. (2021). A historical perspective of explainable artificial intelligence. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11, e1391.
- Corpas, M., Kovalevskaya, N. V., McMurray, A., & Nielsen, F. G. (2018). A FAIR guide for data providers to maximise sharing of human genomic data. *PLoS Computational Biology*, 14, e1005873. <https://doi.org/10.1371/journal.pcbi.1005873>
- Cynthia Wu, J. (2021). *Yield data*. <https://sites.google.com/view/jingcynthiawu/yield-data>
- David, R., Mabile, L., Yahia, M., Cambon-Thomsen, A., Archambeau, A.-S., Bezuidenhout, L., Bekaert, S., Bertier, G., Bravo, E., & Carpenter, J. (2018). How to operationalize and to evaluate the FAIRness in the crediting and rewarding processes in data sharing: A first step towards a simplified assessment grid. *JNSO 2018 – Journées Nationales de la Science Ouverte*.
- Deng, S., Huang, X., Shen, J., Yu, H., Wang, C., Tian, H., Ma, F., & Yang, T. (2019). Prediction and trading in crude oil markets using multi-class classification and multi-objective optimization. *IEEE Access*, 7, 182860–182872. <https://doi.org/10.1109/ACCESS.2019.2960379>
- Deprez, P., Shevchenko, P. V., & Wüthrich, M. V. (2017). Machine learning techniques for mortality modeling. *European Actuarial Journal*, 7, 337–352. <https://doi.org/10.1007/s13385-017-0152-4>
- Didimo, W., Grilli, L., Liotta, G., Menconi, L., Montecchiani, F., & Pagliuca, D. (2020). Combining network visualization and data mining for tax risk assessment. *IEEE Access*, 8, 16073–16086. <https://doi.org/10.1109/ACCESS.2020.2967974>
- Douglas, D. (2024). The transformative impact of AI in finance and banking. *Journal of Investment, Banking and Finance*, 2, 1–8.
- Dunning, A., De Smaele, M., & Böhmer, J. (2017). Are the fair data principles fair? *International Journal of Digital Curation*, 12, 177–195. <https://doi.org/10.2218/ijdc.v12i2.567>
- EIOPA. (2021). *Artificial intelligence governance principles: Towards ethical and trustworthy artificial intelligence in the European insurance sector*.

- El Emam, K., Rodgers, S., & Malin, B. (2015). Anonymising and sharing individual patient data. *BMJ*, 350. <https://doi.org/10.1136/bmj.h1139>
- Eling, M., Nuessle, D., & Staubli, J. (2021). The impact of artificial intelligence along the insurance value chain and on the insurability of risks. *The Geneva Papers on Risk and Insurance – Issues and Practice*, 1–37.
- Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A neural network ensemble with feature engineering for improved credit card fraud detection. *IEEE Access*, 10, 16400–16407. <https://doi.org/10.1109/ACCESS.2022.3148298>
- European Commission. (2019). *Ethics guidelines for trustworthy AI*. High-level expert group on AI, European Commission.
- European Parliament. (2024). *AI Act*.
- European Union. (2016). *General Data Protection Regulations*.
- Ezeife, E. (2021). AI-driven tax technology in the United States: A business analytics framework for compliance and efficiency. *International Journal of Multidisciplinary Research and Growth Evaluation*, 2, 693–701. <https://doi.org/10.54660/IJMRGE.2021.2.1.693-701>
- Fan, Q., Liu, X., Zhang, Y., Bao, F., & Li, S. (2018). Adaptive mutation PSO based SVM model for credit scoring. *Proceedings of the 2nd International Conference on Computer Science and Application Engineering*, 1–7. <https://doi.org/10.1145/3207677.3278014>
- Fang, K., Jiang, Y., & Song, M. (2016). Customer profitability forecasting using big data analytics: A case study of the insurance industry. *Computers & Industrial Engineering*, 101, 554–564. <https://doi.org/10.1016/j.cie.2016.09.011>
- Fares, O. H., Butt, I., & Lee, S. H. M. (2023). Utilization of artificial intelligence in the banking sector: A systematic literature review. *Journal of Financial Services Marketing*, 28, 835–852. <https://doi.org/10.1057/s41264-022-00176-7>
- Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management*, 23, 157–214. <https://doi.org/10.1002/isaf.1386>
- Foucault, M. (2020). Power/knowledge. In S. Seidman & J. C. Alexander (Eds.), *The new social theory reader*. Routledge. <https://doi.org/10.4324/9781003060963-10>
- Furxhi, I., Arvanitis, A., Murphy, F., Costa, A., & Blosi, M. (2021). Data shepherding in nanotechnology. The initiation. *Nanomaterials*, 11, 1520.
- Gajbe, S. B., Tiwari, A., & Singh, R. K. (2021). Evaluation and analysis of data management plan tools: A parametric approach. *Information Processing & Management*, 58, 102480. <https://doi.org/10.1016/j.ipm.2020.102480>
- Gan, G. (2013). Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics*, 53(3), 795–801.
- Gan, G., & Huang, J. X. (2017). A data mining framework for valuing large portfolios of variable annuities. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/3097983.3098013>
- Gao, Y., Zhang, S., & Lu, J. (2021). Machine learning for credit card fraud detection. *Proceedings of the 2021 1st International Conference on Control and Intelligent Robotics*. <https://doi.org/10.1145/3473714.3473749>
- Gibney, E. (2022). Is AI fuelling a reproducibility crisis in science? *Nature*, 608, 250–251. <https://doi.org/10.1038/d41586-022-02035-w>

- Gómez Martínez, R., Prado Román, M., & Plaza Casado, P. (2019). Big data algorithmic trading systems based on investors' mood. *Journal of Behavioral Finance*, 20, 227–238. <https://doi.org/10.1080/15427560.2018.1506786>
- Grize, Y. L., Fischer, W., & Lützelshwab, C. (2020). Machine learning applications in nonlife insurance. *Applied Stochastic Models in Business and Industry*, 36, 523–537. <https://doi.org/10.1002/asmb.2543>
- Grohmann, A., Klühs, T., & Menkhoff, L. (2018). Does financial literacy improve financial inclusion? Cross country evidence. *World Development*, 111, 84–96. <https://doi.org/10.1016/j.worlddev.2018.06.020>
- Groth, P., & Dumontier, M. (2020). *Introduction – FAIR data, systems and analysis*. SAGE Publications. <https://doi.org/10.3233/DS-200029>
- Gu, Y., Yan, D., Yan, S., & Jiang, Z. (2020a). Price forecast with high-frequency finance data. *ACM*. <https://doi.org/10.1145/3340531.3412738>
- Gu, Y., Yan, D., Yan, S., & Jiang, Z. (2020b). Price forecast with high-frequency finance data: An autoregressive recurrent neural network model with technical indicators. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2485–2492. <https://doi.org/10.1145/3340531.3412738>
- Guo, S., He, H., & Huang, X. (2019). A multi-stage self-adaptive classifier ensemble model with application in credit scoring. *IEEE Access*, 7, 78549–78559. <https://doi.org/10.1109/ACCESS.2019.2922676>
- Guptill, S. C. (1999). Metadata and data catalogues. *Geographical Information Systems*, 2, 677–692.
- Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, 11, 181–217. <https://doi.org/10.1002/jrsm.1378>
- Gweon, H., Li, S., & Mamon, R. (2020). An effective bias-corrected bagging method for the valuation of large variable annuity portfolios. *ASTIN Bulletin: The Journal of the IAA*, 50, 853–871. <https://doi.org/10.1017/asb.2020.28>
- Hagger, M. S. (2022). Developing an open science ‘mindset.’ *Health Psychology and Behavioral Medicine*, 10, 1–21. <https://doi.org/10.1080/21642850.2021.2012474>
- Hagras, H. (2018). Toward human-understandable, explainable AI. *Computer*, 51, 28–36. <https://doi.org/10.1109/MC.2018.3620965>
- Halevy, A., Korn, F., Noy, N. F., Olston, C., Polyzotis, N., Roy, S., & Whang, S. E. (2016). Goods: Organizing Google's datasets. *Proceedings of the 2016 International Conference on Management of Data*, 795–806. <https://doi.org/10.1145/2882903.2903730>
- Harjes, J., Link, A., Weibulat, T., Triebel, D., & Rambold, G. (2020). FAIR digital objects in environmental and life sciences should comprise workflow operation design data and method information for repeatability of study setups and reproducibility of results. *Database*, 2020. <https://doi.org/10.1093/database/baaa059>
- Harris, T. (2015). Credit scoring using the clustered support vector machine. *Expert Systems with Applications*, 42, 741–750. <https://doi.org/10.1016/j.eswa.2014.08.029>
- He, X., Liu, D., & Xiao, Y. (2022). Research on building open insurance in the insurance industry based on the concept of open banking. *7th International Conference on Social Sciences and Economic Development (ICSSSED 2022)*, 608–613. <https://doi.org/10.2991/aebmr.k.220405.100>
- Head, B. W. (2016). Toward more “evidence-informed” policy making? *Public Administration Review*, 76, 472–484. <https://doi.org/10.1111/puar.12475>

- Henckaerts, R., Côté, M.-P., Antonio, K., & Verbelen, R. (2021). Boosting insights in insurance tariff plans with tree-based machine learning methods. *North American Actuarial Journal*, 25, 255–285. <https://doi.org/10.1080/10920277.2020.1745656>
- Herland, M., Khoshgoftaar, T. M., & Bauder, R. A. (2018). Big data fraud detection using multiple medicare data sources. *Journal of Big Data*, 5, 1–21. <https://doi.org/10.1186/s40537-018-0138-3>
- Hoffman, H. (1994). *German Credit Dataset* [Dataset]. UCI Machine Learning Repository.
- Höglund, H. (2017). Tax payment default prediction using genetic algorithm-based variable selection. *Expert Systems with Applications*, 88, 368–375. <https://doi.org/10.1016/j.eswa.2017.07.027>
- Holub, P., Kohlmayer, F., Prasser, F., Mayrhofer, M. T., Schlünder, I., Martin, G. M., Casati, S., Koumakis, L., Wutte, A., & Kozera, Ł. (2018). Enhancing reuse of data and biological material in medical research: From FAIR to FAIR-health. *Biopreservation and Biobanking*, 16. <https://doi.org/10.1089/bio.2017.0110>
- Huang, D., Lin, Y., Weng, Z., & Xiong, J. (2021). Decision analysis and prediction based on credit card fraud data. *The 2nd European Symposium on Computer and Communications*, 20–26. <https://doi.org/10.1145/3478301.3478305>
- Huang, S.-H., Miao, Y.-H., & Hsiao, Y.-T. (2021). Novel deep reinforcement algorithm with adaptive sampling strategy for continuous portfolio optimization. *IEEE Access*, 9, 77371–77385. <https://doi.org/10.1109/ACCESS.2021.3082186>
- Huang, Z. (2018). Discussion on the development of artificial intelligence in taxation. *American Journal of Industrial and Business Management*, 8, 1817. <https://doi.org/10.4236/ajibm.2018.88123>
- Ileberi, E., Sun, Y., & Wang, Z. (2021). Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. *IEEE Access*, 9, 165286–165294. <https://doi.org/10.1109/ACCESS.2021.3134330>
- Jacobsen, A., De Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., & Evelo, C. T. (2020). FAIR principles: Interpretations and implementation considerations. *Data Intelligence*, 2, 10–29. https://doi.org/10.1162/dint_r_00024
- Jadhav, V., & Khang, A. (2025). Cutting-edge technologies and applications for digital banking and financial services. In *Shaping Cutting-Edge Technologies and Applications for Digital Banking and Financial Services* (p. 1). <https://doi.org/10.4324/9781003501947-1>
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29, 258–268. <https://doi.org/10.1080/10580530.2012.716740>
- Jati, P. H. P., Lin, Y., Nodehi, S., Cahyono, D. B., & Van Reisen, M. (2022). FAIR versus open data: A comparison of objectives and principles. *Data Intelligence*, 4, 867–881. https://doi.org/10.1162/dint_a_00176
- Jeliazkova, N., Apostolova, M. D., Andreoli, C., Barone, F., Barrick, A., Battistelli, C., Bossa, C., Botea-Petcu, A., Châtel, A., & De Angelis, I. (2021). Towards FAIR nanosafety data. *Nature Nanotechnology*, 16, 644–654. <https://doi.org/10.1038/s41565-021-00911-6>
- Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2023). *Is there a replication crisis in finance?* National Bureau of Economic Research.
- Jin, Y., Zhang, W., Wu, X., Liu, Y., & Hu, Z. (2021). A novel multi-stage ensemble model with a hybrid genetic algorithm for credit scoring on imbalanced data. *IEEE Access*, 9, 143593–143607. <https://doi.org/10.1109/ACCESS.2021.3120086>

- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Medicare fraud detection using neural networks. *Journal of Big Data*, 6, 1–35. <https://doi.org/10.1186/s40537-019-0225-0>
- Junyu, H. (2020). Prediction of financial crisis based on machine learning. *The 4th International Conference on Business and Information Management*, 71–75. <https://doi.org/10.1145/3418653.3418674>
- Kalendralis, P., Sloep, M., Van Soest, J., Dekker, A., & Fijten, R. (2021). Making radiotherapy more efficient with FAIR data. *Physica Medica*, 82, 158–162. <https://doi.org/10.1016/j.ejmp.2021.01.083>
- Kalid, S. N., Ng, K.-H., Tong, G.-K., & Khor, K.-C. (2020). A multiple classifiers system for anomaly detection in credit card data with unbalanced and overlapped classes. *IEEE Access*, 8, 28210–28221. <https://doi.org/10.1109/ACCESS.2020.2972009>
- Kara, A., Zhou, H., & Zhou, Y. (2021). Achieving the United Nations' sustainable development goals through financial inclusion: A systematic literature review of access to finance across the globe. *International Review of Financial Analysis*, 77, 101833. <https://doi.org/10.1016/j.irfa.2021.101833>
- Karkouch, A., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Data quality in internet of things: A state-of-the-art survey. *Journal of Network and Computer Applications*, 73, 57–81. <https://doi.org/10.1016/j.jnca.2016.08.002>
- Katz, D. S., Gruenpeter, M., & Honeyman, T. (2021). Taking a fresh look at FAIR for research software. *Patterns*, 2, 100222. <https://doi.org/10.1016/j.patter.2021.100267>
- Khodairy, M. A., & Abosamra, G. (2021). Driving behavior classification based on oversampled signals of smartphone embedded sensors using an optimized stacked-LSTM neural networks. *IEEE Access*, 9, 4957–4972. <https://doi.org/10.1109/ACCESS.2020.3048915>
- Khuong, M. N., & Tuan, T. M. (2016). A new neuro-fuzzy inference system for insurance forecasting. *International Conference on Advances in Information and Communication Technology*.
- King, D. K., & Hayes, J. (2018). The effects of power relationships: Knowledge, practice and a new form of regulatory capture. *Journal of Risk Research*, 21, 1104–1116. <https://doi.org/10.1080/13669877.2017.1382560>
- Koster, H. (2020). Towards better implementation of the European Union's anti-money laundering and countering the financing of terrorism framework. *Journal of Money Laundering Control*. <https://doi.org/10.1108/JMLC-09-2019-0073>
- Koster, O., Kosman, R., & Visser, J. A. (2021). Checklist for explainable AI in the insurance domain. In *International Conference on the Quality of Information and Communications Technology* (pp. 446–456). Springer. https://doi.org/10.1007/978-3-030-85347-1_32
- Kowolenko, M., & Vouk, M. A. (2018). Developing an open source 'big data' cognitive computing platform: Big data (Ubiquity symposium). *Ubiquity*, 2018, 1–15. <https://doi.org/10.1145/3158344>
- Kowshalya, G., & Nandhini, M. (2018). Predicting fraudulent claims in automobile insurance. *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. <https://doi.org/10.1109/ICICCT.2018.8473034>
- Lahmiri, S., & Bekiros, S. (2019). Can machine learning approaches predict corporate bankruptcy? Evidence from a qualitative experimental design. *Quantitative Finance*, 19, 1569–1577. <https://doi.org/10.1080/14697688.2019.1588468>

- Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., Van De Sandt, S., Ison, J., & Martinez, P. A. (2020). Towards FAIR principles for research software. *Data Science*, 3, 37–59. <https://doi.org/10.3233/DS-190026>
- Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., Da Silva Santos, L. O. B., & Roos, M. (2020). The “A” of FAIR – as open as possible, as closed as necessary. *Data Intelligence*, 2, 47–55. https://doi.org/10.1162/dint_a_00027
- Lee, G. Y., Manski, S., & Maiti, T. (2020). Actuarial applications of word embedding models. *ASTIN Bulletin: The Journal of the IAA*, 50, 1–24. <https://doi.org/10.1017/asb.2019.28>
- Leonelli, S. (2016). *Data-centric biology*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226416502.001.0001>
- Leonelli, S., Lovell, R., Wheeler, B. W., Fleming, L., & Williams, H. (2021). From FAIR data to fair data use: Methodological data fairness in health-related social media research. *Big Data & Society*, 8(1). <https://doi.org/10.1177/20539517211010310>
- Li, L.-H., Lin, C.-T., & Chen, S.-F. (2017). *Micro-lending default awareness using artificial neural network*. ACM Press. <https://doi.org/10.1145/3145511.3145515>
- Li, Z., Liu, X.-Y., Zheng, J., Wang, Z., Walid, A., & Guo, J. (2021). FinRL-Podracar: High performance and scalable deep reinforcement learning for quantitative finance. *Proceedings of the Second ACM International Conference on AI in Finance*, 1–9. <https://doi.org/10.1145/3490354.3494413>
- Lin, W. C., Lu, Y. H., & Tsai, C. F. (2019). Feature selection in single and ensemble learning-based bankruptcy prediction models. *Expert Systems*, 36, e12335. <https://doi.org/10.1111/exsy.12335>
- Liu, Q., Liu, Z., Zhang, H., Chen, Y., & Zhu, J. (2021). Mining cross features for financial credit risk assessment. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 1069–1078. <https://doi.org/10.1145/3459637.3482371>
- Lopez, O., & Milhaud, X. (2021). Individual reserving and nonparametric estimation of claim amounts subject to large reporting delays. *Scandinavian Actuarial Journal*, 34–53. <https://doi.org/10.1080/03461238.2020.1793218>
- Maedche, A., Elshan, E., Höhle, H., Lehrer, C., Recker, J., Sunyaev, A., Sturm, B., & Werth, O. (2024). Open science: Towards greater transparency and openness in science. *Business & Information Systems Engineering*, 1–16. <https://doi.org/10.1007/s12599-024-00858-7>
- Mahalakshmi, V., Kulkarni, N., Kumar, K. P., Kumar, K. S., Sree, D. N., & Durga, S. (2022). The role of implementing artificial intelligence and machine learning technologies in the financial services industry for creating competitive intelligence. *Materials Today: Proceedings*, 56, 2252–2255.
- Mazimwe, A., Hammouda, I., & Gidudu, A. (2021). Implementation of FAIR principles for ontologies in the disaster domain: A systematic literature review. *ISPRS International Journal of Geo-Information*, 10, 324. <https://doi.org/10.3390/ijgi10050324>
- McKinlay, A., Carter, C., & Pezet, E. (2012). *Governmentality, power and organization*. Taylor & Francis. <https://doi.org/10.1177/1744935911429414>
- Mehrban, S., Nadeem, M. W., Hussain, M., Ahmed, M. M., Hakeem, O., Saqib, S., Kiah, M. M., Abbas, F., Hassan, M., & Khan, M. A. (2020). Towards secure fintech: A survey, taxonomy, and open research challenges. *IEEE Access*, 8, 23391–23406. <https://doi.org/10.1109/ACCESS.2020.2970430>
- Meyer, A., Faverjon, C., Hostens, M., Stegeman, A., & Cameron, A. (2021). Systematic review of the status of veterinary epidemiological research in two species regarding the FAIR guiding principles. *BMC Veterinary Research*, 17, 1–14. <https://doi.org/10.1186/s12917-021-02971-1>

- Mökander, J., & Schroeder, R. (2024). Artificial intelligence, rationalization, and the limits of control in the public sector: The case of tax policy optimization. *Social Science Computer Review*, 42(6), 1359–1378. <https://doi.org/10.1177/08944393241235175>
- Mulgan, G. (2005). Government, knowledge and the business of policy making: The potential and limits of evidence-based policy. *Evidence & Policy*, 1, 215–226. <https://doi.org/10.1332/1744264053730789>
- Mushi, G. E., Pienaar, H., & Van Deventer, M. (2020). Identifying and implementing relevant research data management services for the library at the University of Dodoma, Tanzania. *Data Science Journal*, 19, 1–1. <https://doi.org/10.5334/dsj-2020-001>
- Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23, 242–266. <https://doi.org/10.2307/259373>
- Nembe, J. K., Atadoga, J. O., Mhlongo, N. Z., Falaiye, T., Olubusola, O., Daraojimba, A. I., & Oguejiofor, B. B. (2024). The role of artificial intelligence in enhancing tax compliance and financial regulation. *Finance & Accounting Research Journal*, 6, 241–251. <https://doi.org/10.51594/farj.v6i2.822>
- Neumann, Ł., Nowak, R. M., Okuniewski, R., & Wawrzyński, P. (2019). Machine learning-based predictions of customers' decisions in car insurance. *Applied Artificial Intelligence*, 33, 817–828. <https://doi.org/10.1080/08839514.2019.1630151>
- Normore, L. F., & Tebo, M. E. (2011). Assessing user requirements for a small scientific data repository. *Proceedings of the American Society for Information Science and Technology*, 48, 1–4. <https://doi.org/10.1002/meet.2011.14504801286>
- Ntoutsis, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdil, W., Vidal, M. E., Ruggieri, S., Turini, F., Papadopoulos, S., & Krasanakis, E. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10, e1356.
- Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Castignani, G., & Masello, L. (2024). Towards pay-as-you-move (PAYM) insurance: The multimodal mobility risk transfer solution. *Transportation Research Interdisciplinary Perspectives*, 28, 101283. <https://doi.org/10.1016/j.trip.2024.101283>
- Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Ressel, J., & Castignani, G. (2022). Explainable artificial intelligence (XAI) in insurance. *Risks*, 10, 230. <https://doi.org/10.3390/risks10120230>
- Ozili, P. K. (2018). Impact of digital finance on financial inclusion and stability. *Borsa Istanbul Review*, 18, 329–340. <https://doi.org/10.1016/j.bir.2017.12.003>
- Papadimitris, A. G., Klaessig, F. C., Exner, T. E., Hofer, S., Hofstaetter, N., Himly, M., Williams, M. A., Doganis, P., Hoover, M. D., & Afantitis, A. (2020). Metadata stewardship in nanosafety research: Community-driven organisation of metadata schemas to support FAIR nanoscience data. *Nanomaterials*, 10, 2033. <https://doi.org/10.3390/nano10102033>
- Paruchuri, H. (2020). The impact of machine learning on the future of insurance industry. *American Journal of Trade and Policy*, 7, 85–90. <https://doi.org/10.18034/ajtp.v7i3.537>
- Payrovnaziri, S. N., Chen, Z., Rengifo-Moreno, P., Miller, T., Bian, J., Chen, J. H., Liu, X., & He, Z. (2020). Explainable artificial intelligence models using real-world electronic health record data: A systematic scoping review. *Journal of the American Medical Informatics Association*, 27, 1173–1185. <https://doi.org/10.1093/jamia/ocaa053>
- Pellandini-Simányi, L. (2020). The financialization of everyday life. In C. Borch & R. Wosnitzer (Eds.), *The Routledge handbook of critical finance studies*. Taylor & Francis. <https://doi.org/10.4324/9781315114255-18>

- Peng, R. D., & Hicks, S. C. (2021). Reproducible research: A retrospective. *Annual Review of Public Health*, 42, 79–93. <https://doi.org/10.1146/annurev-publhealth-012420-105110>
- Pérez López, C., Delgado Rodríguez, M. J., & De Lucas Santos, S. (2019). Tax fraud detection through neural networks: An application using a sample of personal income taxpayers. *Future Internet*, 11, 86. <https://doi.org/10.3390/fi11040086>
- Prusti, D., Kumar, A., Purusottam, I. S., & Rath, S. K. (2021). A design methodology for web-based services to detect fraudulent transactions in credit card. *14th Innovations in Software Engineering Conference (Formerly Known as India Software Engineering Conference)*, 1–9. <https://doi.org/10.1145/3452383.3452397>
- Quan, Z., & Valdez, E. A. (2018). Predictive analytics of insurance claims using multivariate decision trees. *Dependence Modeling*, 6, 377–407. <https://doi.org/10.1515/demo-2018-0022>
- Rahman, S., Sirazy, M., Das, R., & Khan, R. (2024). An exploration of artificial intelligence techniques for optimizing tax compliance, fraud detection, and revenue collection in modern tax administrations. *International Journal of Business Intelligence and Big Data Analytics*, 7, 56–80.
- Raikov, A. (2021). Decreasing tax evasion by artificial intelligence. *IFAC-PapersOnLine*, 54, 172–177. <https://doi.org/10.1016/j.ifacol.2021.10.440>
- Randhawa, K., Loo, C. K., Seera, M., Lim, C. P., & Nandi, A. K. (2018). Credit card fraud detection using AdaBoost and majority voting. *IEEE Access*, 6, 14277–14284. <https://doi.org/10.1109/ACCESS.2018.2806420>
- Rashid, S. M., McCusker, J. P., Pinheiro, P., Bax, M. P., Santos, H. O., Stingone, J. A., Das, A. K., & McGuinness, D. L. (2020). The semantic data dictionary – an approach for describing and annotating data. *Data Intelligence*, 2, 443–486. https://doi.org/10.1162/dint_a_00058
- Ravi, K., Ravi, V., & Prasad, P. S. R. K. (2017). Fuzzy formal concept analysis based opinion mining for CRM in financial services. *Applied Soft Computing*, 60, 786–807. <https://doi.org/10.1016/j.asoc.2017.05.028>
- Redondo, M., Hernández-Aguado, I., & Lumbreras, B. (2018). The impact of the tax on sweetened beverages: A systematic review. *The American Journal of Clinical Nutrition*, 108, 548–563. <https://doi.org/10.1093/ajcn/nqy135>
- Riikinen, M., Saarijärvi, H., Sarlin, P., & Lähteenmäki, I. (2018). Using artificial intelligence to create value in insurance. *International Journal of Bank Marketing*. <https://doi.org/10.1108/IJBM-01-2017-0015>
- Ritchie, F. (2014). Access to sensitive data: Satisfying objectives rather than constraints. *Journal of Official Statistics*, 30, 533–545. <https://doi.org/10.2478/jos-2014-0033>
- Rodríguez-Iglesias, A., Rodríguez-González, A., Irvine, A. G., Sesma, A., Urban, M., Hammond-Kosack, K. E., & Wilkinson, M. D. (2016). Publishing FAIR data: An exemplar methodology utilizing PHI-base. *Frontiers in Plant Science*, 7, 641. <https://doi.org/10.3389/fpls.2016.00641>
- Romanos, N., Kalogerini, M., Koumoulos, E. P., Morozinis, A., Sebastiani, M., & Charitidis, C. (2019). Innovative data management in advanced characterization: Implications for materials design. *Materials Today Communications*, 20, 100541. <https://doi.org/10.1016/j.mtcomm.2019.100541>
- Rotolo, D., Camerani, R., Grassano, N., & Martin, B. R. (2022). Why do firms publish? A systematic literature review and a conceptual framework. *Research Policy*, 51, 104606. <https://doi.org/10.1016/j.respol.2022.104606>
- Sano, C. (1992). *Japanese Credit Screening Data Set* [Dataset]. UCI Machine Learning Repository.

- Sarkar, A. (2020). Disrupting the insurance value chain. In S. Chishti, I. Bartoletti, A. Leslie, & S. M. Millie (Eds.), *The AI book: The artificial intelligence handbook for investors, entrepreneurs and fintech visionaries* (pp. 89–91). Wiley.
- Schelldorfer, J., & Wuthrich, M. V. (2019). *Nesting classical actuarial models into neural networks*. SSRN 3320525. <https://doi.org/10.2139/ssrn.3320525>
- Schweitzer, B. (2024). Artificial intelligence (AI) ethics in accounting. *Journal of Accounting, Ethics & Public Policy, JAEPP*, 25(1), 67.
- Senave, E., Jans, M. J., & Srivastava, R. P. (2023). The application of text mining in accounting. *International Journal of Accounting Information Systems*, 50, 100624. <https://doi.org/10.1016/j.accinf.2023.100624>
- Shah, P., & Guez, A. (2009). Mortality forecasting using neural networks and an application to cause-specific data for insurance purposes. *Journal of Forecasting*, 28, 535–548. <https://doi.org/10.1002/for.1111>
- Shapiro, A. F. (2007). An overview of insurance uses of fuzzy logic. *Computational Intelligence in Economics and Finance*, 25–61. https://doi.org/10.1007/978-3-540-72821-4_2
- Shawat, R., Wassef, A., & Badawy, H. (2023). Artificial intelligence in financial services: Advantages and disadvantages. In R. Shawat, A. Wassef, & H. Badawy (Eds.), *Artificial Intelligence for Capital Markets*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781003327745-3>
- Singh, D., & Jain, A. (2018). A look into the artificial intelligence and its applications in various fields of life. *JournalNX*, 198–203.
- Sithic, H. L., & Balasubramanian, T. (2013). Survey of insurance fraud detection using data mining techniques. *arXiv Preprint arXiv:1309.0806*.
- Smith, J. A., & Sandbrink, J. B. (2022). Biosecurity in an age of open science. *PLoS Biology*, 20, e3001600. <https://doi.org/10.1371/journal.pbio.3001600>
- Smuha, N. A. (2025). Regulation 2024/1689 of the Eur. Parl. & Council of June 13, 2024 (EU Artificial Intelligence Act). *International Legal Materials*, 1–148.
- Smyth, G. K., & Jørgensen, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: Dispersion modelling. *ASTIN Bulletin: The Journal of the IAA*, 32, 143–157. <https://doi.org/10.2143/AST.32.1.1020>
- Soui, M., Smiti, S., Mkaouer, M. W., & Ejbal, R. (2020). Bankruptcy prediction using stacked auto-encoders. *Applied Artificial Intelligence*, 34, 80–100. <https://doi.org/10.1080/08839514.2019.1691849>
- Splitter, V., Dobusch, L., Von Krogh, G., Whittington, R., & Walgenbach, P. (2023). Openness as organizing principle: Introduction to the special issue. *Organisation Studies*, 44, 7–27. <https://doi.org/10.1177/01708406221145595>
- Sterner, B., & Elliott, S. (2023). How data governance principles influence participation in biodiversity science. *Science as Culture*, 1–26.
- Sun, X., Sun, L., & Wang, H. (2011). Extended k-anonymity models against sensitive attribute disclosure. *Computer Communications*, 34, 526–535. <https://doi.org/10.1016/j.comcom.2010.03.020>
- Suver, C., Thorogood, A., Doerr, M., Wilbanks, J., & Knoppers, B. (2020). Bringing code to data: Do not forget governance. *Journal of Medical Internet Research*, 22, e18087. <https://doi.org/10.2196/18087>
- Tanhua, T., Pouliquen, S., Hausman, J., O'Brien, K., Bricher, P., De Bruin, T., Buck, J. J., Burger, E. F., Carval, T., & Casey, K. S. (2019). Ocean FAIR data services. *Frontiers in Marine Science*, 6, 440. <https://doi.org/10.3389/fmars.2019.00440>

- Tekaya, B., Feki, S. E., Tekaya, T., & Masri, H. (2020). Recent applications of big data in finance. *Proceedings of the 2nd International Conference on Digital Tools & Uses Congress*, 1–6. <https://doi.org/10.1145/3423603.3424056>
- Tripathi, D., Edla, D. R., Cheruku, R., & Kuppili, V. (2019). A novel hybrid credit scoring model based on ensemble feature selection and multilayer ensemble classification. *Computational Intelligence*, 35, 371–394. <https://doi.org/10.1111/coin.12200>
- Tsai, C. F. (2008). Financial decision support using neural networks and support vector machines. *Expert Systems*, 25, 380–393. <https://doi.org/10.1111/j.1468-0394.2008.00449.x>
- Turiel, J., & Aste, T. (2020). Peer-to-peer loan acceptance and default prediction with artificial intelligence. *Royal Society Open Science*, 7, 191649. <https://doi.org/10.1098/rsos.191649>
- Umamaheswari, K., & Janakiraman, S. (2014). Role of data mining in insurance industry. *International Journal of Advanced Computer Technology*, 3, 961–966.
- Verma, A., Taneja, A., & Arora, A. (2017). Fraud detection and frequent pattern matching in insurance claims using data mining techniques. *Tenth International Conference on Contemporary Computing (IC3)*, 1–7. <https://doi.org/10.1109/IC3.2017.8284299>
- Vesteghem, C., Brøndum, R. F., Sønderkær, M., Sommer, M., Schmitz, A., Bødker, J. S., Dybkær, K., El-Galaly, T. C., & Bøgsted, M. (2020). Implementing the FAIR Data Principles in precision oncology: review of supporting initiatives. *Briefings in Bioinformatics*, 21, 936–945. <https://doi.org/10.1093/bib/bbz044>
- Vetter, T. R., McGwin, G., Jr., & Pittet, J.-F. (2016). Replicability, reproducibility, and fragility of research findings—ultimately, caveat emptor. *Anesthesia & Analgesia*, 123, 244–248. <https://doi.org/10.1213/ANE.0000000000001398>
- Viaene, S., Derrig, R. A., Baesens, B., & Dedene, G. (2002). A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection. *Journal of Risk and Insurance*, 69, 373–421. <https://doi.org/10.1111/1539-6975.00023>
- Viaene, S., Derrig, R. A., & Dedene, G. (2004). A case study of applying boosting Naive Bayes to claim fraud diagnosis. *IEEE Transactions on Knowledge and Data Engineering*, 16, 612–620. <https://doi.org/10.1109/TKDE.2004.1277822>
- Walmsley, J. (2021). Artificial intelligence and the value of transparency. *AI & Society*, 36, 585–595. <https://doi.org/10.1007/s00146-020-01066-z>
- Walsh, N., & Taylor, M. (2020). Cutting to the Chase: Mapping AI to the real-world insurance value chain. In S. Chishti, I. Bartoletti, A. Leslie, & S. M. Millie (Eds.), *The AI Book: The artificial intelligence handbook for investors, entrepreneurs and fintech visionaries* (pp. 92–97). Wiley.
- Wang, M., Jian, P., & Li, Y. (2019). Research on financial network loan risk control model based on prior rule and machine learning algorithm. *Proceedings of the 4th International Conference on Mathematics and Artificial Intelligence*, 76–79. <https://doi.org/10.1145/3325730.3325733>
- Wei, C., & Dan, L. (2019). Market fluctuation and agricultural insurance forecasting model based on machine learning algorithm of parameter optimization. *Journal of Intelligent & Fuzzy Systems*, 37, 6217–6228. <https://doi.org/10.3233/JIFS-179204>
- Wei, S., Yang, D., Zhang, W., & Zhang, S. (2019). A novel noise-adapted two-layer ensemble model for credit scoring based on backflow learning. *IEEE Access*, 7, 99217–99230. <https://doi.org/10.1109/ACCESS.2019.2930332>

- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., & Bourne, P. E. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 1–9. <https://doi.org/10.1038/sdata.2016.18>
- Wise, J., De Barron, A. G., Splendiani, A., Balali-Mood, B., Vasant, D., Little, E., Mellino, G., Harrow, I., Smith, I., & Taubert, J. (2019). Implementation and relevance of FAIR data principles in biopharmaceutical R&D. *Drug Discovery Today*, 24, 933–938. <https://doi.org/10.1016/j.drudis.2019.01.008>
- Wolf, M., Logan, J., Mehta, K., Jacobson, D., Cashman, M., Walker, A. M., Eisenhauer, G., Widener, P., & Cliff, A. (2021). Reusability first: Toward FAIR workflows. *IEEE International Conference on Cluster Computing (CLUSTER)*, 444–455. <https://doi.org/10.1109/Cluster48925.2021.00053>
- Wüthrich, M. V. (2018). Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, 2018, 465–480. <https://doi.org/10.1080/03461238.2018.1428681>
- Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Natural language based financial forecasting: A survey. *Artificial Intelligence Review*, 50, 49–73. <https://doi.org/10.1007/s10462-017-9588-9>
- Xu, C. (2021). Image-based candlestick pattern classification with machine learning. *6th International Conference on Machine Learning Technologies*, 26–33. <https://doi.org/10.1145/3468891.3468896>
- Xu, R., Meng, H., Lin, Z., Xu, Y., Cui, L., & Lin, J. (2021). Credit default prediction via explainable ensemble learning. *5th International Conference on Crowd Science and Engineering*, 81–87. <https://doi.org/10.1145/3503181.3503195>
- Xu, W., Wang, S., Zhang, D., & Yang, B. (2011). Random rough subspace based neural network ensemble for insurance fraud detection. *Fourth International Joint Conference on Computational Sciences and Optimization*, 1276–1280. <https://doi.org/10.1109/CSO.2011.213>
- Xu, X., Chen, X., Jia, F., Brown, S., Gong, Y., & Xu, Y. (2018). Supply chain finance: A systematic literature review and bibliometric analysis. *International Journal of Production Economics*, 204, 160–173. <https://doi.org/10.1016/j.ijpe.2018.08.003>
- Ye, Z. (2021). A data slicing method to improve machine learning model accuracy in bankruptcy prediction. *Proceedings of the 2021 5th International Conference on Deep Learning Technologies*, 32–39. <https://doi.org/10.1145/3480001.3480008>
- Yu, H., & Robinson, D. G. (2011). The new ambiguity of open government. *UCLA L. Rev. Discourse*, 59, 178.
- Yuan, M. (2022). A transformer-based model integrated with feature selection for credit card fraud detection. *7th International Conference on Machine Learning Technologies (ICMLT)*, 185–190. <https://doi.org/10.1145/3529399.3529429>
- Zhang, Y., & Lu, S. (2021). Multi-model fusion method and its application in prediction of stock index movements. *6th International Conference on Machine Learning Technologies*, 58–64. <https://doi.org/10.1145/3468891.3468900>
- Zheng, S., Trott, A., Srinivasa, S., Parkes, D. C., & Socher, R. (2022). The AI economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Science Advances*, 8(18). <https://doi.org/10.1126/sciadv.abk2607>
- Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Healey, S. P., Schaaf, C., Hostert, P., & Strobl, P. (2019). Benefits of the free and open Landsat data policy. *Remote Sensing of Environment*, 224, 382–385. <https://doi.org/10.1016/j.rse.2019.02.016>

Zou, H. (2021). Analysis of best sampling strategy in credit card fraud detection using machine learning. *6th International Conference on Intelligent Information Technology*, 40–44. <https://doi.org/10.1145/3460179.3460186>