ISSN: 3033-4136

DOI: https://doi.org/10.17868/strath.00089572

Simplifying Compliance Through Explainable Intelligent Automation

FRIL

Financial Regulation Innovation Lab

FinTech Scotland®

University of Glasgow

University of Strathclyde Glasgow

Financial Regulation Innovation Lab

Shaping the future of financial regulation

Who are we?

The Financial Regulation Innovation Lab (FRIL) is an industry-led collaborative research and innovation programme focused on leveraging new technologies to respond to, shape, and help evolve the future regulatory landscape in the UK and globally, helping to create new employment and business opportunities, and enabling the future talent.

FRIL provides an environment for participants to engage and collaborate on the dynamic demands of financial regulation, explore, test and experiment with new technologies, build confidence in solutions and demonstrate their ability to meet regulatory standards worldwide.

What is Actionable Research?

FRIL will integrate academic research with an industry relevant agenda, focused on enabling knowledge on cutting-edge topics such as generative and explainable AI, advanced analytics, advanced computing, and earth-intelligent data as applied to financial regulation. The approach fosters cross sector learning to produce a series of papers, actionable recommendations and strategic plans that can be tested in the innovation environment, in collaboration across industry and regulators.

Locally-led Innovation Accelerators delivered in partnership with DSIT, Innovate UK and City Regions

Department for Science, Innovation & Technology

UKRI Innovate UK

GLASGOW CITY REGION

FRIL White Paper Series

Simplifying Compliance through Explainable Intelligent Automation

James Bowden* Mark Cummins* Daniel Dao* Kushagra Jain*

** University of Strathclyde*

March 2024

Abstract: We discuss how explainability in AI-systems can deliver transparency and build trust towards greater adoption of automation to support financial regulation compliance among banks and financial services firms. We uniquely propose the concept of *Explainable Intelligent Automation* as the next generation of Intelligent Automation. Explainable Intelligent Automation seeks to leverage emerging innovations in the area of Explainable Artificial Intelligence. AI systems underlying Intelligent Automation bring considerable advantages to the task of automating compliance processes. A barrier to AI adoption though is the black-box nature of the machine learning techniques delivering the outcomes, which is exacerbated by the pursuit of increasingly complex frameworks, such as deep learning, in the delivery of performance accuracy. Through articulating the business value of Robotic Process Automation and Intelligent Automation, we consider the potential for Explainable Intelligent Automation to add value. The solution framework sets out the Explainable Intelligent Automation framework, as the interface of Robotic Process Automation, Business Process Management and Explainable Artificial Intelligence. We discuss key considerations of an organisation in terms of setting strategic priorities around the explainability of AI systems, the technical considerations in Explainable Artificial Intelligence analytics, and the imperative to evaluate explanations.

Strategic Alignment: *FinTech Research & Innovation Roadmap 2021-31; Kalifa Review of UK FinTech; EU's Corporate Sustainability Reporting Directive, subject to European Sustainability Reporting Standards; EU's Sustainable Finance Disclosure Regulation; UK's Task Force on Climate Related Financial Disclosures (FCA); UK's Climate-Related Financial Disclosure (Department for Energy Security and Net-Zero); UK Sustainability Disclosure Standards.*

FinTech Research & Innovation Roadmap 2021-31 Sub-Theme: Simplifying Compliance

Table of Contents

1. Problem Statement

A priority theme in the digital transformation of business is that of automation. The 2022 Deloitte Insights Automation with Intelligence survey[1] notes a number of industry trends in respect of Robotic Process Automation and, in the next phase, Intelligent Automation. The research of Deloitte notes:

- Continued progression of firms along the automation maturity curve;
- The need to move away from task-based automation towards end-to-end automation;
- The benefits of insight-driven transformation gained from process intelligence approaches;
- The use of Automation-as-a-Service as a delivery mode of automation solutions;
- The emergence of citizen-led development as a human-computer framework that enables users to create new task-based automations for their own use, which helps to break the misconception of automation replacing humans.

Automation is noted by the Deloitte analysis as offering significant commercial benefits in the form of increased productivity, cost reduction, improved accuracy, and better customer experience. In the automation space, we are observing a gradual move from Robotic Process Automation (RPA) to Intelligent Automation (IA). RPA is well-established as involving the deployment of technology to automate routine tasks that typically are done by employees of organisations. Towards smarter end-to-end automation and intelligence-based approaches to automation, as called for by the above Deloitte analysis, IA seeks to leverage advanced, sophisticated artificial intelligence capability. IA is defined by IBM[1] as bringing together the domains of Robotic Process Automation, Business Process Management, and Artificial Intelligence (Figure 1). The augmentation of RPA with AI capability offers significant advantages in allowing for complex business processes and procedures that leverage large volumes of data to support decision making.

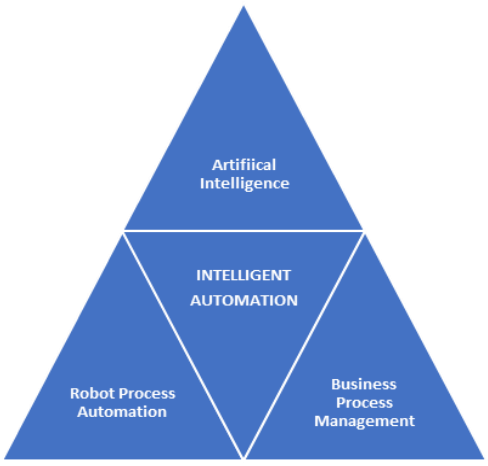


Figure 1: Intelligent Automation Framework

[1] https://www.ibm.com/cloud/learn/intelligent-automation.

There are, however, various barriers to adoption of AI. The 2022 Deloitte Insights Automation with Intelligence survey identifies barriers to innovation adoption. Specifically, the study emphasises the barriers to automation adoption as being:

- Process fragmentation
- Lack of a clear vision
- Lack of IT readiness
- Resistance to change.

In respect of IA, the latter barrier often manifests as distrust within the organisation around the adoption of AI systems. Trust is a key behavioural barrier to the adoption of innovation. AI systems create trust issues for users due to the black-box nature of the underlying AI algorithms. We seek to tackle this problem and demonstrate how trust can be engendered through providing explainability to the outcomes of AI systems underlying IA. We argue that emerging Explainable AI techniques can deliver greater transparency into AI-based automation. Indeed, we propose a new concept, *Explainable Intelligent Automation*, as a next phase of IA.

For banking and financial institutions, automation is currently playing a role in supporting financial regulation compliance processes. In the context of regulatory reporting, Deloitte in 2017[2]articulated where automation may provide the greatest value to an organisation:

- Optimisation of data extraction that would otherwise be performed manually
- Standardisation of data aggregation
- Enhancing regulatory report capabilities
- Streamlining and enhancing data quality and data lineage documentation
- Development of regulatory report review and analysis capabilities

Against this backdrop of regulatory reporting use cases, Explainable Intelligent Automation has the potential to deliver transparency and explainability around the use of AI systems to support the above automation benefits.

[2] https://www2.deloitte.com/content/dam/Deloitte/us/Documents/regulatory/us-regulatoryhttps://www2.deloitte.com/content/dam/Deloitte/us/Documents/regulatory/us-regulatory-automating-regulatory-reporting-banking-securities.pdfautomating-regulatory-reporting-banking-securities.pdf

2. Literature Review

2.1 The business utility of Robotic Process Automation and Intelligent Automation

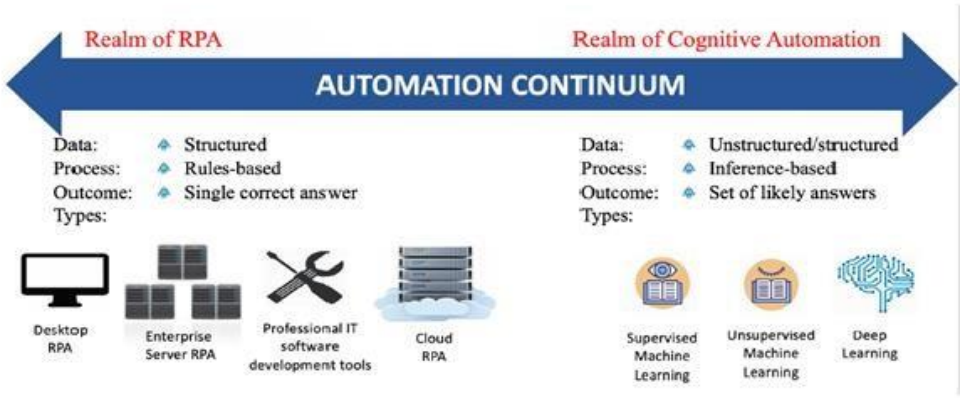
Bot utilization is rapidly increasing in business in general. Though several entities have begun embracing RPA, there is an insufficiency of knowledge in choosing suited processes for automation. Technological advances, coupled with cost cutting business automation, have led to a considerable rise in RPA use. The global RPA software market is assessed at a $1.89 billion value as of 2021, an increase of 118% since 2018. Furthermore, large companies are forecasted to triple the capacity of their existing RPA portfolios by 2024 (Eulerich et al., 2022).

In an audit context, RPA has been shown to increase task efficiency and effectiveness (developed bots saved time net of their creation time and eliminated human errors). However, to prevent failure to meet expectations from emerging audit RPA, more guidance was needed, and simply reusing general or other industry RPA guidance was unlikely to be optimal (Eulerich et al., 2022). Cooper et al. (2019) similarly report on opportunities, and challenges to implementing RPA in accounting. They find sizable efficiency and effectiveness gains from RPA implementation, with highest adoption in tax services, followed by advisory and assurance services. However, they also highlight concerns around future fee reductions sought by clients owing to decreased employee hours. Other fields such as travel, tourism, supply chain management, etc. are also expected to be significantly impacted in their way of operation by intelligent automation and responsible AI (Behl et al., 2023; Rydzik and Kissoon, 2022; Tussyadiah, 2020). Moreover, the contemporary growth of AI capability and scope is likely to continue to expand with language translation, truck driving, retail work, surgery, office, administrative and service work envisioned to be automated significantly (Coombs et al., 2020). Such breakthroughs in the so-called ``fourth industrial revolution'' are envisaged to impact value creation and distribution, forever altering work, interactions and living through automation. Such automation is thus a key ingredient of the digital transformation occurring in many sectors. Rather than direct human labour substitution by machines, such automation is machine integration into self-governing systems (Tussyadiah, 2020).

From a management information systems perspective, Lacity and Wilcocks (2021) present exhaustive evidence on intelligence automation, RPA, and Cognitive Automation (CA), highlighting both their successes and failures in achieving business value. To arrive at their conclusions, they review hundreds of intelligent automation implementations across geographies, industries, and processes across six years. They also note larger digital transformation programs are more and more integrated with intelligent automation programs, with several entities aiming to automate processes across firm boundaries.

More specifically, Lacity and Wilcocks (2021) note a paradigm shift from 2014 onwards, based on which they identify a continuum of automation as seen in Figure 2. From an RPA/CA historical perspective, they note the first use of the term RPA in 2012 by Phil Fersht, founder of an outsourcing consulting firm Horses for Sources (HfS), in a report ``Greetings from Robotistan, Outsourcing's Cheapest New Destination''. It highlighted Blue Prism, a UK start-up incorporated in 2001. Blue Prism came into the limelight when Patrick Geary, its chief marketing officer, began terming its product ``RPA'' in 2012. This term resonated with practitioners and other automation companies rebranded their tools with the same label. By 2016, over two dozen companies indicated they provided RPA tools, with a claimed market

size of $600 million. Owing to this rapid growth, a need for RPA standards arose, and Lee Coulter, then CEO of Ascension Shared Services, began an IEEE initiative for the same, and in December 2016 became chairman of the IEEE Working Group on Standards in Intelligent Process Automation. The group published the first standard in 2017, demarcating enterprise RPA (developed for organizations) and robotic desktop automation (RDA) (intended for single desktop use).



Source: Lacity and Wilcocks (2021)

Figure 2: Automation Continuum

As of 2020, Lacity and Wilcocks (2021) noted the RPA market's value was estimated between $2-4 billion, based on various consulting reports. They observed a consensus among most sources' forecasts on its yearly growth rate from 30% to 50% in the foreseeable future. C-suite priorities for such emerging technologies were seen to change considerably due to the pandemic, which elicited a sharp emphasis on fast Return on Investment (ROI) generating technologies such as process automation. Many claims and foresee mass unemployment, job eliminations from such automation developments first through predictable, repetitive work, and eventually from AI outperformance relative to humans in many activities (Coombs et al., 2020).

However, Lacity and Wilcocks (2021) deduce that many falsely assume automation ROI arises from firing employees. The primary value addition from service automation is undoubtedly freeing up human labour, this is more accurately viewed as ``hours back to the business'' (hours taken if humans still performed automated tasks representing human capacity freed for different work). Most cases they investigate use freed-up labour capacity for people redeployment to other tasks within the work unit. Such entities were able to take on more work without hiring proportionally more workers. It is evidently more valuable to grow efficiently by redeploying existing employees rather than searching for, vetting, onboarding, and training new ones. To illustrate, if 800,000 is the hours back to the business for a company, dividing it by 2,000 (the average annual number of employee work hours) gives us a value of 400 ``Full Time Equivalents (FTEs)''. This does not mean 400 employees are no longer needed, but more likely that 20% of 2,000 people's jobs have been automated, most often repetitive work.

The usual outcome is such partial task automation, rather than pure job losses, which is often offset by the extra work taken on by businesses, assessed to be between 8 to 12% annually. Such work comes from exponentially rising data volumes and regulatory needs, among other causes. Further, skill shortages, backlogs are also offset through automation, with most companies seeking employee retraining, complex-task assignment, or early retirements rather than layoffs. Moreover, while automation undoubtedly results in job losses, evidence indicates these are compensated for by skill development and net positive new job creations and changes. Finally, HR involvement is thought to be essential when embracing automation as productivity may seem to drop on account of more complex tasks being assigned to humans, which take longer to execute (Lacity and Wilcocks, 2021).

Lastly, the business value and utility of BPM has also been studied extensively, and numerous demonstrable instances of this can be highlighted. Owing to the similarity of BPM with the other automation and AI concepts discussed and given that it is a relatively older concept, it is not expanded upon extensively here. Interested readers are directed to Mendling et al. (2020) for a comprehensive overview.

2.2. Motivation for Explainable Intelligent Automation

2.2.1 Robotic Process Automation Context

RPA is both a standalone and combinable technology. For example, both local and cloud-based RPA usage is possible. However, there is limited insight into how interactions occur when RPA is combined with other technologies. RPA may thus be useful to investigate how multiple technologies interact to influence organizations, people, tasks, and structures (Eulerich et al., 2022), perhaps through Explainable Intelligent Automation. Similarly, RPA's implementation flexibility ranging from low-/no-code to high-code solutions can balance usage ease versus user task precision (Eulerich et al., 2022), and it may be possible to study this trade-off with Explainable Intelligent Automation. Studying the effects of both types of RPA this way might help assess the importance of flexibility, usage ease and other system acceptance and use principles (Eulerich et al., 2022). Finally, a current RPA limitation is only being able to perform rules-based tasks; however, as AI progresses, RPA may be able to perform more complex tasks requiring judgment (Eulerich et al., 2022), possibly with combined AI and RPA usage that can be understood with an explainability layer.

2.2.2 Cognitive Automation Context

Likewise, companies are challenged in finding alternate use for specifically designed CA tools. Largely due to data challenges, early adopters experience expensive and painful implementations. Firm case studies reveal CA tool adoptions employ supervised machine learning algorithms. These require thousands of labelled training examples for acceptable proficiency levels. Given 80% of corporate data is “dark,” i.e., untagged, untapped, or unlocatable, tool adopters first must create new data, and clean up dirty (inconsistent, incorrect, outdated, duplicated, or missing) data. ``Difficult data'' - hard for a machine to read but valid and accurate (e.g., sophisticated natural language text, unexpected data types and fuzzy images) - is another significant challenge. Laborious human intervention was required in these circumstances to sort out these data problems (Lacity and Wilcocks,

2021). With judicious deployment of Explainable Intelligent Automation, it may be possible to ameliorate these difficulties.

2.2.3 Intelligent Automation Context

CA tools usage in conjunction with RPA software is gaining traction - serving as an execution engine, especially in banking, insurance, and financial services organizations. For example, a bank, may deploy an interactive front-end chat bot for customer dialogues, but draw upon RPA for ensuring conversational accuracy, say if the topic is a stolen credit card. Even better CA and RPA integration is augured, giving rise to an increase in cloud-based automation exchange platforms (Lacity and Wilcocks, 2021). With such platforms, it may be useful to monitor and understand the decision making involved in such intelligent automation through the lens of explainable artificial intelligence, clearly indicating what variables drive the decision making of the CA tools and how they subsequently engage the relevant RPA tool.

Smart organizations now look to assimilate (intelligent) automation into grander digital transformations, and RPA and CA are fundamental to these. However, this is increasingly difficult in such applications, and a long-term, complex, large-scale process in any sizeable long-standing organization. Scaling automation is another crucial challenge. By 2019's end, just 13% of companies had RPA deployments which were industrialized and scaled, and only 12% had an enterprise automation approach without much change by late 2020. While top software providers have several customers, very few customers deploy over 100 software robots. This may be partly as the next stage's cost looks steep, as suggested earlier in this paragraph, though evidence suggests exponential benefits. Problems arise in integrating RPA with existing/new IT and are exacerbated when considered across the enterprise. Preexisting process fragmentation is vicissitudinous. RPA deployment concerns are worsened where executives do not see strategic value, are too far removed from the programs, or underinvest. CA deployment faces issues of an even greater magnitude. Progress has been slow and challenging to date (Lacity and Wilcocks, 2021). Once more, Explainable Intelligent Automation may pave the way to resolving these systemic issues, possibly reducing complexity through explainability, and providing explanations on how this might be achieved. Thereby, it may be possible to enable access to the sizable advantages promised by scaled, integrated enterprise-wide intelligent automation.

2.2.4 Business Process Management Context

Contemporary research suggests digital innovation may benefit from business process management (BPM), perhaps the most prominent management practice to improve operational efficiency. Digital innovation is catalysing useful change in work for the modern world, and BPM can speed this process up even further, with several instances of the impact of both shown to revolutionize several walks of life (Mendling et al., 2020). Unfortunately, research on digital innovation and BPM has been conducted separately under orthogonal assumptions thus far (Mendling et al., 2020). Once again, the synthesis of both these concepts might perhaps be facilitated with the aid of Explainable Intelligent Automation.

3. Solution Framework

3.1 Explainable Intelligent Automation Framework

As articulated in the problem statement, trust is a key behavioural barrier to the adoption of automation-based innovation. AI systems are known to create trust issues for users, particularly in commercial settings, due to the black box nature of the underlying AI algorithms. However, trust may be engendered through providing explainability to the outcomes of AI systems. Cutting-edge Explainable AI techniques offer the potential to deliver transparency and build trust in IA.

Explainability is vital to ensuring that automation systems are doing what they are expected to, and that outcomes can be explained via a transparent and trustworthy evidence base throughout an organisation's management structure. Trustworthiness in automation systems inspires confidence in individuals and organisations that they can benefit from, and rely on, the efficiencies that automation delivers.

This is the basis of the Explainable Intelligent Automation (EIA) concept that we propose. We define Explainable Intelligent Automation as the convergence of Robotic Process Automation, Business Process Management, and Explainable Artificial Intelligence (Figure 3).

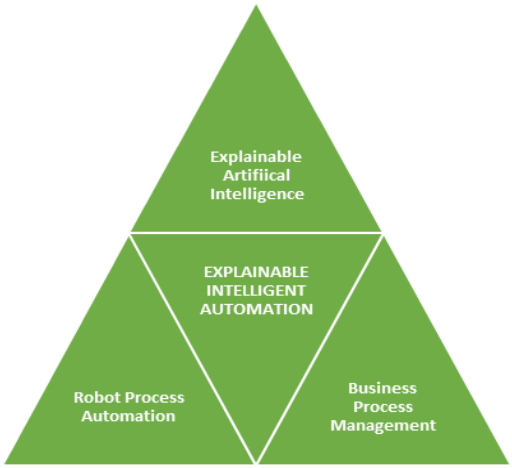


Figure 3: Explainable Intelligent Automation Framework

Depending on the maturity of a firm's digital transformation programme, we envisage a staged process in the automation adoption journey. Firms are required to phase their transition, at an appropriate pace, from Robotic Process Automation to Intelligent Automation to Explainable Intelligent Automation (Figure 4).

Figure 4: Staged Automation Adoption

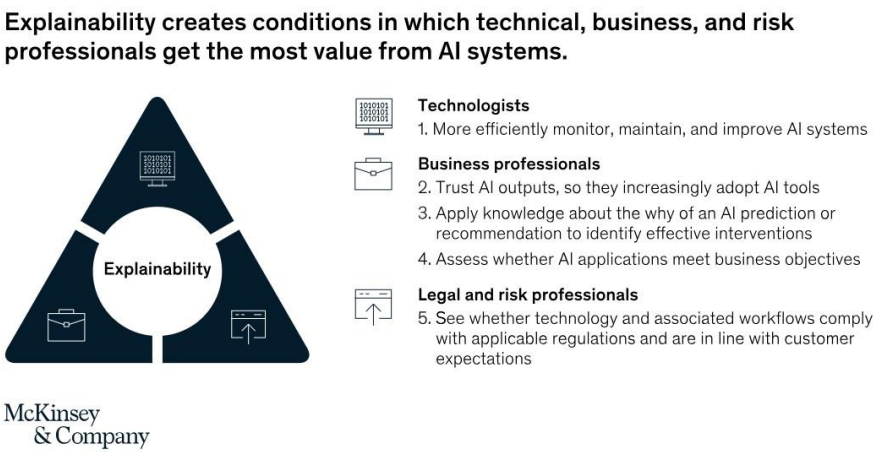
In the forthcoming sections, we explore the concept of Explainable AI as the key innovation that underlies Explainable Intelligent Automation design. We outline (i) the strategic imperative that needs to be placed on explainability in the deployment of Explainable Intelligent Automation, (ii) approaches to explanation generation and (iii) approaches to explanation evaluation.

3.2 Corporate Strategy

In considering XAI integration into AI system deployment within a financial services firm, it is necessary to consider the importance of explainability strategically and to connect this explicitly with the firm's overall digital strategy. Grennan et al. (2022), in a McKinsey article outline the business case for explainable AI. In particular, the following benefits are identified:

- Increased productivity through better monitoring, maintenance, and enhancement of AI systems;
- Building trust and adoption rates among key stakeholders through the transparency that explanations provide;
- Identifying new value creation opportunities from the insights that explanations provide;
- Articulating the business value of AI systems through explanations that connect investment to outcomes more closely.
- Better risk mitigation and regulatory compliance outcomes afforded by AI system explanations.

Placing strategic importance on the explainability of AI systems has the potential to impact various key users across an organisation. Figure 5 from Grennan et al. (2022) summarises this impact for several professional roles – technologists, business professionals and legal and risk professionals. It can be seen that XAI can benefit users through delivering efficiencies, building trust, facilitating human-in-the-loop interventions, aligning with business objectives and complying with regulations. This latter point is extremely important in the context of, on the one hand, using AI towards simplifying compliance, and, on the other hand, complying with regulation pertaining to AI systems usage within financial services organisations.



Source: Grennan et al. (2022) [McKinsey]

Figure 5: Impact of Explainability on AI System Users

3.3 Approaches to Explanation Generation

With strategy and governance structures in place, the organisation needs to then focus on engineering explainability into AI systems through the choice of specific XAI approaches. This choice may depend on the nature of the problem space and the materiality attached to this. We provide an overview of the main considerations in respect of XAI techniques.

What is an explanation, and what are its properties? **(Molnar, 2020)**

An explanation usually relates feature values of an instance to its model prediction in a humanly understandable way. To explain an ML model's predictions, some explanation method is relied on, such as an algorithm that generates explanations. Other explanation types consist of a set of data instances (e.g., for the k-nearest neighbour model). For example, a support vector machine can be used to predict cancer risk, and explain predictions with the local surrogate method, that generates decision trees as explanations. Alternately, a linear regression model may be used that is already equipped with an explanation method (interpreting weights). Certain properties have been identified for explanation methods, and explanations. These may be used to assess how good they are. It is unclear how these properties may be measured correctly, so formalizing how they could be calculated is a vicissitude. (Molnar, 2020).

Properties of Explanation Methods (Molnar, 2020):

- Expressive Power - "Language" or structure of explanations the method generates. An explanation method may generate natural language, a weighted sum, decision trees, IFTHEN rules, or something else (Molnar, 2020)

- Translucency - Describes the extent of reliance on the explanation method to look into the ML model, like its parameters. E.g., Intrinsically interpretable models like the linear regression model (model-specific) with explanations reliant on them are highly translucent. Conversely, methods solely dependent on manipulating inputs and observing predictions have zero translucency. Different scenario-dependent translucency levels may be desirable. High translucency methods can rely on more information to generate explanations.

Meanwhile, low translucency explanation methods are more portable (Molnar, 2020)

- Portability - Describes how many ML models with which the explanation method may be used. Low translucency methods have higher portability: they treat ML models as black boxes. Surrogate models may be the explanation method with highest portability. Model specific methods (only work for that model e.g., recurrent neural networks) have low portability (Molnar, 2020).

- Algorithmic Complexity - Describes computational complexity of the explanation generating method. Important when computation time bottlenecks generating explanations (Molnar, 2020).

Properties of Individual Explanations (Molnar, 2020):

- Accuracy: How well is unseen data predicted? High accuracy is particularly valuable if the explanation is used for predictions in place of the ML model. Low accuracy may be acceptable if the ML model's accuracy is also low, and if the goal is to explain what the black box model does. In this case, only fidelity is important (Molnar, 2020).

- Fidelity: How well is the black-box model's prediction approximated? High fidelity is one of the most important explanation properties, as low fidelity explanations have no value in explaining ML models. Accuracy and fidelity are closely related. If the black box model has high accuracy its explanation also usually has high fidelity and accuracy. Some explanations only offer local fidelity, i.e., explanation only approximates well to model prediction for a data subset (e.g., local surrogate models) or individual data instance (e.g., local Shapley Values) (Molnar, 2020).

- Consistency: Differences between models trained on the same task and producing similar predictions? E.g., assume a support vector machine and a linear regression model are trained on the same task and both produce very similar predictions. Using a method of choice, if the explanations are very similar, they are highly consistent. This is somewhat subtle, as two models may use different features, with similar predictions (also called "Rashomon Effect"). A high consistency is undesirable here as the explanations must be very different. High consistency is desirable if models really rely on similar relationships (Molnar, 2020).

- Stability: Similarity across similar instances. Stability juxtaposes explanations between similar instances for a model, whereas consistency contrasts explanations between models. High stability means slight variations in an instance's features do not substantially change the explanation (unless these slight variations also strongly change the prediction). Instability may be due to high variance of the explanation method. Put differently, strong effects on explanations are seen from slight changes to feature values of the instance to be explained. Non-deterministic components of the explanation method may also drive instability, like a data sampling step, which the local surrogate method uses. High stability is always desirable (Molnar, 2020).

- Comprehensibility: How well do humans understand? While seemingly like the other properties, this one is particularly important. It is difficult to measure and define, but very crucial to get right. Comprehensibility is broadly agreed to depend on the audience. Measurement ideas include measuring the explanation size (number of features with non-zero weights in a linear model, number of decision rules, etc.) or testing how well people predict ML model behaviour from explanations. Comprehensibility of features used in explanations also should be considered. Complex feature

transformations may be less comprehensible than the originals (Molnar, 2020).

- Certainty: Is the certainty of the ML model reflected? Many ML models only predict without stating the confidence of correct predictions. If a 4% cancer probability is predicted for a patient, is it as certain as the 4% probability another patient, with different feature values, received? Explanation incorporating model certainty is very useful (Molnar, 2020).

- Degree of Importance: How well is importance of features or parts of the explanation reflected? If a decision rule explanation for instance is generated for an individual prediction, is it clear which rule conditions were the most important (Molnar, 2020)?

- Novelty: Is it evident if a data instance to be explained is sampled from a "new" region, far removed from the training data's distribution? If not, the model may be inaccurate and explanation useless. Novelty is conceptually related to certainty. Higher novelty, Implied higher likelihood of low model certainty due to lack of data (Molnar, 2020).

- Representativeness: How many instances are covered? Explanations may cover an entire model (e.g., linear regression model weights interpretation) or represent individual predictions (e.g., local Shapley Values) (Molnar, 2020).

***What are good or human-friendly explanations?* (Molnar, 2020)**

There can be far-reaching consequences for interpretable machine learning based on "good" explanations, as defined by humans. Concise and single (or at most double) cause explanations which juxtapose the treatment group with a counterfactual group are preferred by humans. Good explanations are provided particularly by abnormal causes. Explanations are also "social interactions between the explainer and explanation recipient", where a human being or a machine is the explainer. This implies the actual content of the explanation is significantly impacted by the social context. Alternately, they may refer to "the social and cognitive process of explaining, but also to the product of these processes". Furthermore, a careful distinction needs to be made when comparing explanations that are "human-friendly", and complete causal attribution, where all factors for a particular prediction or behaviour need explaining. The latter may be preferred in legal contexts, where one is mandated to debug an ML model or indicate all influencing sources (Molnar, 2020).

Conversely, where non-experts or time-starved individuals are the explanation's target audience, an alternative definition applies, which defines an explanation as "the answer to a why-question", which can be answered with an "everyday"-explanation. Instances of such questions may include why a loan was rejected, or why a treatment lacked efficacy for a patient. Such "why" questions can usually be reformulated as questions beginning with "how" as well (Molnar, 2020).

A deeper dive into what constitutes a "good" explanation yields certain criteria which have definite implications for interpretable ML. These can be listed as follows - for more detail on these and their implications with examples, interested readers are referred to (Molnar, 2020):

- Contrastive – Answers why this prediction was made *instead of another prediction*. The implication is a requirement for application-dependent explanations because a point of reference for comparison is needed. This may depend on the data point to be explained, but also on the user receiving the explanation. The solution for automated creation of contrastive explanations might also involve

finding proto/archetypes in the data (Molnar, 2020).

- Selected - Select one or two causes from various possible causes as "THE" explanation, rather than covering an actual complete list of event causes (Molnar, 2020).

  This implies a preference for brevity in explanation with 1-3 reasons, even if reality is more complex (Molnar, 2020).

- Social - Part of a conversation or interaction between the explainer and explanation receiver.

  The implication is that attention to the social environment and intended recipients for explanations is needed. Getting this right may depend entirely on the specific application (Molnar, 2020).

- Focus on the abnormal - People focus more on abnormal causes in any sense (like a rare category of a categorical feature) to explain events, that had a small probability but nevertheless happened, without which the outcome would have greatly changed (counterfactual explanation) (Molnar, 2020).

  If an input feature for a prediction is abnormal, and it influenced the latter, it should be included in an explanation, even if other 'normal' features have the same influence (Molnar, 2020).

- Truthful - Prove to be true in reality (i.e., in other situations), but selectiveness seems more important, which is troubling (Molnar, 2020).

  This implies events should be predicted as truthfully as possible (also called fidelity), with less relative importance given to it than contrast, social aspect, and selectivity (Molnar, 2020).

- Consistent with explainee's prior beliefs - Humans tend to devalue or ignore information inconsistent or in disagreement with prior beliefs, also called confirmation bias. Thus, this bias logically also extends to explanations (Molnar, 2020).
  This implies using specific ways to deal with inconsistent explanations, although difficult to integrate into ML, and may come at a heavy cost to predictive performance (Molnar, 2020).

- General and probable - A cause that can explain many events is very general and could be considered a good explanation. Although this contradicts the claim that abnormal causes make good explanations, as a rule of thumb, abnormal causes trump general causes, and in the absence of the former, the latter comes to the fore (Molnar, 2020).

  Implies measurement of generality should happen, which is easily achieved by the feature's support: the number of instances to which the explanation applies, divided by the total number of instances (Molnar, 2020).

Methods for machine learning interpretability can be classified according to various criteria (Molnar, 2020):

Intrinsic or post hoc: Criterion distinguishes based on how interpretability is achieved by restricting the model complexity (intrinsic) or analysing the model by applying methods after training (post hoc). Intrinsic interpretability describes models deemed interpretable owing to their simplicity, e.g., sparse linear models or short decision trees. Post hoc interpretability implies interpretability methods applied after model training, e.g., permutation feature importance. Post hoc methods may also be applied to intrinsically interpretability models, like computing permutation feature importance for decision trees (Molnar, 2020).

Model-specific or model-agnostic: Interpretability tools confined to specific model classes are considered model-specific. Linear regression model weights are interpreted this way, as their intrinsic

interpretation is always model-specific. Similarly, tailored tools for interpreting machine learning models such as neural networks are also considered model specific. In contrast, model-agnostic interpretability tools may be deployed on any model and are used post hoc, after model training. Generally, such agnostic methods function through feature input and output pairs' analysis. These methods cannot access model internals like weights or structural information (Molnar, 2020).

Scope of interpretability: Each algorithmic step in training a predictive model can be evaluated in terms of transparency and interpretability (Molnar, 2020):

- **Algorithm Transparency:** *Assesses how an algorithm creates the model.* This relates to how an algorithm learns a model from data and the relation types it is capable of learning. Using convolutional neural networks to classify images, one may explain the learning of edge detectors and filters on the lowest layers by the algorithm. This is comprehension of how the algorithm works, but not the specific model that learned in the end, and the individual prediction process. Such transparency only requires algorithmic knowledge rather than knowing data or the learned model. Algorithms like the least squares method are well studied and understood. They characterize high transparency. Deep learning approaches (pushing a gradient through a network with millions of weights) are in contrast less well understood. Research is ongoing on their inner workings and are thus opaquer (Molnar, 2020).

- **Global, Holistic Model Interpretability:** *This distinction focuses on how the trained model makes predictions.* A model may be called interpretable if it can be comprehended entirely at once. To explain the global model output, knowing the trained model, algorithm and data are prerequisites. This interpretability level considers how the model decisions are made, from a holistic features' view, and each learned component e.g., weights, other parameters, and structures. Global interpretability answers the question: which features are important, and what kind of interactions between them take place? In other words, it helps comprehend the target outcome distribution based on features and is exceedingly difficult to achieve pragmatically. Any model beyond a limited number of parameters or weights cannot fit into an average human's short-term memory. One cannot imagine a five-feature linear model as it implies drawing the estimated hyperplane in a five-dimensional space. Any space over three dimensions cannot be conceived by humans. Thus, model comprehension by humans is generally limited to parts, such as linear model weights (Molnar, 2020).

- **Global Model Interpretability on a Modular Level:** *At a modular level global explanations determine how model parts impact predictions.* A Naive Bayes model with several hundreds of features is far too large for a human's working memory. Even with memorization, quick predictions for new data points would be impractical. The joint distribution of all features is needed over and above this to estimate each feature's importance and how they affect predictions on average, making it impossible. But a single weight is easily understood. Thus, understanding some models at a modular level is probable. Not all models can be interpreted at a parameter level. For linear models, the interpretable parts are weights, for trees they are splits (selected features + cut-off points) and leaf node predictions. Linear models may seem perfectly interpretable on a modular level, but a single weight's interpretation is

inextricably linked with all other weights. This is why such an interpretation is prefaced by saying other input features remain the same, which is not realistic in most cases. A linear model predicting a house's value, accounts for both its size and number of rooms, and may negative weight the room feature. This is as it is highly correlated with the house size feature. Where people prefer larger rooms, fewer rooms in a house may be valued over a house with more rooms, if both are of the same size. Weights only make sense after contextualizing other model features. But linear model weights may still be interpreted better than deep neural network weights (Molnar, 2020).

- **Local Interpretability for a Single Prediction:** *Investigates why the model made a certain prediction for a certain instance.* This entails homing in on a single instance and examining what the model predicts for it and explaining why. For individual predictions, an otherwise complex model might behave more accessibly. Locally, predictions may only be linearly or monotonically dependent on some features, rather than complexly so. Say a house's value depends nonlinearly on its size. But when examining one particular 100 square meter house, it is possible for that subset, prediction depends linearly on size. This can be deduced by simulating how predicted price changes upon increasing or decreasing size by 10 square meters.
  Local explanations may therefore be more accurate than global ones. (Molnar, 2020).

- **Local Interpretability for a Group of Predictions -** *Answers why a model made specific predictions for a group of instances.* Multiple instance predictions may be explained either with global (modular level) interpretation methods or with individual instances. Global methods can be applied by taking the group, treating them as the complete dataset, and using global methods with the subset. Individual explanation methods can be used on each instance, then listed or aggregated for the entire group (Molnar, 2020).

Interpretation method: Various interpretation methods can be broadly distinguished based on their results. These can be summarised as follows (Molnar, 2020):

- *Feature summary statistics* - Several methods give summary statistics for every feature, with some providing a single number per feature (like feature importance), or more complex output, (e.g., pairwise feature interaction strengths) (Molnar, 2020).

- *Feature summary visualization* - Most feature summary statistics may also be visualized. Certain summaries only become meaningful if visualized, and a table would be the wrong choice. A feature's partial dependence is such a case, where plots are curves depicting a feature and the average predicted outcome. Partial dependences are ideally presented with the drawn curve rather than printed coordinates (Molnar, 2020).

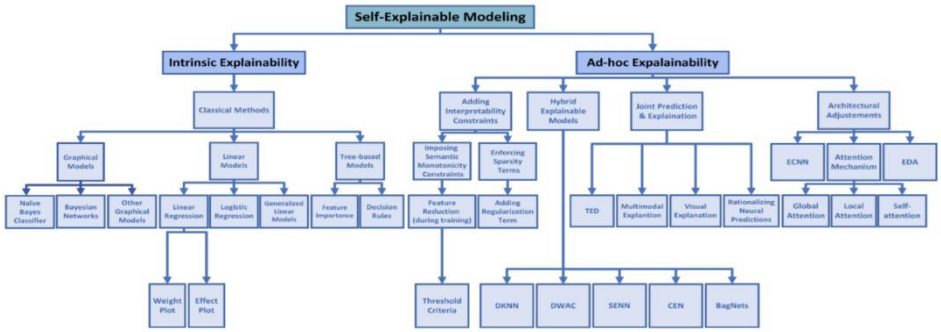
- *Model internals (e.g., learned weights)* - Intrinsically interpretable models fall into this category; for instance, linear models' weights or learned decision trees' structure (features, and thresholds for splits). There is no clear distinction between feature summary statistic and model internals in cases like linear models, as weights represent them simultaneously. Another method eliciting model internals is the feature detectors visualization in convolutional neural networks. Such methods are, by definition, model-specific (Molnar, 2020).

- *Data point* - This category comprises all methods with data points (already existent or newly created) as outputs to facilitate interpretability. One such method is counterfactual explanations. To explain a data instance forecast, the method changes some features where the predicted outcome changes accordingly (like a class prediction change), to find a similar data point. Another instance is identifying predicted class prototypes. For utility, interpretation methods returning new data points need data points that themselves are interpretable. This has limited relevance for tabular data with hundreds of features but works well for images and texts (Molnar, 2020).

- *Intrinsically interpretable model* - One black box model interpretation solution is (global or local) approximations with interpretable models. The model itself is interpreted through internal feature summary statistics or model parameters (Molnar, 2020).

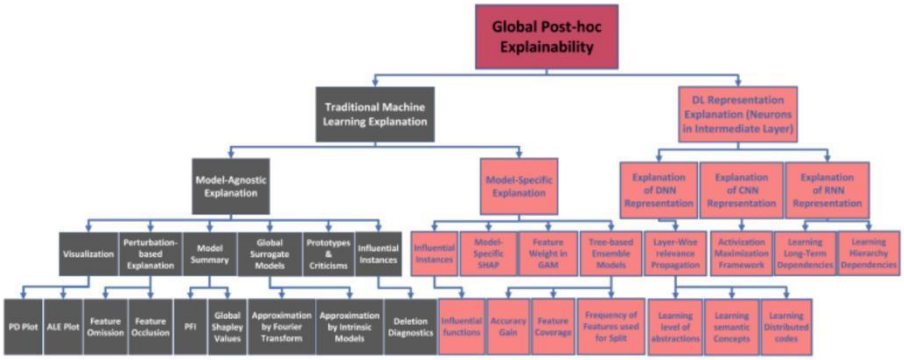
3.3.1 Explainable AI Techniques

While a technical review of XAI techniques is beyond the scope of this white paper, Nagahisarchoghaei et al. (2023) in their survey paper provide a useful visualisation of existing XAI techniques across three broad categories: (i) self-explainability (Figure 6); (ii) global post hoc explainability (Figure 7) and (iii) local post hoc explainability (Figure 8).



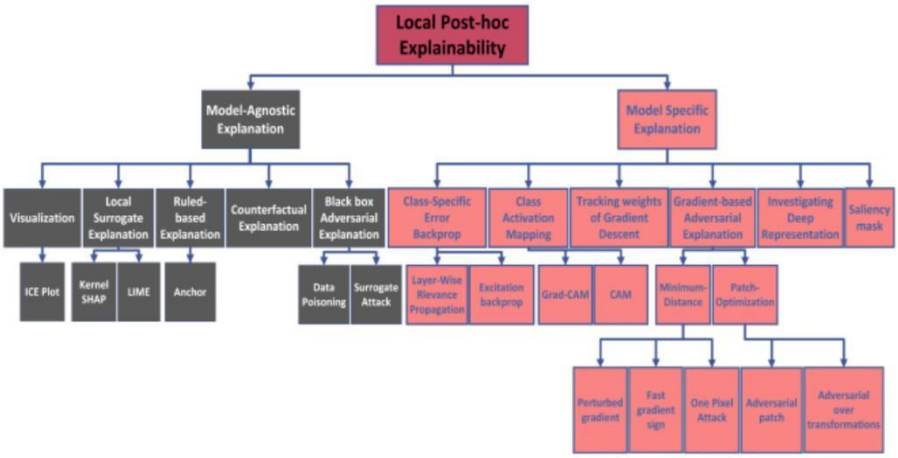
Source: Nagahisarchoghaei et al. (2023)

Figure 6: Self-Explainability Techniques



Source: Nagahisarchoghaei et al. (2023)

Figure 7: Global Post Hoc Explainability Techniques



Source: Nagahisarchoghaei et al. (2023)

Figure 8: Local Post Hoc Explainability Techniques

3.4 Approaches to Evaluating Explanations

In advance of EU laws regulating AI and some associated standards, a careful evaluation of XAI is essential to outline specific desirable properties. Given that the overarching goal of XAI is to establish trust among humans, it is crucial to prioritize properties such as human-friendliness, privacy, and non-discrimination (Robnik et al., 2018; Miller, 2019). Ali et al. (2023) document five aspects of XAI evaluations.

First, explanation evaluation can be built up on cognitive psychology theories to articulate a general formal system of how humans can interpret. By examining the cognitive state of human users, investigations can improve efficiency of explanations and enhance user understanding of AI systems. To determine what kinds of XAI are preferred, measures of understandability of users on AI agents and algorithms are imperative (Dodge et al., 2018; Penney et al., 2018; Rader and Gray, 2015). It is

also essential to consider users' attention and expectation in the process of incorporating explainability into AI systems (Stumpf et al., 2018).

Satisfaction is the second aspect of XAI evaluations. A diverse array of metrics, encompassing both subjective and objective measures, has been adopted to assess the clarity and adequacy of explanations (Miller, 2019). Curran et al. (2012) utilize a method involving ranking and coding of user transcripts to evaluate the effectiveness of explanations within a computer vision challenge. Lage et al. (2019) illustrate the importance of complexity of XAI model (length, intricacy) in affecting satisfaction. Confalonieri et al. (2021) gauge users' perceived understanding of explanations through task performance metrics, including accuracy and response time, as well as subjective measures like confidence level of user's responses.

The next aspect of XAI evaluation is trust and transparency. Cahour and Forzy (2009) adopt three trust scales in trust assessment of users. Nothdurft et al. (2014) examine the relationship between user trust and AI decision explanations, particularly focusing on transparency. Bussone et al. (2015) utilize a Likert scale and think-aloud protocols to appraise user trust in a clinical decision-support system, revealing that factual explanations contribute to an enhancement in user trust. Recently, Stepin et al. (2022) employed Likert scales to measure human perceptions of the trustworthiness of automated counterfactual explanations.

Assessment of human-AI interface is one aspect to evaluate XAI. Myers et al. (2006) introduce a framework allowing users to pose "why" and "why not" questions for coherent responses Lim et al. (2009) assess human performance using AI systems with varied explanations, considering task completion time and success rates. Evaluating the human-AI interface helps verify model outputs and debug specific AI models (Kulesza et al., 2015). Visual analytics tools like TopicPanaroma, FairSight, DGMTracker, aid domain experts in evaluating and reducing biases for fair data-driven decision-making.

The last aspect that Ali et al. (2023) propose for XAI evaluation is computational assessment. Not only human assessment, but system transparency may also be prioritized. In response, Herman (2017) advocates for computational approaches to evaluate explanation fidelity, focusing on the accuracy of saliency maps as indicators. Various computational methods have emerged to assess the validity, consistency, and fidelity of explainability techniques compared to the original blackbox model. Zeiler and Fergus (2014) demonstrate improved prediction outcomes through evaluating a CNN visualization tool's fidelity in detecting model flaws. Ross et al. (2017) evaluate the consistency and computing cost of explanations using LIME as a baseline, while Schmidt and Biessmann (2019) introduce an explanation quality score based on human intuition

4. Conclusion

In this white paper, we discuss how explainability in AI-systems can deliver transparency and build trust towards greater adoption of automation to support financial regulation compliance among banks and financial services firms. We uniquely propose the concept of *Explainable Intelligent Automation* as the next generation of Intelligent Automation. Explainable Intelligent Automation seeks to leverage emerging innovations in the area of Explainable Artificial Intelligence. AI systems underlying Intelligent Automation bring considerable advantages to the task of automating compliance processes. A barrier to AI adoption though is the black-box nature of the machine learning techniques delivering the outcomes, which is exacerbated by the pursuit of increasingly complex frameworks, such as deep learning, in the delivery of performance accuracy. Through articulating the business value of Robotic Process Automation and Intelligent Automation, we consider the potential for Explainable Intelligent Automation to add value. The solution framework sets out the Explainable Intelligent Automation framework, as the interface of Robotic Process Automation, Business Process Management and Explainable Artificial Intelligence. We also discuss key considerations of an organisation in terms of setting strategic priorities around the explainability of AI systems, the technical considerations in Explainable Artificial Intelligence analytics, and the imperative to evaluate explanations.

Bibliography

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., ... & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99, 101805.

Behl, A., Sampat, B., Pereira, V., & Chiappetta Jabbour, C. J. (2023). The role played by responsible artificial intelligence (RAI) in improving supply chain performance in the MSME sector: an empirical inquiry. *Annals of Operations Research*, 1-30.

Bussone, A., Stumpf, S., & O'Sullivan, D. (2015, October). The role of explanations on trust and reliance in clinical decision support systems. In *2015 International Conference on Healthcare Informatics* (pp. 160-169). IEEE.

Cahour, B., & Forzy, J. F. (2009). Does projection into use improve trust and exploration? An example with a cruise control system. *Safety Science*, 47(9), 1260-1270.

Coombs, C., Hislop, D., Taneva, S. K., & Barnard, S. (2020). The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review. *The Journal of Strategic Information Systems*, 101600.

Cooper, L. A., Holderness Jr, D. K., Sorensen, T. L., & Wood, D. A. (2019). Robotic process automation in public accounting. *Accounting Horizons*, 15-35.

Curran, W., Moore, T., Kulesza, T., Wong, W. K., Todorovic, S., Stumpf, S., ... & Burnett, M. (2012, February). Towards recognizing" cool" can end users help computer vision recognize subjective attributes of objects in images? In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces* (pp. 285-288).

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(1), e1391.

Dodge, J., Penney, S., Anderson, A., & Burnett, M. M. (2018). What Should Be in an XAI Explanation? What IFT Reveals. In *IUI Workshops* (pp. 1-4).

Eulerich, M., Pawlowski, J., Waddoups, N. J., & Wood, D. A. (2022). A framework for using robotic process automation for audit tasks. *Contemporary Accounting Research*, 691-720.

Grennan, L., Kremer, A., Single, A., & Zipparo, P. (2022). Why businesses need explainable AI – and how to deliver it: https://www.mckinsey.com/capabilities/quantumblack/ourhttps://www.mckinsey.com/capabilities/quantumblack/our-insights/why-businesses-need-explainable-ai-and-how-to-deliver-itinsights/why-businesses-need-explainable-ai-and-how-to-deliver-it.

Herman, B. (2017). The promise and peril of human evaluation for model interpretability. *arXiv preprint arXiv:1711.07414*.

Kulesza, T., Burnett, M., Wong, W. K., & Stumpf, S. (2015, March). Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 126-137).

Lacity, M., & Willcocks, L. (2021). Becoming strategic with intelligent automation. *MIS Quarterly Executive*, 1-14.

Lage, I., Chen, E., He, J., Narayanan, M., Kim, B., Gershman, S. J., & Doshi-Velez, F. (2019, October). Human evaluation of models built for interpretability. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* (Vol. 7, pp. 59-67).

Lim, B. Y., Dey, A. K., & Avrahami, D. (2009, April). Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2119-2128).

Mendling, J., Pentland, B. T., & Recker, J. (2020). Building a complementary agenda for business process management and digital innovation. *European Journal of Information Systems*, 208-219.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267, 1-38.

Molnar, C. (2020). *Interpretable machine learning.* Lulu. com.

Myers, B. A., Weitzman, D. A., Ko, A. J., & Chau, D. H. (2006, April). Answering why and why not questions in user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 397-406).

Nagahisarchoghaei, M., Nur, N., Cummins, L., Nur, N., Karimi, M.M., Nandanwar, S., Bhattacharyya, S. and Rahimi, S., 2023. An empirical survey on explainable ai technologies: Recent trends, use-cases, and categories from technical and application perspectives. *Electronics*, 12(5), p.1092.

Nothdurft, F., Richter, F., & Minker, W. (2014, June). Probabilistic human-computer trust handling. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (SIGDIAL) (pp. 51-59).

Penney, S., Dodge, J., Hilderbrand, C., Anderson, A., Simpson, L., & Burnett, M. (2018, March). Toward foraging for understanding of StarCraft agents: An empirical study. In *23rd International Conference on Intelligent User Interfaces* (pp. 225-237).

Rader, E., & Gray, R. (2015, April). Understanding user beliefs about algorithmic curation in the Facebook news feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 173-182).

Robnik-Šikonja, M., & Bohanec, M. (2018). Perturbation-based explanations of prediction models. *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent*, 159-175.

Ross, A. S., Hughes, M. C., & Doshi-Velez, F. (2017). Right for the right reasons: Training differentiable models by constraining their explanations. *arXiv preprint arXiv:1703.03717.*

Rydzik, A., & Kissoon, C. S. (2022). Decent work and tourism workers in the age of intelligent automation and digital surveillance. *Journal of Sustainable Tourism*, 2860-2877.

Schmidt, P., & Biessmann, F. (2019). Quantifying interpretability and trust in machine learning systems. *arXiv preprint arXiv:1901.08558.*

Stepin, I., Alonso-Moral, J. M., Catala, A., & Pereira-Fariña, M. (2022). An empirical study on how humans appreciate automated counterfactual explanations which embrace imprecise information. *Information Sciences*, 618, 379-399.

Stumpf, S., Skrebe, S., Aymer, G., & Hobson, J. (2018, March). Explaining smart heating systems to discourage fiddling with optimized behavior. In *CEUR Workshop Proceedings* (Vol. 2068).

Tussyadiah, I. (2020). A review of research into automation in tourism: Launching the Annals of Tourism Research Curated Collection on Artificial Intelligence and Robotics in Tourism. *Annals of Tourism Research*, 102883.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In

Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13 (pp. 818-833). Springer International Publishing.

About the Authors

Dr. James Bowden is Lecturer in Financial Technology at the Strathclyde Business School, University of Strathclyde, where he is the programme director of the MSc Financial Technology. Prior to this, he gained experience as a Knowledge Transfer Partnership (KTP) Associate at Bangor Business School, and he has previous industry experience within the global financial index team at FTSE Russell. Dr Bowden's research focusses on different areas of financial technology (FinTech), and his published work involves the application of text analysis algorithms to financial disclosures, news reporting, and social media. More recently he has been working on projects incorporating audio analysis into existing financial text analysis models and investigating the use cases of satellite imagery for the purpose of corporate environmental monitoring. Dr Bowden has published in respected international journals, such as the European Journal of Finance, the Journal of Comparative Economics, and the Journal of International Financial Markets, Institutions and Money. He has also contributed chapters to books including "Disruptive Technology in Banking and Finance", published by Palgrave Macmillan. His commentary on financial events has previously been published in The Conversation UK, the World Economic Forum, MarketWatch and Business Insider, and he has appeared on international TV stations to discuss financial innovations such as non-fungible tokens (NFTs).

Email: james.bowden@strath.ac.uk

Professor Mark Cummins is Professor of Financial Technology at the Strathclyde Business School, University of Strathclyde, where he leads the FinTech Cluster as part of the university's Technology and Innovation Zone leadership and connection into the Glasgow City Innovation District. As part of this role, he is driving collaboration between the FinTech Cluster and the other strategic clusters identified by the University of Strathclyde, in particular the Space, Quantum and Industrial Informatics Clusters. Professor Cummins is the lead investigator at the University of Strathclyde on the newly funded (via UK Government and Glasgow City Council) Financial Regulation Innovation Lab initiative, a novel industry project under the leadership of FinTech Scotland and in collaboration with the University of Glasgow. He previously held the posts of Professor of Finance at the Dublin City University (DCU) Business School and Director of the Irish Institute of Digital Business. Professor Cummins has research interests in the following areas: financial technology (FinTech), with particular interest in Explainable AI and Generative AI; quantitative finance; energy and commodity finance; sustainable finance; model risk management. Professor Cummins has over 50 publication outputs. He has published in leading international discipline journals such as: European Journal of Operational Research; Journal of Money, Credit and Banking; Journal of Banking and Finance; Journal of Financial Markets; Journal of Empirical Finance; and International Review of Financial Analysis. Professor Cummins is co-editor of the open access Palgrave title *Disrupting Finance: Fintech and Strategy in the 21st Century*. He is also co-author of the Wiley Finance title *Handbook of Multi-Commodity Markets and Products: Structuring, Trading and Risk Management*.

Email: mark.cummins@strath.ac.uk

Daniel Dao is a Research Associate at the Financial Regulation Innovation Lab (FRIL), University of Strathclyde. Besides, he is Doctoral Researcher in Fintech at Centre for Financial and Corporate Integrity, Coventry University, where his research topics focus on fintech (crowdfunding), sustainable finance and entrepreneurial finance. He is also working as an Economic Consultant at World Bank Group, Washington DC Headquarters, where he has been contributing to various policy publications and reports, including World Development Report 2024; Country Economic Memorandum of Latin American and Caribbean countries; Policy working papers of labor, growth, and policy reforms, etc. Regarding professional qualifications and networks, he is CFA Charter holder and an active member of CFA UK. He has earned his MBA (2017) in Finance from Bangor University, UK, and his MSc (2022) in Financial Engineering from WorldQuant University, US. He has shown a strong commitment and passion for international development and high-impact policy research. His proficiency extends to data science techniques and advanced analytics, with a specific focus on artificial intelligence, machine learning, and natural language processing (NLP).

Email: daniel.dao@strath.ac.uk

Kushagra Jain is a Research Associate at the Financial Regulation Innovation Lab (FRIL), University of Strathclyde. His research interests include artificial intelligence, machine learning, financial/regulatory technology, textual analysis, international finance, and risk management, among others. He is a recipient of doctoral scholarships from the Financial Mathematics and Computation Cluster (FMCC), Science Foundation Ireland (SFI), Higher Education Authority (HEA) and Michael Smurfit Graduate Business School, University College Dublin (UCD). Previously, he worked within wealth management and as a statutory auditor. He is due to complete his doctoral studies in Finance from UCD in 2024, and obtained his MSc in Finance from UCD, his Accounting Technician accreditation from the Institute of Chartered Accountants of India and his undergraduate degree from Bangalore University. He was formerly a FMCC Database Management Group Data Manager, Research Assistant, PhD Representative and Teaching Assistant for undergraduate, graduate and MBA programmes.

Email: kushagra.jain@strath.ac.uk

Get in touch

FRIL@FinTechscotland.com

FinTech Scotland®

University of Glasgow

University of Strathclyde Glasgow