Proc. of Int. Conf. on Computing Electrical and Electronics Engineering, CEEE

Explainable Artificial Intelligence for Banking and Financial Sector: Emerging Research Directions

Pavitha N[1] and Shounak Sugave[2]
[1-2]School of Computer Engineering and Technology, MIT World Peace University, Pune, Maharashtra, India
Email: pavithanrai@gmail.com

***Abstract*—Explainability in automated decision-making confirms transparency and trust among various stakeholders of the artificial intelligence (AI) ecosystem. The banking and finance sector (BFS) is the lifeline of an economy; the sector is experiencing massive technological disruption and innovations. The adoption of new technologies such as explainable artificial intelligence (XAI) has promising features at various levels of decision-making in BFS. This paper contributes to elaborating the emerging paradigms of XAI and presents future research directions in BFS. In conclusion, the paper recommends explainable, transparent, and trustworthy AI models for a stable, healthy and contributory BFS.**

***Index Terms*— Explainability, Banking and Finance Sector, Explainable Artificial Intelligence.**

I. INTRODUCTION

The advancement in Artificial Intelligence (AI) has enormous opportunities in the Banking and Financial Sector (BFS) for scrupulous decision making with the advancement of Machine Learning (ML) and Deep Learning (DL) technologies. However, in dynamic and complex environments, ML and DL decisions are extremely 'Black Box' in nature [56][18][19] [9][55]. BFS is a strictly regulated segment, wherein the interpretability/explainability of models is a prerequisite for the decision-making process [3][4][5][6][36]. In the recent past, to address the limitations of the Black Box models, researchers scouted for new approaches with the help of explainability techniques. This has comprehended the application of AI in BFS for a higher degree of explainability with germane decision-making power. Researchers' novel contribution and continuous expansion of AI applications with the greater notch on explainability lead to Explainable Artificial Intelligence (XAI).

Interpretability is one of the established properties of XAI to demonstrate the model's ability to disseminate information to human understanding and for appropriate decision-making [1][55]. 'Explainability', 'Fairness' and 'Transparency' are the principal forebodings in BFS for the automated decision-making process. The robustness and trustworthiness of deployed model are directly associated with an appropriate explanation for the outcome. The enhanced trust of the model envisages automatic decision-making with a higher degree of confidence and transparency.

BFS is evolving with dynamic regulatory principles due to volatility in business cycles and disruptive market behaviors from the supply and demand side. The emerging-market conditions are directing the governments, central banks, and other critical players in the BFS sector to employ data-driven decision-making for suitable financial market operations. BFS is driven through enormous structured and unstructured data/information, without much advancement in applications of XAI models for sound decision making.

Grenze ID: 02.CEEE.2022.7.506

Explainability is necessary for BFS and is motivated from different perspectives, such as (i) regulator's perspective, (ii) customers perspective, (iii) bank/institutional perspective, and (iv) employs perspective (v) developer's perspective. The regulator's intent of deploying the AI models precludes a comprehensive understanding of the technology process and its reliability in meeting the regulatory compliances. In a service lead industry like BFS customers' choice and freedom are defined with accurate decisions and implications to value creation. Commercial decisions of the banks are defined through profit maximization for the shareholders. The employees of the BFS sector shall take a fair decision and their intuitions shall propel from accurate and congenial technical models. Developers want to test and debug the models.

The paper is divided into VI sections. Section II elaborates on XAI emerging paradigm and section III covers explainability types with existing XAI models considering all the sectors. Section IV provides XAI emerging methods in BFS. Section V includes discussion and ample of future research directions in XAI for BFS and section VI makes the concluding remarks.

II. XAI An Emerging Paradigm

There are various initiatives by different organizations demonstrating the requirement of XAI and this section summarizes some of the milestone missions. The United States Defense Advanced Research Projects Agency (DARPA) has conducted one of the pioneering works in the field of XAI. The contribution of DARPA in the field of XAI is acknowledged by many researchers. DARPA's XAI "program endeavors to create AI systems whose learned models and decisions can be understood and appropriately trusted by end-users" [56][59][43][44][58].

In 2017, The United States' auspicious body of public policy in its "Statement on algorithmic transparency and accountability" [46] identified expl ainability as a critical component for effective public policy.

United States Office on Science and Technology released its report on "Preparing for the Future of Artificial Intelligence" in 2016. According to this report "AI-enabled systems are governable; that they are open, transparent, and understandable" [45][57].

European Commission issued a policy called "Algorithmic Awareness-Building" in march 2018 and states the necessity of XAI as "Algorithmic transparency is an important safeguard for accountability and fairness in decision-making" [51][52]. In continuation to this European Union Commission identifies the need for explainability in its report on "Responsible AI & National AI Strategies" [53]. Finally, guidelines released by the AI expert group on ethical and safe AI [54] and also accentuates its importance to the research domain. In addition, European Union in May 2018 executed a law on data safety and privacy [37], which necessitates algorithm decision explanations [38]. On-demand, the decisions or results should be re-traceable [39] and do not mandate to explain everything at all times, which may become a complex technical task [40]. Moreover, it is better to have an explanation for each instance [41] [42] which may be utilized on-demand.

The Academy of Sciences of the United Kingdom in its machine learning report envisioned the gravity of maintaining transparency and interpretability in managing social issues [49].

In 2018, the Prime minister task force on the National Strategy for AI of French highlighted for opening up of black boxes with obvious explicable models and satisfactory explanations [48].

The Monetary Authority of Singapore in 2018 documented a set of fundamental principles on responsible use of AI through "Fairness, Ethics, Accountability and Transparency" for the financial sector [61].

The Netherlands "Special Interest Group on Artificial Intelligence" in its "Dutch Artificial Intelligence Manifesto" prioritized the future of AI "models for making these systems socially aware, explainable and responsible" [47].

'AI Portugal 2030' documents XAI as the focal point of the national strategy to achieve an ethical and safe society with transference and accountability in decision making [50].

National Institution for Transforming India (NITI Aayog) of the Government of India released a "working document towards responsible #AIforAll" for stakeholders' response in 2020 desires explainability of AI models with ensuring trust by users [60].

The Organization for Economic Co-operation and Development (OECD) adopted principles on AI for its member countries in 2019. It is recommended for transparent and responsible AI systems to ensure decisions are democratic and explainable [62].

III. EXPLAINABILITY TYPES

The Explainability of an AI model can be categorized in 2 ways (i) Model integrated explainability (ii) Model agnostic methods. Table1 summarizes existing methods explainability types under both categories.

A. Model integrated explainability

Model integrated explainability indicates the developed model is self-explanatory in terms of the outcomes or decisions/predictions that are generated. Molnar and Christoph list the interpretable models as "linear regression, logistic regression, other linear regression extensions, decision trees, decision rules and the Rule Fit algorithms, Naive Bayes Classifier, K-Nearest Neighbors" [2]. Linear regression is used to solve regression problems and it is a linear model. Logistic regression is used to solve classification problems and it is a nonlinear model. Decision trees can be used for both classification and regression problems and it is a nonlinear model. Rule Fit also can solve both classification and regression problems but it is a linear model. Naïve Bayes algorithm can be applied to classification problems and it is nonlinear in nature. K-Nearest Neighbors can solve both classification and regression problems and it is nonlinear. The explainability of the model can be at the global level that is interpreting the entire model or it can be at the local level that is interpreting the instances. Post-hoc explainability can be applied to the interpretable models also.

TABLE I: EXISTING METHODS EXPLAINABILITY TYPES

sr No	References	Is model integrated explainability supported	Is pot-hoc interpretability supported	
1	[10][11][12][13][14][15]	Yes		
2	[16][17]	Yes	Yes	
3	[18][19][20][21][22][23]		Yes	Yes
4	[24][25][26][27][28][29][30]		Yes	

B. Model Agnostic methods

Model integrated explainability provides the greatest explainability however as the complexity of the problem increases, these models cannot solve the problem effectively. Post-hoc explainability on the other hand gives the flexibility to select any machine learning model for training and explainable algorithms are applied post-training. Molnar and Christoph list the post-hoc explainable models as "PDPs, LIME, SHAP, Anchors" [2], etc. It is very easy to plot categorical variables using partial dependence plots (PDP) and plots become self-explanatory. In the case of LIME, a local surrogate model is built and explanations are based on the local surrogate model. In SHAP the instance is created with help of shapely values grounded with contributions. Anchors make use of easy-to-explain if-then rules to explain the model. Post-hoc algorithms work independently and they are not depending on models used for training purposes, so for any complex machine learning algorithms, the post-hoc methods can be applied to get explainability.

IV. XAI MODELS FOR BFS

The economic growth and development of a nation vitally depend on robust BFS and the explainability of the models is very significant for this sector. The research studies conducted by individuals and organizations are summarized in the subsequent part of the section.

An eminent credit scoring organization Fair Issac Corporation (FICO) from the United States issued an open competitive challenge for researchers to develop XAI models in the year 2018. The HELLOC dataset was supplied to the participants of the competition to generate machine learning models that are accurate in decision-making with a higher degree of explanation. Data scientists from FICO declared Dash et al. [32] as winners for their research contribution in developing 'a rule-based classifier'. In the same challenge, Gomez et al. [33] had proposed a Support Vector Machine-based solution for training, and for getting explainability they used a new version of Anchors [22]. Chen et al., designed a 'two-layer additive risk model' which is a rule-based globally and locally interpretable model [35]. In this study, the authors formed small groups from each of the features, and finally, these groups are merged for the final outcome.

Niklas et al., [3] developed an XAI-based 'credit risk management model' for P2P lending platforms with the help of the European Conference on Artificial Intelligence (ECAI) dataset. In this work, training is performed by the XGBoost algorithm and the TreeSHAP method is employed for getting explainability.

Janet et al., [4] created an XAI fuzzy model in the financial sector for prudential compliance, fair dealing with customers, and enhanced risk management. This research concentrates on financial institutions by focusing on various stakeholders like customers, markets, etc. Authors assert that fuzzy models perform comparatively better as against neural network and logistic regression models [4].
Lara et al., [5] developed an XAI-based credit scoring model by using HELLOC and lending club datasets. The authors made use of the XGBoost model for training and three different XAI models for explanations. The global explanations are achieved through SHAP+GIRP, local feature explanation through anchors and local instance explanations attained through protoDash.
Miller et al., [6] developed a scoring model using XAI for P2P lending with the help of a lending club dataset. Authors claim that SHAP-based explanations perform better than linear approaches. With the help of "hyperparameter optimization" [6] the authors performed algorithm formations. To deal with 'data imbalance nature' authors considered resampling and/or weighting scheme [6] methods.
Neus et al., [7] developed an XAI-based PSD2 model for credit score generation by using the synthetic database. Authors claim that catboost gives the best performance for credit scoring. SHAP based explanation method is used by authors for providing both global and local explanations. In the study, Shapely values are used for feature selection [7].
Antonio et al., [8] used a prediction approach for assessing the performance of crowdlending platforms for peer-to-peer business in their research by using logistic regression. The aim of the study was to predict whether the loan request is subscribed by the crowd or not? Authors claim that their models would help financial institutions to alter loan contracts and offer the customers an attractive financial option.
BFS is undergoing rapid transformation through disruptive technologies at regulatory compliance and business management. Our study accomplishes that the explainability of models in BFS is abundantly essential for robust economic growth and development through financial stability. The future research directions in XAI for BFS are in detail discussed in section 5.

V. Discussion and Future Research Directions in BFS

The global financial crisis of 2008 shook the financial sector at all corners leading to a lack of trust in the financial institutions and their services. Since then, international institutions like the Committee on Banking Regulations and Supervisory Practices (BASEL), the international monetary fund, the world bank, and the central banks had taken several efforts to rebuild the financial ecosystem with stability and transparency. The adoption of new technologies played a great role in reporting, practicing international standard norms, anti-money laundering, etc., to the global statutory bodies like BASEL. Technology also played a significant role in the recent past in digital financial inclusion and Financial Technologies (Fintech) revolution. The world is witnessing from conventional banking to totally technology-driven banking systems like 'open banking'. The Fintech and the Application Programming Interface (API) movement had taken a paramount shift in onboarding customers, compliance management, lending, risk management, remittance, grievance redressal, and many other places in BFS.
Emerging technologies of AI are increasingly deployed at various layers of operations in BFS. According to the estimates of McKinsey advanced AI technologies will contribute additionally $1trillion in each year [63][64]. Adopting responsible, explainable and trustworthy AI by the BFS will be the foundation for value creation and eccentric customer experiences. The BFS ecosystem interacts with various stakeholders such as customers, regulators, governments, investors/shareholders, employees, and so on. The XAI-enabled interactions across the stakeholders would bring more transparency and trust in the governance and management of financial institutions. Figure 1 illustrates various research directions in BFS using XAI.
The regulation and supervision of BFS is taking a paradigm shift from a conventional manual defined process to new technology-driven models for accuracy in decisions and transparency in the process. The central bank of the countries is looking forward to technological innovations RegTech [34] and SupTech [34] for transparent and data-rich automated decision-making techniques in regulation and supervision. Developing and integrating such new technologies by using XAI would benefit the regulator in a swift and transparent process of regulation and supervision.
The global financial crisis of 2008 led to prioritizing compliance management in BFS for financial stability. The rapidly growing concern towards management of Know Your Customer (KYC), anti-money laundering (AML), terrorism financing, and so on are looking forward to using XAI-based compliance process management in BFS. The explainable models would reduce the cost of management of compliance, faster responses, and efficiency in

decision making. There is a large scope to intervene in data-driven XAI models to competitively manage internal compliance as well as regulatory compliance reporting of the banks and financial institutions.

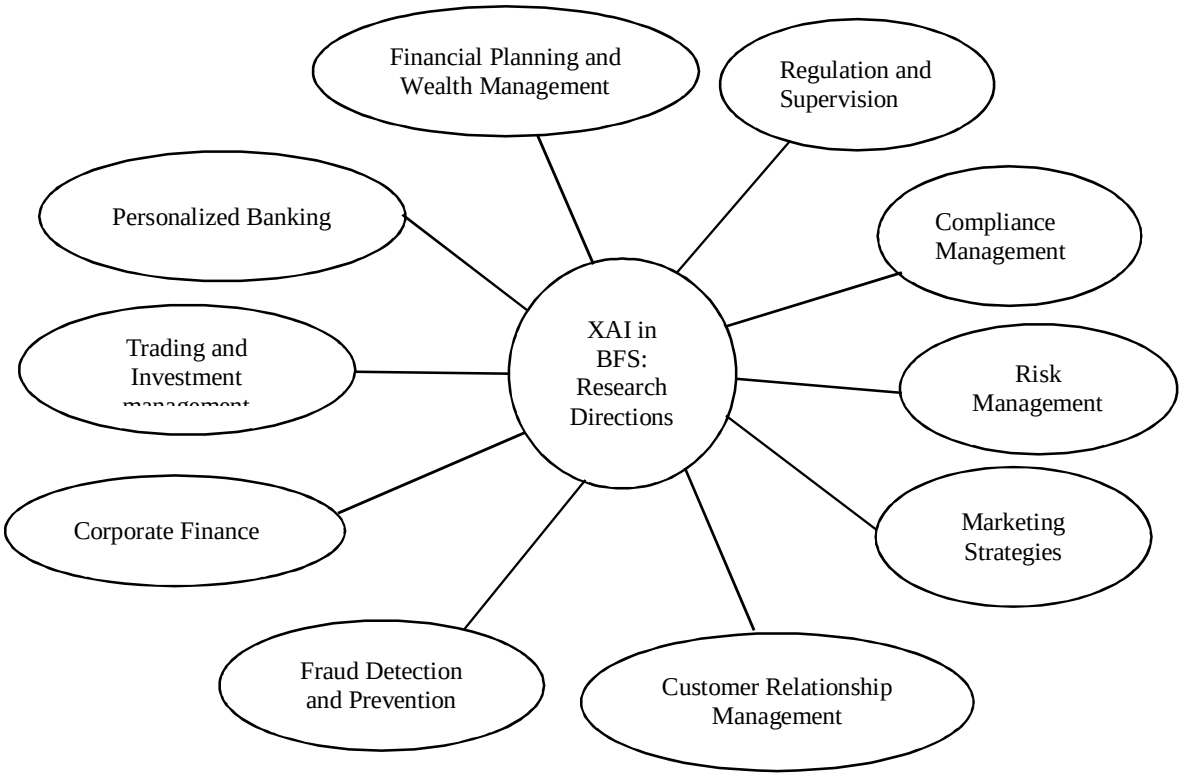

Figure 1: XAI in BFS: Future Research Directions

The BFS faces a wide range of risks due to market imperfection and obsolete method of risk assessment, as a result, the banks and financial institutions fall below the level of critical minimum business survival profit. The robust XAI models would enable banks and financial institutions test and forecast the comprehensive capital requirement to address market volatility. As banks and financial institutions are working on thin margins due to credit risk, market risk and operational risk finding suitable risk management tools through XAI would lead to profitability and efficiency in the future.

Customer is the central point of BFS's business model. Adopting new technologies like XAI for Customer Relationship Management (CRM) would allow effective customer interaction and proficiency in customer services. New technologies demystify the potential customer behaviors and permit to interact with them for better services. Utilization of XAI models in CRM will dramatically reduce the operational time of the bank executives.

Data driven marketing strategies are growing very fast in sectors like Fast Moving Consumer Goods (FMCGs), retail sector, banking, etc. Innovative technical models like XAI will predict accurate customer buying behavior based on their socio-economic characteristics. Modern banking survival ultimately depends on advanced technology-driven marketing strategies for delivering of financial products and services. Researchers in banking domain needs to innovate unique ways to reach the diversified customer segment with least cost and efficiency with the help of responsible AI.

Technical advancement in banking and financial services had opened up many new opportunities and at the same time 'digital fraudulence behavior' is one of the critical threats for the sector. There are continuous efforts from researchers and data scientists to detect and prevent fraudulence activities with the help of data science. As the sector is having very rich time series data on various segments of banking activities along with possible fraud. XAI models can detect and can generate an early warning signal to prevent fraudulence activities in BFS.

In modern banking a significant amount of business is carried by the corporate finance for private and public sector companies/corporates. The new technologies will enable banks to develop robust valuation models, due-

diligence processes, personalized banking services, investment opportunities etc. for the corporates. There is a dearth for research to evolve transparent and trustworthy AI models in corporate finance at various levels.

Trade and investment banking currently using AI analytics for data collection, predictive analytics, and trade processing. This can be enhanced further for larger integration of trade-related business and investment opportunities with the use of XAI. There are wide range of research opportunities to explore various XAI models to determine faster and more accurate bond pricing, strategies for effective fund allocation, healthy portfolio management, efficient management in front, middle and back-office of financial institutions etc.

The proliferation of the retail banking segment in the areas of consumer credit, home loan, vehicle loan, credit card services, and other numerous personal banking services getting into the limelight in the banking sector. The digital footprints of customers in the financial and non-financial sectors give room for more and more personalized services through cutting-edge technologies. The advent of XAI would enable 'right product for right customer' with seamless and tailored online experiences in retail banking services. The future studies in XAI need to get into a deeper analysis of complex customer behavior to predict the choice of markets, channels, products, and services. This will harness personalized/retail financial services with higher degree of customer experience and market penetration.

Disruptive technologies are rapidly transforming the Wealth Management and Financial Planning (WMFP) segment of BFS in recent years. WMFP is using AI-based solutions for value creation and to improve customer experiences on a real-time basis. However, effective use of XAI will benefit in customer offerings and efficiency in internal decision-making for wealth managers.

VI. Conclusion

Despite numerous advancements in AI methods, their efficiency in solving real-world problems is questioned by researchers, data scientists, and policymakers due to the 'black box' nature. Building "transparency", "explainability" and "trustworthiness" across various stakeholders of new technologies is the thrust area of research. The explainability in AI models is popularly used in areas such as medical sciences, financial sector, agriculture, and so on. The banking and financial ecosystem is massively disrupted by modern technologies like AI. A present array of literature questions the efficacy of these models for their insufficiency, explainable power in decision making in BFS. Against this backdrop, the paper presents various emerging paradigms in XAI along with explainability types. Further, the research paper summarizes various XAI models in BFS. The main aim of the paper is to present future research directions using XAI in BFS as a promising new technology model in regulations, compliance management, risk management, business management, excellence in customer experiences, and so on. The paper concludes with considerable efforts to develop XAI models in BFS to attain a larger goal of financial stability and economic growth.

References

[1] Finale Doshi Velez, Been Kim. "Towards A Rigorous Science of Interpretable Machine Learning". arXiv preprint: arXiv:1702.08608v2, 2017.

[2] Molnar, Christoph. "Interpretable machine learning. A Guide for Making Black Box Models Explainable", 2019.

[3] Bussmann, N., Giudici, P., Marinelli, D. et al. Explainable Machine Learning in Credit Risk Management. Comput Econ (2020). https://doi.org/10.1007/s10614-020-10042-0.

[4] J. Adams and H. Hagras, "A Type-2 Fuzzy Logic Approach to Explainable AI for regulatory compliance, fair customer outcomes and market stability in the Global Financial Sector," 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, United Kingdom, 2020, pp. 1-8, doi: 10.1109/FUZZ48607.2020.9177542.

[5] Lara Marie Demajo, Vince Vella and Alexiei Dingli. "EXPLAINABLE AI FOR INTERPRETABLE CREDIT SCORING". arXiv preprint: arXiv:2012.03749v1, 2020.

[6] M. J. Ariza-Garzón, J. Arroyo, A. Caparrini and M. Segovia-Vargas, "Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending," in IEEE Access, vol. 8, pp. 64873-64890, 2020, doi: 10.1109/ACCESS.2020.2984412

[7] Neus Llop Torrent, Giorgio Visani and Enrico Bagli. "PSD2 Explainable AI Model for Credit Scoring". arXiv preprint: arXiv:2011.10367v2, 2020.

[8] ANTONIO-M. MORENO- MORENO, CARLOS SANCHÍS-PEDREGOSA AND EMMA BERENGUER., "Success Factors in Peer-to-Business (P2B) Crowdlending: A Predictive approach," in IEEE Access, vol.7, pp 148586- 148593, 2019, doi: 10.1109/ACCESS.2019.2946858.

[9] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," in IEEE Access, vol. 6, pp. 52138-52160, 2018, doi: 10.1109/ACCESS.2018.2870052.

[10] Himabindu Lakkaraju, Stephen H Bach, and Jure Leskovec. 2016. “Interpretable Decision Sets: A Joint Framework for Description and Prediction,”. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 1675–1684.
[11] Xiyang Hu, Cynthia Rudin, and Margo Seltzer. 2019. “Optimal Sparse Decision Trees,”. In Advances in Neural Information Processing Systems 32. 7267–7275.
[12] Berk Ustun, Alexander Spangher, and Yang Liu. 2019. “Actionable Recourse in Linear Classification,”. In Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency. 10–19.
[13] Berk Ustun, Stefano Tracà, and Cynthia Rudin. 2013. “Supersparse Linear Integer Models for Interpretable Classification,”. arXiv:1306.6677.
[14] Rich Caruana, Yin Lou, Johannes Gehrke Microsoft, Paul Koch, Marc Sturm, and Noémie Elhadad. 2015. “Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission,”. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 1721–1730.
[15] Yin Lou, Rich Caruana, and Johannes Gehrke. 2012. “Intelligible Models for Classification and Regression,”. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 150–158.
[16] Gregory Plumb, Denali Molitor, and Ameet Talwalkar. 2018. “Model Agnostic Supervised Local Explanations. In Advances in Neural Information Processing Systems,”. 2515–2524.
[17] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 1135–1144.
[18] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex Degrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. “From local explanations to global understanding with explainable AI for trees,”. Nature Machine Intelligence 2, 1 (2020), 56–67.
[19] Scott M Lundberg, Gabriel G Erion, and Su-In Lee. 2018. “Consistent Individualized Feature Attribution for Tree Ensembles,”. arXiv:1802.03888.
[20] Scott M Lundberg and Su-In Lee. 2017. “A unified approach to interpreting model predictions,”. In Advances in neural information processing systems. 4765–4774.
[21] Scott M. Lundberg, Bala Nair, Monica S. Vavilala, Mayumi Horibe, Michael J. Eisses, Trevor Adams, David E. Liston, Daniel King Wai Low, Shu Fang Newman, Jerry Kim, and Su In Lee. 2018. “Explainable machine-learning predictions for the prevention of hypoxaemia during surgery,”. Nature Biomedical Engineering 2, 10 (oct 2018), 749–760.
[22] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2018. “Anchors: High-Precision Model-Agnostic Explanations,”. In AAAI, Vol. 18. 1527–1535.
[23] Hiroshi Tsukimoto. 2000. “Extracting Rules from Trained Neural Networks. Transactions on Neural Networks,” 11, 2 (2000), 512–519.
[24] Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus Robert Müller, and Wojciech Samek. 2015. “On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation,”. PLoS ONE 10, 7 (2015), 1–46.
[25] David Baehrens, Timon Schroeter, Stefan Harmeling, Motoaki Kawanabe, Katja Hansen, and Klaus-Robert Müller. 2010. “How to explain individual classification decisions,”. The Journal of Machine Learning Research 11 (2010), 1803–1831.
[26] Yin Lou, Rich Caruana, Johannes Gehrke, and Giles Hooker. 2013. “Accurate Intelligible Models with Pairwise Interactions,”. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 623–631.
[27] Matthew D Zeiler and Rob Fergus. 2014. “Visualizing and understanding convolutional networks. In European conference on computer vision,”. Springer, 818–833.
[28] Pang Wei Koh and Percy Liang. 2017. “Understanding Black-box Predictions via Influence Functions,”. In Proceedings of the 34th International Conference on Machine Learning. 1885–1894.
[29] Rafael Poyiadzi, Kacper Sokol, Raul Santos-Rodriguez, Tijl De Bie, and Peter Flach. 2020. “FACE: feasible and actionable counterfactual explanations,”. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. 344–350.
[30] Berk Ustun and Cynthia Rudin. 2019. “Learning Optimized Risk Scores,”. Journal of Machine Learning Research 20, 150 (2019), 1–75.
[31] Kasun Amarasinghe, Kit Rodolfa, Hemank Lamba, Rayid Ghani. “Explainable Machine Learning for Public Policy: Use Cases, Gaps, and Research Directions,”. arXiv preprint: arXiv:2010.14374v1, 2020.
[32] Dash, S., Gunluk, O., & Wei, D. (2018). Boolean decision rules via column generation. In Advances in Neural Information Processing Systems (pp. 4655-4665).
[33]] Gomez, O., Holter, S., Yuan, J., & Bertini, E. (2020, March). ViCE: visual counterfactual explanations for machine learning models. In Proceedings of the 25th International Conference on Intelligent User Interfaces (pp. 531-535).
[34] Shaktikanta Das, Governor, Reserve Bank of India, “Opportunities and Challenges of FinTech,” Keynote Address Delivered at the NITI Aayog’s FinTech Conclave 2019. Available online: https://rbidocs.rbi.org.in/rdocs/Speeches/PDFs/GSFNA250319A D0EE1F30EB746028A177251138EC297.PDF
[35] Chen, C., Lin, K., Rudin, C., Shaposhnik, Y., Wang, S., & Wang, T. (2018). An interpretable model with globally consistent explanations for credit risk. arXiv preprint arXiv:1811.12615.

[36] Roberto Confalonieri, Ludovik Coba, Benedikt Wagner, Tarek R. Besold, “A historical perspective of explainable Artificial Intelligence,” 2020, DOI: 10.1002/widm.139.
[37] European Commission. General Data Protection Regulation. 2016. Available online: https://eur-lex.europa. eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.
[38] Weller, A. Challenges for transparency. arXiv 2017, arXiv:1708.01870.
[39] Holzinger, A.; Biemann, C.; Pattichis, C.S.; Kell, D.B. What do we need to build explainable AI systems for the medical domain? arXiv 2017, arXiv:1712.09923.
[40] Wachter, S.; Mittelstadt, B.; Russell, C. Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. (2017). Harv. J. Law Technol. 2017, 31, 841.
[41] Goodman, B.; Flaxman, S. EU regulations on algorithmic decision-making and a “right to explanation”. arXiv 2016, arXiv:1606.08813.
[42] Wachter, S.; Mittelstadt, B.; Floridi, L. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. Int. Data Priv. Law 2017, 7, 76–99.
[43] Gunning, D. Explainable Artificial Intelligence (XAI); Defense Advanced Research Projects Agency: Arlington, VA, USA, 2017; Volume 2.
[44] Gunning, D. Explainable Artificial Intelligence (XAI). Available online: https://www.darpa.mil/program/explainable-artificial-intelligence.
[45] CTNSTCPHP- Committee on Technology National Science and Technology Council and Penny Hill Press (2016). Preparing for the Future of Artificial Intelligence; CreateSpace Independent Publishing Platform: Scotts Valley, CA, USA, 2016.
[46] ACM US Public Council (2017). Statement on Algorithmic Transparency and Accountability. 2017. Avaivlable online: https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf.
[47] IPN SIG AI. Dutch Artificial Intelligence Manifesto. 2018. Available online: http://ii.tudelft.nl/bnvki/wpcontent/uploads/2018/09/Dutch-AI-Manifesto.pdf.
[48] Cédric Villani. AI for Humanity—French National Strategy for Artificial intelligence. 2018. Available online: https://www.aiforhumanity.fr/en/
[49] Royal Society. Machine Learning: The Power and Promise of Computers that Learn by Example. 2017. Available online: https://royalsociety.org/topics-policy/projects/machine-learning/.
[50] Portuguese National Initiative on Digital Skills. AI Portugal 2030. 2019. Available online: https://www. incode2030.gov.pt/sites/default/filess/draft_ai_portugal_2030v_18mar2019.pdf.
[51] European Commission. Artificial Intelligence for Europe. 2018. Available online: https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence europe.
[52] European Commission. Algorithmic Awareness-Building. 2018. Available online: https://ec.europa.eu/digital-single-market/en/algorithmic-awareness-building.
[53] Rao, A.S. Responsible AI & National AI Strategies. 2018. Available online: https://ec.europa.eu/growth/tools/databases/dem/monitor/sites/default/files/4%20International%20initiatives%20v3_0.pdf.
[54] High-Level Expert Group on Artificial Intelligence (AI HLEG). Ethics Guidelines for Trustworthy Artificial Intelligence. 2019. Available online: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines.
[55] Carvalho, D.V.; Pereira, E.M.; Cardoso, J.S. Machine Learning Interpretability: A Survey on Methods and Metrics. Electronics 2019, 8, 832. doi:10.3390/electronics8080832.
[56] David Gunning, David W. Aha. “DARPA’s Explainable Artificial Intelligence Program”, in Association for the Advancement of Artificial Intelligence, 2019.
[57] https://www.canada.ca/en/treasury-board-secretariat/corporate/reports/treasury-board-canada-secretariat-2018-19-departmental-plan.html.
[58] https://iui.asscm.org/2019/toc.html.
[59] https://iui.acm.org/2019/keynotes.html
[60] https://niti.gov.in/sites/default/files/2020-11/Towards_Responsible_AIforAll_Part1.pdf
[61] https://www.mas.gov.sg/publications/monographs-or-information-paper/2018/feat
[62] https://www.oecd.org/going-digital/ai/principles/
[63] https://www.mckinsey.com/industries/financial-services/our-insights/ai-bank-of-the-future-can-banks-meet-the-ai-challenge
[64] https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-executives-ai-playbook?page=industries/banking/