[1] Zhang, Yu, Peter Tiňo, Aleš Leonardis, and Ke Tang. "A survey on neural network interpretability." IEEE Transactions on Emerging Topics in Computational Intelligence 5, no. 5 (2021): 726-742.

[2] Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." Information fusion 58 (2020): 82-115.

[3] Angelov, Plamen P., Eduardo A. Soares, Richard Jiang, Nicholas I. Arnold, and Peter M. Atkinson. "Explainable artificial intelligence: an analytical review." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 11, no. 5 (2021): e1424.

[4] Yang, Wenli, Yuchen Wei, Hanyu Wei, Yanyu Chen, Guan Huang, Xiang Li, Renjie Li et al. "Survey on Explainable AI: From Approaches, Limitations and Applications Aspects." Human-Centric Intelligent Systems 3, no. 3 (2023): 161-188.

[5] Xu, Feiyu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. "Explainable AI: A brief survey on history, research areas, approaches and challenges." In Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8, pp. 563-574. Springer International Publishing, 2019.

[6] Gunning, David, and David Aha. "DARPA's explainable artificial intelligence (XAI) program." AI magazine 40, no. 2 (2019): 44-58.

[7] Saeed, Waddah, and Christian Omlin. "Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities." Knowledge-Based Systems 263 (2023): 110273.

[8] Adadi, Amina, and Mohammed Berrada. "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)." IEEE access 6 (2018): 52138-52160.

[9] Nagahisarchoghaei, Mohammad, Nasheen Nur, Logan Cummins, Nashtarin Nur, Mirhossein Mousavi Karimi, Shreya Nandanwar, Siddhartha Bhattacharyya, and Shahram Rahimi. "An empirical survey on explainable ai technologies: Recent trends, use-cases, and categories from technical and application perspectives." Electronics 12, no. 5 (2023): 1092.

[10] Preece, Alun. "Asking 'Why'in AI: Explain-ability of intelligent systems–perspectives and challenges." Intelligent Systems in Accounting, Finance and Management 25, no. 2 (2018): 63-72.

[11] Miller, Tim. "Explanation in artificial intelligence: Insights from the social sciences." Artificial intelligence 267 (2019): 1-38.

[12] Lundberg, Scott M., Bala Nair, Monica S. Vavilala, Mayumi Horibe, Michael J. Eisses, Trevor Adams, David E. Liston et al. "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery." Nature biomedical engineering 2, no. 10 (2018): 749-760.

[13] Vilone, Giulia, and Luca Longo. "Classification of explainable artificial intelligence methods through their output formats." Machine Learning and Knowledge Extraction 3, no. 3 (2021): 615-661.

[14] Agarwal, Garvita, Lauren Hay, Ia Iashvili, Benjamin Mannix, Christine McLean, Margaret Morris, Salvatore Rappoccio, and Ulrich Schubert. "Explainable AI for ML jet taggers using expert variables and layerwise relevance propagation." Journal of High Energy Physics 2021, no. 5 (2021): 1-36.

[15] Saranya, A., and R. Subhashini. "A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends." Decision analytics journal (2023): 100230.

[16] Minh, Dang, H. Xiang Wang, Y. Fen Li, and Tan N. Nguyen. "Explainable artificial intelligence: a comprehensive review." Artificial Intelligence Review (2022): 1-66.

[17] Walia, Savita, Krishan Kumar, Saurabh Agarwal, and Hyunsung Kim. "Using xai for deep learning-based image manipulation detection with shapley additive explanation." Symmetry 14, no. 8 (2022): 1611.

[18] Meena, Jaishree, and Yasha Hasija. "Application of explainable artificial intelligence in the identification of Squamous Cell Carcinoma biomarkers." Computers in Biology and Medicine 146 (2022): 105505.

[19] Batarseh, Feras A., Laura Freeman, and Chih-Hao Huang. "A survey on artificial intelligence assurance." Journal of Big Data 8, no. 1 (2021): 60.

[20] Chen, Han-Yun, and Ching-Hung Lee. "Vibration signals analysis by explainable artificial intelligence (XAI) approach: Application on bearing faults diagnosis." IEEE Access 8 (2020): 13324246-134256.

[21] Dağlarli, Evren. "Explainable artificial intelligence (xAI) approaches and deep meta-learning models." Advances and applications in deep learning 79 (2020).

[22] Dodge, Jonathan, and Margaret Burnett. "Position: We Can Measure XAI Explanations Better with Templates." In ExSS ATEC@ IUI, pp. 1-13. 2020.

[23] Carrillo, Alfredo, Luis F. Cantú, and Alejandro Noriega. “Individual explanations in machine learning models: A survey for practitioners.” arXiv preprint arXiv:2104.04144 (2021).

[24] Adadi, Amina, and Mohammed Berrada. “Peeking inside the black-box: a survey on explainable artificial intelligence (XAI).” IEEE access 6 (2018): 52138-52160.

[25] Apley, Daniel W., and Jingyu Zhu. “Visualizing the effects of predictor variables in black box supervised learning models.” Journal of the Royal Statistical Society Series B: Statistical Methodology 82, no. 4 (2020): 1059-1086.

[26] Dabkowski, Piotr, and Yarin Gal. “Real time image saliency for black box classifiers.” Advances in neural information processing systems 30 (2017).

[27] Došilović, Filip Karlo, Mario Brčić, and Nikica Hlupić. “Explainable artificial intelligence: A survey.” In 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO), pp. 0210-0215. IEEE, 2018.

[28] Samek, Wojciech, Thomas Wiegand, and Klaus-Robert Müller. “Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models.” arXiv preprint arXiv:1708.08296 (2017).

[29] Cortez, Paulo, and Mark J. Embrechts. “Using sensitivity analysis and visualization techniques to open black box data mining models.” Information Sciences 225 (2013): 1-17.

[30] Saeed, Waddah, and Christian Omlin. “Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities.” Knowledge-Based Systems 263 (2023): 110273.

[31] Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García et al. “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI.” Information fusion 58 (2020): 82-115.

[32] Gade, Krishna, Sahin Geyik, Krishnaram Kenthapadi, Varun Mithal, and Ankur Taly. “Explainable AI in industry: Practical challenges and lessons learned.” In Companion Proceedings of the Web Conference 2020, pp. 303-304. 2020.

[33] Lundberg, Scott M., and Su-In Lee. “A unified approach to interpreting model predictions.” Advances in neural information rocessing systems 30 (2017). [34] Štrumbelj, Erik, and Igor Kononenko. “Explaining prediction models and individual predictions with feature contributions.” Knowledge and information systems 41 (2014): 647-665.

[35] Marcinkevičs, Ričards, and Julia E. Vogt. “Interpretable and explainable machine learning: A methods-centric overview with concrete examples.” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery (2023): e1493.

[36] Burkart, Nadia, and Marco F. Huber. “A survey on the explain-ability of supervised machine learning.” Journal of Artificial Intelligence Research 70 (2021): 245-317.

[37] Su, Guolong, Dennis Wei, Kush R. Varshney, and Dmitry M. Malioutov. “Learning sparse two-level boolean rules.” In 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1-6. IEEE, 2016.

[38] Samek, Wojciech, Thomas Wiegand, and Klaus-Robert Müller. “Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models.” arXiv preprint arXiv:1708.08296 (2017).

[39] Islam, Sheikh Rabiul, William Eberle, Sheikh Khaled Ghafoor, and Mohiuddin Ahmed. “Explainable artificial intelligence approaches: A survey.” arXiv preprint arXiv:2101.09429 (2021).

[40] Došilović, Filip Karlo, Mario Brčić, and Nikica Hlupić. “Explainable artificial intelligence: A survey.” In 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO), pp. 0210-0215. IEEE, 2018.

[41] Tjoa, Erico, and Cuntai Guan. “A survey on explainable artificial intelligence (xai): Toward medical xai.” IEEE transactions on neural networks and learning systems 32, no. 11 (2020): 4793-4813.

[42] Verma, Sahil, Varich Boonsanong, Minh Hoang, Keegan E. Hines, John P. Dickerson, and Chirag Shah. “Counterfactual explanations and algorithmic recourses for machine learning: A review.” arXiv preprint arXiv:2010.10596 (2020).

[43] Vellido, Alfredo, José David Martín-Guerrero, and Paulo JG Lisboa. “Making machine learning models interpretable.” In ESANN, vol. 12, pp. 163-172. 2012.

[44] Doshi-Velez, Finale, and Been Kim. “Towards a rigorous science of interpretable machine learning.” arXiv preprint arXiv:1702.08608 (2017).

[45] Bhatt, Umang, McKane Andrus, Adrian Weller, and Alice Xiang. “Machine learning explainability for external stakeholders.” arXiv preprint arXiv:2007.05408 (2020). [46] Weller, Adrian. “Transparency: motivations and challenges.” In Explainable AI: interpreting, explaining and visualizing deep learning, pp. 23-40. Cham: Springer International Publishing, 2019.

[47] Doshi-Velez, Finale, and Been Kim. “Towards a rigorous science of interpretable machine learning.” arXiv preprint arXiv:1702.08608 (2017).

[48] Amarasinghe, Kasun, Kit T. Rodolfa, Hemank Lamba, and Rayid Ghani. “Explainable machine learning for public policy: Use cases, gaps, and research directions.” Data & Policy 5 (2023): e5.

[49] González-Nóvoa, José A., Laura Busto, Juan J. Rodríguez-Andina, José Fariña, Marta Segura, Vanesa Gómez, Dolores Vila, and César Veiga. “Using explainable machine learning to improve intensive care unit alarm systems.” Sensors 21, no. 21 (2021): 7125.

[50] Tiddi, Ilaria, and Stefan Schlobach. “Knowledge graphs as tools for explainable machine learning: A survey.” Artificial Intelligence 302 (2022): 103627.

[51] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. “Anchors: High-precision model-agnostic explanations.” In Proceedings of the AAAI conference on artificial intelligence, vol. 32, no. 1. 2018.

[52] Biran, Or, and Courtenay Cotton. “Explanation and justification in machine learning: A survey.” In IJCAI-17 workshop on explainable AI (XAI), vol. 8, no. 1, pp. 8-13. 2017.

[53] Miller, Tim. “Explanation in artificial intelligence: Insights from the social sciences.” Artificial intelligence 267 (2019): 1-38.

[54] Bellucci, Matthieu, Nicolas Delestre, Nicolas Malandain, and Cecilia Zanni-Merk. “Une terminologie pour une IA explicable contextualisée.” In EXPLAIN'AI Workshop EGC 2022. 2022.

[55] Bussmann, Niklas, Paolo Giudici, Dimitri Marinelli, and Jochen Papenbrock. “Explainable machine learning in credit risk management.” Computational Economics 57 (2021): 203-216.

[56] Molnar, Christoph. Interpretable machine learning. Lulu. com, 2020.

[57] Murdoch, W. James, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu. “Definitions, methods, and applications in interpretable machine learning.” Proceedings of the National Academy of Sciences 116, no. 44 (2019): 22071-22080.

[58] Bracke, Philippe, Anupam Datta, Carsten Jung, and Shayak Sen. “Machine learning explainability in finance: an application to default risk analysis.” (2019).

[59] Beckh, Katharina, Sebastian Müller, Matthias Jakobs, Vanessa Toborek, Hanxiao Tan, Raphael Fischer, Pascal Welke, Sebastian Houben, and Laura von Rueden. “Explainable machine learning with prior knowledge: an overview.” arXiv preprint arXiv:2105.10172 (2021).

[60] Wexler, James, Mahima Pushkarna, Tolga Bolukbasi, Martin Wattenberg, Fernanda Viégas, and Jimbo Wilson. “The what-if tool: Interactive probing of machine learning models.” IEEE transactions on visualization and computer graphics 26, no. 1 (2019): 56-65.

[61] Von Rueden, Laura, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch et al. “Informed machine learning–a taxonomy and survey of integrating prior knowledge into learning systems.” IEEE Transactions on Knowledge and Data Engineering 35, no. 1 (2021): 614-633.

[62] Samek, Wojciech, Thomas Wiegand, and Klaus-Robert Müller. “Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models.” arXiv preprint arXiv:1708.08296 (2017). 1139