Annex 1. How Machine Learning Algorithms Work

While machine learning (ML) is a common fixture of today's artificial intelligence (AI) systems, this has not always been the case. Early AI relied heavily on expert systems consisting of a set of precise rules created by human experts. These were rules that a computer could follow, step by step, often giving the impression that it was responding to changing situations. However, given that these rules were often expressed in an "if-then" format, they imposed heavy limitations, basically rendering such systems useless when facing situations outside of the prescribed environment. Advances in ML offer AI systems the ability to learn from experience, thus making them more adaptable to new states of the world.

The key element that constitutes the basis of ML is that all learning can be reduced to learning the representation of a mapping between some input and output. For example, if a classification problem is considered in which objects are split into predetermined classes, the objective of ML becomes learning a mapping between the features of the objects and the predetermined classes. An important fact to understand is that the mapping does not necessarily have to have a functional form, that is, a mathematical expression. There is an important distinction between fitting and learning processes. If the functional forms for the mapping were known, with the only unknowns being its parameters, then the problem becomes a fitting task, which means that all that is left to do is to estimate the parameters. When the functional form is not known, or does not exist, fitting is not helpful, and a learning process needs to be employed.

A ML model usually consists of the selected relevant features from the input data along with a learning algorithm and performance criteria. Regardless of the learning paradigm, the learning process requires the selection of a set of features from the available data on which a learning algorithm, or even a set of learning algorithms, is applied. The algorithms are evaluated based on some performance criteria that vary with the problem for which a solution needs to be learned. The combination of features, learning algorithms, and performance criteria can be thought of as the ML model that provides the learned mapping from the inputs to the outputs.

The learning paradigms depend on the data available for the learning process:

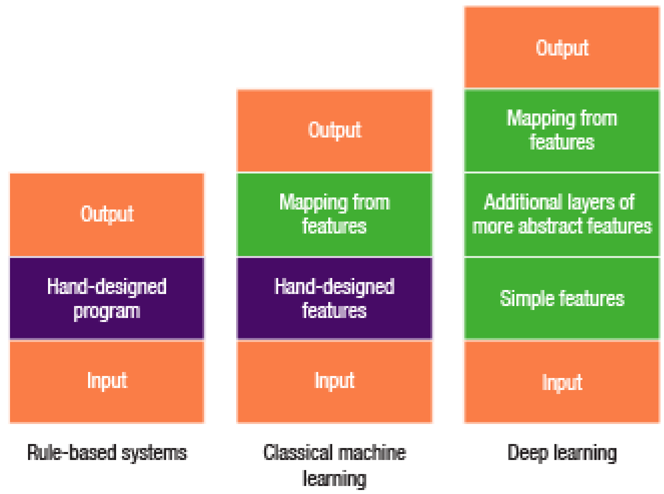
- *Supervised learning* requires that the training data contain both the inputs (or features) and the correct associated outputs (labels, or targets). In this case, the objective of ML is to find a mapping between the known inputs and outputs through the use of an algorithm. The idea is that the mapping (or the model)[1] learned this way can be used on unseen inputs and accurately predict the outputs. Overfitting is one of the main challenges in ML and it occurs when the model performs very well on the training data set, but poorly on new, unseen data (test data). The better a model performs on the training data, the less biased it is. However, its performance can change dramatically on a set of unseen data. In this case, in order to achieve robustness in performance, the researcher may opt to choose a model that does not perform as well on the training set but maintains its performance on the test set. This situation is known as the bias-variance trade-off.
- *Unsupervised learning* occurs when the available data does not contain the outputs. Given that only the features are available (and without an associated label), the algorithm employed will not only have to learn the mapping, but also generate the labels. For example, in anomaly detection problems, no prior information exists about the data points that represent an anomaly. In this situation, using a clustering algorithm is a common approach to learn a viable mapping. Under the assumption that all normal activity occurs in clusters, the data points that are identified as belonging to clusters can be labeled as normal activity, while the rest of the data points will be labeled as anomalies.

[1]In practice, it is common to refer to the mapping as a model; the terms are used interchangeably from here on.

- *Reinforcement learning* is used in situations when the available data is not fixed, and it changes through a feedback loop from the environment to the AI system. In addition, the labels or outputs associated with inputs are not necessarily correct or desired. The most common example of reinforcement learning is learning to play a game, such as chess.

Deep learning is an approach to AI/ML that has been responsible for most of the recent developments of AI/ML systems, including image and speech recognition (Hinton, LeCun, and Bengio 2015). The main difference between deep learning and the classical ML approach resides in the way they pre-process the raw input data before feeding it into the algorithm (Annex Figure 1.1). Classical ML modeling requires the modeler to transform the original raw data into a set of variables (features) that can be used by the AI/ML system to train the model. For instance, if the task is to recognize handwritten numbers, the modeler would have to transform the image pixels into a set of variables, such as curvature, size, and density, and then use this transformed data to train a ML model. Deep learning, conversely, embeds this process into the AI/ML system, which would be trained using the raw data. One of its tasks would be to extract from the data the important features required to maximize its performance criteria. For instance, considering the aforementioned task, the modeler would feed the images directly into the system and it would select the features set that works best (Goodfellow, Bengio, and Courville 2016).

Annex Figure 1.1. Machine Learning Paradigms



Sources: IMF staff; adapted from Goodfellow, Bengio, and Courville (2016).

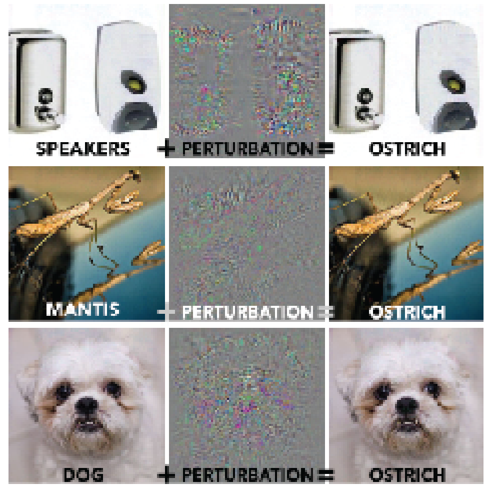
Another feature of deep learning models is that they are comprised of several learning layers, where each layer is responsible for representing concepts that are expressed in terms of previous layers that represent simpler concepts (Hinton, LeCun, and Bengio 2015). For example, the first layer of the system can recognize limbs, the second one may recognize human limbs, and the third and final one can leverage these previous layers and recognize humans in a picture. The number of layers is related to the depth of the deep learning system. In general, the deeper the model, the more flexible (and more complex) it is. On one hand, it enables great performance in complex tasks, such as face recognition and translation but, on the other hand, it is usually referred to as a black box system and is at the lower end of explainable models (Guidotti and others 2019). The depth and complexity of such models has implications for the security of the system, given that it could be more prone to adversarial attacks, which consist of building examples with small perturbations that can easily trick the AI system (Goodfellow, Shlens, and Szegedy 2014). A well-known example is in Annex Figure 1.2, where a human imperceptible perturbation is added to a panda picture that tricks the AI to classify it as a gibbon with high level of confidence.

Natural language processing (NLP) is a set of computational techniques that allow machines to learn, understand, and produce human language contents. It dates to the 1940s, when scientists began to work with populating computers with vocabularies and human language rules. From the 1990s onward scientists started to favor the currently used statistical or corpus-based methods largely because of the difficulties of interpreting human language in a structured manner (Hirshberg and Manning 2015). Nevertheless, NLP suffers from a major limitation: only a limited set of the languages currently spoken generate online material in the necessary magnitude to train algorithms. Even though many of the remaining languages are spoken and written by a large share of the world's

population, they lack proper online resources for NLP algorithms, yielding inaccurate models for such tasks as translations or conversational agents on such languages, therefore contributing to a widening of the digital divide. Researchers aim to rectify this situation by developing multilingual NLP using unsupervised models and novel techniques to transfer resources between various languages (Ponti and others 2019).

Annex Figure 1.2. Example of an Input Attack

In an input (or adversarial example) attack, researchers add an imperceptible perturbation (middle column, pictures magnified by 10×) to a previously recognized image (left column), causing the machine learning model to misclassify all the resulting images as ostriches.



Source: Adapted from "Adversarial examples generated for AlexNet" by Szegedy and others (2014), licensed under CC BY 3.0.

NLP is currently used in several applications, such as:

- To facilitate human-to-human communication via automated language translation, a capability that was boosted by the exceptional quantities of parallel text brought by the Internet in the 1990s. In a context-aware blind evaluation in 2020, human judges thought a deep learning system outperformed news translation in accuracy, with comparable fluency.
- To provide spoken dialogue systems and conversational agents, leading to such devices as Apple Siri or Amazon Alexa personal assistants. Although dialogue management and contextualization are generally deficient at this moment, conversational technology is widespread and performs well under suitable conditions, such as when conversation topics are known in advance.
- To extract knowledge by reading and understanding free text. Such capability may be used to create structured information databases from textual records. For instance, in 2015 researchers were able to predict drug-gene relationships by using algorithms to read and analyze over 23 million articles containing the results of biomedical research abstracts (Percha and Altman 2015).
- To analyze social media, by retrieving information from sources such as Twitter, Facebook, YouTube, among other. Even though data from these sources is often unreliable, it enables the extraction of useful analytics from language used by users in forums and discussions, such as demographics. Furthermore, it is often used to analyze speaker states: the opinions, beliefs, emotions, and other personal views expressed in language. This is usually done through sentiment analysis, whereby one may determine the positive or negative emotional attitude of subjects toward a predetermined subject.

Annex 2. Artificial Intelligence in Finance—Risk Profile

Sources of Risk	Risk Category	Risk Materialization
Data	Quality and availability	Inaccurate or suboptimal outcome
	Privacy	Data leaks undermining public trust
	Embedded bias	Discriminatory practices
Algorithm	Performance robustness	Inability to recognize structural shifts in the data
	Explainability	Difficulties in explaining or detecting the appropriateness of machine learning (ML) decisions
	Privacy	ML operations creating inadvertent data leaks undermining public trust
	Embedded bias	Discriminatory practices
	Development governance	Inadequate deployment of ML systems in production environments
Cybersecurity	Traditional cyber threats	Human or software failures
	Data poisoning	Corrupting artificial intelligence (AI)/ML training
	Input manipulation	Misleading AI/ML systems during operations
	Model extraction/inversion	Recover training data, or the model itself
Technology Management	Operational capacity	Inadequate deployment, operation, and risk management of AI/ML systems
	Outsourcing AI/ML services	Exposure to third-party risks
Systemic	Network effects	Higher concentration of AI/ML service providers who could become systemically important
	Risk assessment	Buildup of systemic risk due to greater homogeneity in risk assessments and credit decisions
	Interconnectedness	Inaccurate ML risk assessment and reaction could quickly amplify and spread in the financial system
	Regulatory gaps	Development outpaces regulations and providers who may fall outside regulatory perimeter

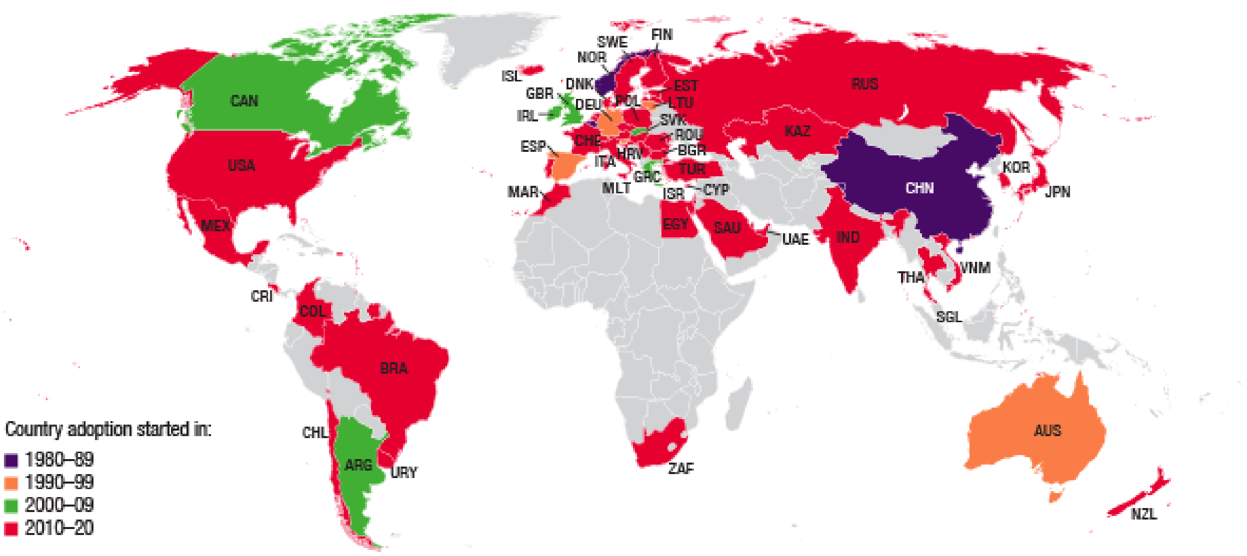
Annex 3. National Artificial Intelligence Strategies

This annex provides an overview of national artificial intelligence (AI) strategies. The overview is intended to highlight efforts and approaches to developing AI strategies across the globe as well as priorities and key lessons. This overview is not intended to provide details on specific national AI strategies or views on their qualities.

Considering the promise and complexity of AI, a rising number of countries are developing national AI strategies. About 60 countries around the world have developed national AI strategies (most since 2010 (OECD 2021)). Some national authorities have pursued overarching multiyear strategies of AI implementation that are intended to attract private investment, foster innovation, and develop a skilled workforce for the future. Success indicators often include mapping AI national strategies to the Sustainable Development Goals adopted by United Nations member states in 2015 (Stanford University 2019).

Developing national strategies on AI is not new but has accelerated significantly in recent years (Annex Figure 3.1). Recent years have witnessed a rapid increase in the number of countries developing national AI strategies, driven by technological advances that are facilitating fast deployments of AI systems across a wide range of sectors (for example, security, financial, health).

Annex Figure 3.1. National Artificial Intelligence Strategy Landscape



Source: IMF staff, based on information in OECD (2021).
Note: Data labels use International Organization for Standardization (ISO) country codes.

Approaches to national AI strategies vary across countries, reflecting differences in national priorities, as well as available skills and resources. The coverage of a national strategy on AI is defined by the number of policy initiatives a country develops and the spread of these policies across multiple themes (for example, finance, education, health, climate change). AI strategies are sometimes embedded in broader science and technology initiatives. Furthermore, AI strategies are linked to strategies and regulations related to, among others, data access, sharing and privacy frameworks, intellectual property management, national security, and ethical use of AI. Regulations, such as the US Health Insurance Portability and Accountability Act and General Data Protection Regulation, will need to be continuously reviewed to address new concerns arising from AI systems deployment.

OECD Artificial Intelligence Policy Observatory data show that 60 countries and the European Union have developed AI strategies that encompass, collectively, more than 600 policy initiatives covering a broad range of issues. The widest coverage is attributed to the United States, where more than 47 initiatives have been developed on AI usage and coverage. Regionally, the European Union has more coverage than any other region, with 51 initiatives.[1] Common elements of these strategies touch on leadership and vision, focus and specialization, partnership and collaboration, including with academia, AI research and development, developing human capital, governance, and risk management (Annex Figure 3.2).

A review of current AI national and regional strategies reveals several common lessons in designing and implementing AI development strategies:

- Designing and implementing a successful AI strategy is linked to clear objectives, often the United Nations Sustainable Development Goals. This linkage enables robust assessment and monitoring, which in turn translates to valuable outcomes on national, regional, and global levels.
- There is a need for enhanced open access to research on AI. Publishing and sharing AI research, including data and code, would aid practitioners and policymakers.
- Stronger emphasis should be placed on developing human capital. Those skilled in AI are heavily recruited by major companies, resulting in a brain drain. The slow growth in skilled graduates with formal AI education must be addressed.
- Focus on ethics in AI. While technical standards are designed to address concerns about using technology and managing associated risks, they do not address, by design, the ethical aspects of developing AI technologies, which need to be considered when designing AI strategies.

Collaboration at the international level is picking up. The European Commission and its 28 state members, in addition to Norway, signed the "Declaration of Cooperation on AI" in 2018 (Stix 2020). It aims to define and implement a comprehensive and integrated approach and to review and modernize individual state

Annex Figure 3.2. Key Features of National Artificial Intelligence Strategies

(Number of policy initiatives embedded in 60 national AI strategies)

AI strategies' priorities focus on developing an enabling environment, including, in particular, the governance of AI development, digital infrastructure, upgrading skills, and promoting innovative businesses and sectors.

Source: IMF staff, based on information provided by OECD (2021).
Note: AI = artificial intelligence.

[1] Of the 60 countries and regions that have developed AI strategies, 11 (Argentina, Brazil, Colombia, Estonia, Germany, India, Israel, Japan, Lithuania, Turkey, and the United States) have included initiatives that aim to address the COVID-19 pandemic (OECD 2021).

policies “to ensure that opportunities arising from AI are seized and the emerging challenges addressed.” The Chief Executive Board within the United Nations has launched an initiative with a concrete roadmap to build capacity in harnessing the potential of AI and raise awareness regarding its risks (UN 2019). The Group of Twenty (G20) Digital Economy Task Force has taken the lead in advancing the G20 AI principles. The main objective of those principles is to foster public trust and confidence in AI technologies and realize their potential through a human-centered approach (G20 2019).

References

Alonso, C., A. Berg, S. Kothari, C. Papageorgiou, and S. Rehman. 2020. "Will the AI Revolution Cause a Great Divergence?" IMF Working Paper 20/184, International Monetary Fund, Washington, DC.

Arner, D., J. Barberis, and R. Buckley. 2017. "FinTech, RegTech, and the Reconceptualization of Financial Regulation." *Northwestern Journal of International Law & Business* 37 (3). https://scholarlycommons.law.northwestern.edu/njilb/vol37/iss3/2/.

Bank of England (BoE). 2020. "The Impact of COVID on Machine Learning and Data Science in UK Banking." *Quarterly Bulletin 2020 Q4*. https://www.bankofengland.co.uk/quarterly-bulletin/2020/2020-q4/the-impact-of-covid-on-machine-learning-and-data-science-in-uk-banking.

Bazarbash, M. 2019. "FinTech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk." IMF Working Paper 19/109, International Monetary Fund, Washington, DC.

Bolhuis, M. A., and B. Rayner. 2020. "Deus ex Machina? A Framework for Macro Forecasting with Machine Learning." IMF Working Paper 20/45, International Monetary Fund, Washington, DC.

Chandola, V., A. Banerjee, and V. Kumar. 2009. "Anomaly Detection: A Survey." *ACM Computing Surveys* 41 (3). https://doi.org/10.1145/1541880.1541882.

Cloudera Fast Forward. 2020. "Interpretability." Cloudera Fast Forward Labs Research, Santa Clara, CA. https://ff06-2020.fastforwardlabs.com/ff06-2020-interpretability.pdf.

Corbett-Davies, S., E. Pierson, A. Feller, S. Goel, and A. Huq. 2017. "Algorithmic Decision Making and the Cost of Fairness." In *Proceedings of the 23rd International Conference on Knowledge Discovery and Data Mining,* 797–806.

Comiter, M. 2019. "Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It." Belfer Center for Science and International Affairs, Harvard Kennedy School. https://www.belfercenter.org/publication/AttackingAI.

Danielsson, J., R. Macrae, and A. Uthemann. 2020. "Artificial Intelligence As a Central Banker." VOX CEPR Policy Portal, March 6. https://voxeu.org/article/artificial-intelligence-central-banker.

De Nederlandsche Bank (DNB). 2019. "General Principles for Use of Artificial Intelligence in Finance." DNB, Amsterdam.

di Castri, S., S. Hohl, A. Kulenkampff, and J. Prenio. 2019. "The Suptech Generations." FSI Insights on Policy Implementation, Bank for International Settlements, Basel. https://www.bis.org/fsi/publ/insights19.pdf.
Digalaki, E. 2021. "The Impact of Artificial Intelligence in the Banking Sector & How AI is Being Used in 2021." *Insider*, January 13. https://www.businessinsider.com/ai-in-banking-report.

Doerr, S., L. Gambacorta, and J. M. Serena. 2021. "Big Data and Machine Learning in Central Banking." BIS Working Paper 930, Bank for International Settlements, Basel.

European Central Bank (ECB). 2019. "Bringing Artificial Intelligence to Banking Supervision." https://www.bankingsupervision.europa.eu/press/publications/newsletter/2019/html/ssm.nl191113_4.en.html.

Friedman, B., and H. Nissenbaum. 1996. "Bias in Computer Systems." *ACM Transactions on Information Systems* 14 (3): 330–47.

Financial Stability Board (FSB). 2020. "The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions: Market Developments and Financial Stability Implications." Financial Stability Board, Basel, Switzerland. https://www.fsb.org/wp-content/uploads/P091020.pdf.

FinCoNet. 2020. "SupTech Tools for Market Conduct Supervisors." Paris.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. 2020. “Predictably Unequal? The Effects of Machine Learning on Credit Markets.” Mimeo. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3072038.

Gates, S. W., V. G. Perry, and P. M. Zorn. 2002. “Automated Underwriting in Mortgage Lending: Good News for the Underserved?”, Housing Policy Debate, 13 (2): 369–91.

Gambacorta, L., Y. Huang, H. Qiu, and J. Wang. 2019. “How Do Machine Learning and Nontraditional Data Affect Credit Scoring? New Evidence from a Chinese Fintech Firm.” BIS Working Paper 834, Bank for International Settlements, Basel.

Gensler, G., and L. Bailey. 2020. “Deep Learning and Financial Stability.” MIT Working Paper, Massachusetts Institute of Technology, Boston, MA. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3723132.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep Learning*. Cambridge, MA: MIT Press.

Goodfellow, I. J., J. Shlens, and C. Szegedy. 2014. “Explaining and Harnessing Adversarial Examples.” Cornell University. https://arxiv.org/abs/1412.6572.

Goodman, B., and S. Flaxman. 2016. “European Union Regulations on Algorithmic Decision Making and a ‘Right to Explanation’.” https://arxiv.org/pdf/1606.08813.pdf.

Google and International Finance Corporation. 2020. “e-Conomy Africa 2020.” https://www.ifc.org/wps/wcm/connect/e358c23f-afe3-49c5-a509-034257688580/e-Conomy-Africa-2020.pdf?MOD=AJPERES&CVID=nmuGYF2.

Group of Twenty (G20). 2019. “G20 Ministerial Statement on Trade and Digital Economy.” https://www.mofa.go.jp/files/000486596.pdf.

Guidotti, R., A. Monreale, S. Ruggieri, F. Turini, D. Pedreschi, and F. Giannotti. 2019. “A Survey of Methods for Explaining Black Box Models.” *ACM Computing Surveys* 51 (5): 1–42.

Haksar, V., Y. Carrière-Swallow, E. Islam, A. Giddings, K. Kao, E. Kopp, and G. Quiros. 2021. “Towards A Global Approach to Data in the Digital Age.” IMF Staff Discussion Note, International Monetary Fund, Washington, DC.

Hao, K. 2019. “This is How AI Bias Really Happens—And Why It’s So Hard to Fix.” *MIT Technology Review*, February 4, Massachusetts Institute of Technology, Cambridge, MA.

Harker, P. 2020. “The Economics of Artificial Intelligence and Machine Learning.” Speech at the Official Monetary and Financial Institutions Forum, September 29, Philadelphia, PA.

Hinton, G., Y. LeCun, and Y. Bengio. 2015. “Deep Learning.” *Nature 521* (7553): 436–44.

Hirshberg, J., and C. D. Manning. 2015. “Advances in Natural Language Processing.” *Science* 349 (6245): 261–66.

Hong Kong Monetary Authority (HKMA). 2020. “Reshaping Banking with Artificial Intelligence.” https://www.hkma.gov.hk/media/eng/doc/key-functions/finanical-infrastructure/Whitepaper_on_AI.pdf.

Institute of International Finance (IIF). 2017. “Deploying Regtech against Financial Crime.” Report of the Regtech Working Group, IIF, Washington, DC. https://www.iif.com/portals/0/Files/private/32370132_aml_final_id.pdf.

International Monetary Fund (IMF). 2020. *Regional Economic Outlook for Sub-Saharan Africa*, April, Washington, DC.

Khan, A. 2018. “A Behavioral Approach to Financial Supervision, Regulation, and Central Banking.” IMF Working Paper 18/178, International Monetary Fund, Washington, DC.

Khan, A., and M. Malaika. 2021. “Central Bank Risk Management, Fintech, and Cybersecurity.” IMF Working Paper 2021/105, International Monetary Fund, Washington DC.

Khandani, A., K. Adlar, and A. Lo. 2010. “Consumer Credit-Risk Models via Machine Learning Algorithms.” *Journal of Banking & Finance* 34 (11): 2767–787.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J., Ludwig, J., and S. Mullainathan. 2018a. “Human Decisions and Machine Predictions.” *The Quarterly Journal of Economics* 133 (1): 237–93.

Kleinberg, J., J. Ludwig, S. Mullainathan, A. Rambachan. 2018b. “Algorithmic Fairness.” *AEA Papers and Proceedings* 108: 22–27.

Kleinberg, J., J. Ludwig, S. Mullainathan, C. Sunstein. 2019. “Discrimination in the Age of Algorithms.” *Journal of Legal Analysis* 10: 113–74.

Lehmann, E. L. 1951. “A General Concept of Unbiasedness.” *The Annals of Mathematical Statistics* 22 (4): 587–92.

Lessmann, S., B. Baesens, H. V. Seow, and L. C. Thomas. 2015. “Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring: An Update of Research.” *European Journal of Operational Research* 247 (1): 124–36.

Liu, K., B. Dolan-Gavitt, and S. Garg. 2018. “Fine-Pruning: Defending against Backdooring Attacks on Deep Neural Networks.” In *Research in Attacks, Intrusions, and Defenses*, edited by M. Bailey, T. Holz, M. Stamatogiannakis, and S. Ioannidis, 273–94. Cham, Switzerland: Springer.

Mayson, S. G. 2019. “Bias In, Bias Out.” *The Yale Law Journal* 128 (8): 2218–300.

McKinsey. 2020a. “AI-Bank of the Future: Can Banks Meet the AI Challenge?” https://www.mckinsey.com/industries/financial-services/our-insights/ai-bank-of-the-future-can-banks-meet-the-ai-challenge#.

McKinsey. 2020b. “The State of AI in 2020: Survey.” https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020.

Miller, A. P. 2018. “Want Less-Biased Decisions? Use Algorithms.” *Harvard Business Review*, July 26. https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms.

Molnar, C. 2021. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.* https://christophm.github.io/interpretable-ml-book/.

Monetary Authority of Singapore (MAS). 2018. “Principles to Promote Fairness, Ethics, Accountability, and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore’s Financial Sector.”

Murphy, K. P. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.

Organisation for Economic Co-operation and Development (OECD). 2019. “Recommendation of the Council on Artificial Intelligence.” OECD Legal Instruments, OECD, Paris. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

Organisation for Economic Co-operation and Development (OECD). 2021. “STIP Compass Database.” OECD, Paris. https://stip.oecd.org/stip.htm.

Percha, B., and R. B. Altman. 2015. “Learning the Structure of Biomedical Relationships from Unstructured Text.” *PLoS Computational Biology* 11 (7): e1004216.

Petropoulos, A., V. Siakoulis, E. Stavroulakis, and A. Klamargias. 2019. “A Robust Machine Learning Approach for Credit Risk Analysis of Large Loan Level Datasets Using Deep Learning and Extreme Gradient Boosting.” *IFC Bulletin* 49: 1486–506, Bank for International Settlements, Basel.

Plous, S. 2002. “The Psychology of Prejudice, Stereotyping, and Discrimination: An Overview.” In *Understanding Prejudice and Discrimination*, edited by S. Plous. New York: McGraw-Hill.

Ponti, E.M., H. O’Horan, Y. Berzak, I. Vulić, R. Reichart, T. Poibeau, E. Shutova, and A. Korhonen. 2019. “Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing.” *Computational Linguistics* 45 (3): 559–601.

Ribeiro, M. T., S. Singh, and C. Guestrin. 2016. "'Why Should I Trust You?': Explaining the Predictions of Any Classifier." Cornell University. https://arxiv.org/abs/1602.04938.

Sahay, R., U. E. von Allmen, A. Lahreche, P. Khera, S. Ogawa, M. Bazarbash, and K. Beaton. 2020. "The Promise of Fintech: Financial Inclusion in the Post–COVID-19 Era." Departmental Paper 20/09, International Monetary Fund, Washington, DC.

Schizas, E., G. McKain, B. Zhang, A. Ganbold, P. Kumar, H. Hussain, K. J. Garvey, et al. 2019. "The Global Regtech Industry: Benchmark Report." Cambridge Centre for Alternative Finance. University of Cambridge, UK. https://www.jbs.cam.ac.uk/wp-content/uploads/2020/08/2019-12-ccaf-global-regtech-benchmarking-report.pdf.

Shapley, Lloyd S. 1953. "A Value for N-Person Games." *Contributions to the Theory of Games* 2 (28): 307–17.

Silberg, J., and J. Manyika. 2019. "Notes from the AI Frontier: Tackling Bias in AI (and in Humans)." McKinsey Global Institute.

Stanford University. 2019. "The AI Index 2019 Annual Report." AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA. https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf.

Stix, C. 2020. "A Survey of the European Union's Artificial Intelligence Ecosystem." JRC Science Hub Communities, European Commission, Brussels. https://ec.europa.eu/jrc/communities/sites/default/files/ff3afe_1513c6bf2d81400eac182642105d4d6f.pdf.

Sy, A., R. Maino, A. Massara, H. Perez-Saiz, and P. Sharma. 2019. "Fintech in Sub-Saharan African Countries: A Game Changer?" Departmental Paper 19/04, International Monetary Fund, Washington, DC.

Szegedy, C., W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. 2014. "Intriguing Properties of Neural Networks." Cornell University. https://arxiv.org/abs/1312.6199.

Triepels, R., H. Daniels, and R. Heijmans. 2018. "Detection and Explanation of Anomalous Payment Behavior in Real Time Gross Settlement Systems." In *Enterprise Information Systems*, edited by S. Hammoudi, M. Śmiałek, O. Camp, and J. Filipe, 145–61. Cham: Springer Verlag.

United Nations (UN). 2019. A United Nations System-Wide Strategic Approach and Road Map for Supporting Capacity Development on Artificial Intelligence, CEB/2019/1/Add.3 (17 June 2019). rom undocs.org/en/CEB/2019/1/Add.3.

United Nations Educational, Scientific and Cultural Organization (UNESCO). 2021. Intergovernmental Meeting of Experts (Category II) related to a Draft Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000373434.

Wang, J. C., and C. B. Perkins. 2019. "How Magic a Bullet Is Machine Learning for Credit Analysis? An Exploration with FinTech Lending Data." Federal Reserve of Boston Working Paper 19–16.

Wang, T. 2016 "The Human Insights Missing from Big Data." https://www.ted.com/talks/tricia_wang_the_human_insights_missing_from_big_data.

World Economic Forum (WEF). 2018. "The New Physics of Financial Services: Understanding How Artificial Intelligence is Transforming the Financial Ecosystem." http://www3.weforum.org/docs/WEF_New_Physics_of_Financial_Services.pdf.

World Economic Forum (WEF). 2020. "Transforming Paradigms: A Global AI in Financial Services Survey." World Economic Forum and Cambridge Centre for Alternative Finance. http://www3.weforum.org/docs/WEF_AI_in_Financial_Services_Survey.pdf.

PUBLICATIONS

Powering the Digital Economy

Departmental Paper No. DP/2021/024