



USING MACHINE LEARNING FOR ANTI-CORRUPTION RISK AND COMPLIANCE

Coalition for Integrity



Coalition for Integrity



The Coalition for Integrity is a non-profit, non-partisan 501(c)(3) organization. We work in coalition with a wide range of individuals and organizations to combat corruption and promote integrity in the public and private sectors. www.coalitionforintegrity.org

Every effort has been made to verify the accuracy of the information contained in this report.

© 2021 Coalition for Integrity. All rights reserved.

USING MACHINE LEARNING FOR ANTI-CORRUPTION RISK AND COMPLIANCE

Researched and Produced By
Coalition for Integrity ©2021



Table of Contents

I. Introduction	1
A. Background.....	1
B. Regulatory Expectations.....	2
C. Artificial Intelligence and Anti-Corruption.....	3
D. Purpose of This Guidance	6
II. Executive Summary.....	11
III. Evaluating the Use of Machine Learning in Anti-Corruption Risk and Compliance Functions: Analysis and Recommendations.....	13
A. Stating the Business Case for Anti-Corruption Machine Learning.....	15
1. Rule-Based ANI Versus Machine Learning	15
2. Staffing, Training, and Experience	17
3. Scope of the Solution	17
4. Cost and Return on Investment of Machine Learning Solutions	17
B. Framing the Problem	18
1. Articulate the Problem	18
2. See What Data It Has That Could Be Used in That Solution	22
3. Design the Data for the Model.....	23
4. Determine Where the Data Come From	24
5. Determine and Prioritize Easily Obtained Inputs	24
6. Determine Quantifiable Outputs	25
C. Constructing the Dataset	26
D. Transforming the Data.....	27
E. Training the Model	28
1. Class Imbalance	28
2. Overfitting and Underfitting	30
3. Bias	31
4. Accuracy, Precision, Recall, and F1 Score	33

F. Making Predictions and Assessing Performance	36
G. Examples.....	39
1. AB InBev.....	39
2. Microsoft.....	45
3. Alexion Pharmaceuticals	51
IV. Ethical, Legal, and Governance Issues in Implementation and Operation of Anti-Corruption Machine Learning: Analysis and Recommendations	55
A. Ethical Issues.....	55
1. Responsible Design and Use.....	56
2. Ethical Use	57
B. Legal Issues	58
1. Data Privacy	58
2. Cybersecurity	62
3. Use for Lawful Purposes	63
C. Governance Issues.....	65
Conclusions	67
Appendices	70
Appendix 1: Glossary	71
Appendix 2: Sources	78

I. Introduction

A. Background

In every region of the world, companies of all types need to pay close attention to the perennial problem of corruption and how to avoid becoming entangled in it. While media reporting often focuses on cases in which major companies pay substantial criminal and civil penalties to resolve investigations into bribery of foreign government officials, bribery and corruption risks are by no means limited to transnational activities and senior officials. Companies in the United States and other countries may also encounter corruption risks from domestic officials at all levels of government,¹ from within their own ranks,² and in some cases even from their competitors.³

Moreover, prevention of corruption is a societal interest “of the highest importance.”⁴ From a public-policy standpoint, it is important for companies to have effective anti-corruption programs not only to reduce their own risks, but collectively to play a role with the public sector in limiting the pernicious effects of corruption.

To maintain effective anti-corruption programs, companies must be attentive to a variety of risk and compliance obligations. As reflected in guidance by the U.S. Department of Justice and the Securities and Exchange Commission (SEC), these obligations include establishing and maintaining a risk-assessment process that timely identifies potential bribery and corruption concerns; overseeing and monitoring gifts, travel, and entertainment expenses by employees; making and tracking charitable and political donations; and conducting due

1 See, e.g., Deferred Prosecution Agreement, *United States v. Commonwealth Edison Co.* (N.D. Ill., signed July 17, 2020) (agreement by utility with U.S. Department of Justice to pay \$200 million criminal fine with respect to utility’s efforts to influence and reward Speaker of Illinois House of Representatives), available at <https://www.justice.gov/usao-ndil/press-release/file/1295241/download>.

2 See, e.g., *United States v. Stanford*, 805 F.3d 557 (5th Cir. 2015), cert. denied, 137 S. Ct. 491 (2016).

3 See Roger Alford, Deputy Assistant Attorney General, Antitrust Division, U.S. Department of Justice, *Antitrust Enforcement and the Fight Against Corruption*, Remarks for the Conference on Rule of Law and Anti-Corruption Challenges, University of Notre Dame and University of Sao Paulo (October 3, 2017), <https://www.justice.gov/opa/speech/file/1001076/download>.

4 *First National Bank of Boston v. Bellotti*, 435 U.S. 765, 789 (1978).

diligence on and engaging third parties to handle outsourced functions or provide various products and services.⁵

B. Regulatory Expectations

Furthermore, two sets of formal guidance by government entities indicate that companies need to be increasingly attentive to data from across an enterprise that is relevant to risk and compliance functions. First, a 2020 addition to the U.S. Department of Justice’s formal guidance to federal prosecutors on evaluating corporate compliance programs states that prosecutors should inquire into whether compliance and control personnel “have sufficient direct or indirect access to relevant sources of data to allow for timely and effective monitoring and/or testing of policies, controls, and transactions,” and whether “any impediments exist that limit access to relevant sources of data.”⁶ Subsequently, in a deferred prosecution agreement between the Justice Department and a global financial institution to resolve a Foreign Corrupt Practices Act investigation, the agreed statement of facts stated that the financial institution’s control functions “did not engage in electronic surveillance of [an institution executive’s] correspondence or activities to determine whether [a third party, since indicted] was involved” in a particular deal, and noted that the institution’s remedial measures ultimately included “implementing heightened controls and additional procedures and policies relating to electronic surveillance and investigation.”⁷

Second, in its 2018 Due Diligence Guidance for Responsible Business Conduct, the Organisation for Economic Co-operation and Development (OECD) stated that

systems that collect information at a local level (e.g., supplier assessment data), but then are aggregated at a centralised department (e.g., headquarters or regional

5 See CRIMINAL DIVISION, U.S. DEPARTMENT OF JUSTICE AND ENFORCEMENT DIVISION, SECURITIES AND EXCHANGE COMMISSION, A RESOURCE GUIDE TO THE U.S. FOREIGN CORRUPT PRACTICES ACT 59-65 (2d ed. 2020) [hereinafter “FCPA Resource Guide”], <https://www.justice.gov/criminal-fraud/file/1292051/download>.

6 Criminal Division, U.S. Department of Justice, *Evaluation of Corporate Compliance Programs* (updated June 2020), <https://www.justice.gov/criminal-fraud/page/file/937501/download>.

7 Statement of Facts, Deferred Prosecution Agreement, *United States v. Goldman Sachs Group*, Criminal No. 20-437 (MKB), ¶¶4(f) at 5 and 54 at 19 (E.D.N.Y., filed October 22, 2020), <https://www.justice.gov/usao-edny/press-release/file/1329961/download>.

office) may help to identify trends more widely and can be used as a basis for sharing lessons learned across the enterprise.⁸

Neither of these guidance documents states that all companies are required to adopt anti-corruption artificial intelligence. Together, however, they strongly indicate that companies must be prepared to demonstrate that they can and do draw on all relevant data within their enterprise to manage their compliance programs effectively.

Because many companies and financial institutions have large quantities of data that warrant review in anti-corruption risk and compliance functions, many firms are using various types of data analytics and programming to identify risk and compliance concerns and predict potentially risky transactions and relationships.⁹ The larger the company and the volume of its internal data, however, the greater the potential burden on corporate anti-corruption teams in reviewing data and identifying patterns and anomalies that require further inquiry.

C. Artificial Intelligence and Anti-Corruption

As a result, companies, governments, nonprofit entities, and academic researchers are showing increased interest in the potential use of artificial intelligence (AI) for anti-corruption purposes.¹⁰ While there is still no “precise, universally accepted definition of AI,”¹¹ AI can be broadly divided into two categories: **artificial general intelligence** (AGI), which has been broadly defined as “a machine capable of understanding the world as well as any human,

8 ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, OECD DUE DILIGENCE GUIDANCE FOR RESPONSIBLE BUSINESS CONDUCT at 84 (2018).

9 Note: In the context of machine learning, the term “predict” does not mean that the person using a machine learning solution can prophesy or predict a specific event in the future, or decide that a specific individual or relationship should be maintained or terminated. In the simplest terms, the word “prediction,” “in the context of machine learning, is an information output that comes from entering some data and running an algorithm.” Ajay Agrawal, Joshua Gans, and Avi Goldfarb, *How to Win with Machine Learning*, HARVARD BUSINESS REVIEW, September–October 2020, <https://hbr.org/2020/09/how-to-win-with-machine-learning>. That output can enable a company to deduce or infer, based on empirical data, that specific transactions or relationships pose various levels of risk that may require further scrutiny and evaluation by human beings with appropriate knowledge and expertise. See *predict*, v. Oxford English Dictionary.

10 See André Petheram and Isak Nti Asare, *From open data to artificial intelligence: the next frontier in anti-corruption*, Oxford Insights, July 27, 2018, <https://www.oxfordinsights.com/insights/aiforanticorruption>.

11 See, e.g., ONE HUNDRED YEAR STUDY ON ARTIFICIAL INTELLIGENCE, STANFORD UNIVERSITY, ARTIFICIAL INTELLIGENCE AND LIFE IN 2030: REPORT OF THE 2015 STUDY PANEL 12 (September 2016), https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fni_singles.pdf; Vinay Sharma, *Can artificial intelligence stop corruption in its tracks?*, Governance for Development, November 15, 2018, <https://blogs.worldbank.org/governance/can-artificial-intelligence-stop-corruption-its-tracks>.

and with the same capacity to learn how to carry out a huge range of tasks”¹²; and **artificial narrow intelligence** (ANI), which can be defined as a machine and or system “that can perform only one narrowly defined task (or a small set of related tasks).”¹³

AGI remains a tantalizing but elusive goal for researchers¹⁴ and one that has no current utility for practical application in the business world. For those reasons, this guidance will refer hereafter only to ANI. Under the broad heading of ANI, there are two main categories of programming to which this guidance will refer.

The first is **rule-based programming**. For a rule-based solution, human programmers write code that establishes rules defining all aspects of a specific task (e.g., “If A occurs then do X, if something other than A occurs, then do Y”) and install them in a computer system.¹⁵

The second is **machine learning**. In machine learning, “machines” (computer systems) are not dependent on rules that human programmers write, but rather “‘learn’ from data or from their own ‘experiences’.”¹⁶ A subfield of machine learning is **deep learning**, which makes use of a **neural network**. A neural network is one type of machine learning model that typically consists of tens to hundreds of layers of simple processing units (called **neurons**), in which the simple processing units in the later layers learn useful attributes derived from attributes that processing units in the earlier layers learned.¹⁷

12 Nick Heath, *What is artificial general intelligence?*, ZDNet, August 22, 2018, <https://www.zdnet.com/article/what-is-artificial-general-intelligence/>.

13 MELANIE MITCHELL, *ARTIFICIAL INTELLIGENCE: A GUIDE FOR THINKING HUMANS* 46 (2019). Note: In each instance where a key word or phrase relating to AGI or ANI is used, this document will highlight the first instance of that word or phrase in bold italics. A glossary containing definitions and explanations of such words and phrases is included as Appendix 1 in this guidance.

14 See, e.g., Ragnar Fjelland, *Why general artificial intelligence will not be realized*, 7 HUMANITIES AND SOCIAL SCIENCES COMMUNICATIONS 10 (June 17, 2020), available at <https://www.nature.com/articles/s41599-020-0494-4>.

15 Elana Krasner, *How to choose between rule-based AI and machine learning*, TechTalks, June 5, 2020, <https://bdtechtalks.com/2020/06/05/rule-based-ai-vs-machine-learning/>.

16 MELANIE MITCHELL, *supra* note 13, at 21. Two main categories of machine learning are **supervised learning**, in which human beings create labels for categories of data in a dataset and the machine learning **algorithm** learns on that labeled dataset and can evaluate its accuracy on training data; and **unsupervised learning**, in which human programmers do not label data for the machine learning algorithm and the algorithm identifies features and patterns in the dataset. Isha Salian, *SuperVize Me: What’s the Difference Between Supervised, Unsupervised, Semi-Supervised and Reinforcement Learning?*, NVIDIA, August 2, 2018, <https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/#:~:text=In%20a%20supervised%20learning%20model,and%20patterns%20on%20its%20own>. The Glossary in Appendix 1 for further information regarding supervised and unsupervised learning, as well as other forms of machine learning such as **semi-supervised** and **reinforcement learning**.

17 See JOHN D. KELLEHER AND BRENDAN TIERNEY, *DATA SCIENCE* 241-242 (2018).

One category of machine learning that has increasingly benefited from deep learning is **natural-language processing** (NLP).¹⁸ NLP involves interaction between computers and humans using natural (i.e., human) language in order “to read, decipher, understand, and make sense of the human languages in a manner that is valuable.”¹⁹ Such interaction can include review of **structured data** (i.e., clearly defined types of data, such as addresses, that are easily searchable in relational databases) and **unstructured data** (i.e., various types of less defined categories of data, such as text files and email, that are more difficult to search).²⁰

Machine learning is already becoming well-established in the corporate world to address various types of financial crime compliance, such as fraud, identity theft, and anti-money laundering (AML).²¹ With regard to fraud, AI and machine learning platforms reportedly

are capable of combining supervised and unsupervised machine learning that can deliver a weighted score for any digital business’ activity in less than a second. . . . Fraud prevention systems can examine years and in some cases, decades of transaction data in a 250-millisecond response rate to calculate risk scores using AI.²²

One vendor that offers fraud detection solutions to banks asserted that it helped a well-known European bank reduce its false positives by 60 percent and increased actual fraud detection by 50 percent.²³

18 See MELANIE MITCHELL, *supra* note 13, at 178 and 180.

19 Dr. Michael J. Garbade, *A Simple Introduction to Natural Language Processing*, *Becoming Human*, October 15, 2018, <https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32>.

20 See, e.g., Christine Taylor, *Structured vs. Unstructured Data*, *Datamation*, March 28, 2018, <https://www.datamation.com/big-data/structured-vs-unstructured-data.html>.

21 See GUIDEHOUSE, *USING MACHINE LEARNING TO THWART FINANCIAL CRIME* (2020), <https://guidehouse.com/-/media/www/site/insights/financial-services/2020/ai-financial-crime-final.pdf>; *Enlisting AI And Biometrics In The Fight Against Digital Identity Theft*, PYMNTS, May 18, 2020, <https://www.pymnts.com/authentication/2020/enlisting-artificial-intelligence-biometrics-fight-against-digital-identity-theft/>; Chartis, *AI in RegTech: A quiet upheaval* (2018), available at <https://www.ibm.com/downloads/cas/NAJXEKE6>.

22 Louis Columbus, *Top 9 Ways Artificial Intelligence Prevents Fraud*, *Forbes*, July 9, 2019, <https://www.forbes.com/sites/louiscolumnbus/2019/07/09/top-9-ways-artificial-intelligence-prevents-fraud/#7fdca95514b4>. See Finextra with Feedzai, *Utilising AI to Prevent Financial Crime* (May 2019), <https://www.finextra.com/researcharticle/51/utilising-ai-to-prevent-financial-crime>.

23 Tibi Puiu, *Artificial intelligence for fraud detection is bound to save billions*, *ZME Science*, March 23, 2020, <https://www.zmescience.com/science/ai-fraud-detection-0942323/>.

Similarly, with regard to AML, financial institutions—in part because of regulators’ increasing inclination to encourage the use of machine learning to detect suspicious activity²⁴—are employing machine learning to improve the timeliness and accuracy of their customer due diligence and detection of money laundering-related activity. In one case, a global bank that regulators had directed to review some 20 million transactions dating back several years used a machine-learning approach that ultimately not only satisfied the regulator, but significantly decreased the number of alerts being generated and significantly increased the productivity of the remaining alerts.²⁵

Because machine learning has proven both feasible and valuable for other areas of compliance such as fraud and AML, a number of companies have also begun to consider whether they should pursue the use of machine learning in the context of anti-corruption. Some of those companies, however, have indicated that they do not have a clear understanding of how machine learning can be used in the anti-corruption context, or how they should explore that issue in a systematic fashion .

D. Purpose of This Guidance

This document is intended to provide companies in multiple sectors with guidance on whether and how they should consider developing or acquiring anti-corruption machine learning. Four developments indicate that such guidance is timely.

First, three companies in three different industries—Microsoft, AB InBev, and Alexion Pharmaceuticals—have already developed and deployed machine learning solutions in direct support of their anti-corruption compliance programs. (Details regarding several of

24 For example, since 2017 the United Kingdom Financial Conduct Authority has conducted a series of “TechSprints,” two-day sessions “that bring together participants from across and outside of financial services to develop technology based ideas or proof of concepts to address specific industry challenges.” Financial Conduct Authority, TechSprints (updated March 3, 2020), <https://www.fca.org.uk/firms/innovation/regtech/techsprints>.

25 Ellen Zimiles and Tim Mueller, *How AI is transforming the fight against money laundering*, World Economic Forum, January 17, 2019, <https://www.weforum.org/agenda/2019/01/how-ai-can-knock-the-starch-out-of-money-laundering/>. It is noteworthy that U.S. and United Kingdom regulators have been paying closer attention to the utility of anti-fraud and AML machine learning. See, e.g., United Kingdom Financial Conduct Authority, Global AML and Financial Crime TechSprint (reporting on May 2018 Techsprint), <https://www.fca.org.uk/events/techsprints/aml-financial-crime-international-techsprint>.

these companies' development and implementation of machine learning will be discussed in Section III G.)²⁶

Second, a number of non-governmental organizations have generally touted the use of AI by governments for anti-corruption purposes. For example, a U4 Anti-Corruption Resource Center report advocated the use of AI by government agencies “to uncover corruption that was previously difficult to detect” and to develop “novel artificial intelligence-assisted processes with the aim to avoid previously corruption-prone procedures.”²⁷ In particular, it proposed the use of AI and machine learning

to harmonise, categorise, or sort large datasets to select the records interesting for further investigation. The data may be a stream of input such as continuous transaction data, analysed in real time to flag suspicious transactions, or it may be millions of data-records sorted for detailed inspection, for example by tax authorities.²⁸

Similarly, a United Kingdom consulting firm issued a research report in which it suggested “that the conditions are right to test artificial intelligence tools for anti-corruption” in ten specified countries in North and South America and Europe because those countries “score highly in open data rankings, but have high levels of perceived corruption,” indicating that “there is a potentially large amount of data available for developers to use to train an AI anti-corruption tool.”²⁹

Third, academic studies have demonstrated the utility of machine learning in analyzing public data to make predictive judgments relating to corruption. One study used a neural-network approach (one facet of machine learning)³⁰ to predict public corruption based on economic

26 It should be noted that these companies are not the only ones that have deployed anti-corruption related machine learning solutions. Although this Guidance mentions these companies by name, any references herein to these companies, or other companies that are offering compliance-related ANI, are not intended to constitute an endorsement or recommendation of the specific solutions they have deployed.

27 Per Aarvik, *Artificial Intelligence—a promising anti-corruption tool in development settings?*, U4 Report 2019:1, <https://www.u4.no/publications/artificial-intelligence-a-promising-anti-corruption-tool-in-development-settings.pdf>.

28 *Id.*

29 OXFORD INSIGHTS, *THE NEXT GENERATION OF ANTI-CORRUPTION TOOLS: BIG DATA, OPEN DATA & ARTIFICIAL INTELLIGENCE* 3 (May 2019), <https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/5ced49ccc8302518cb27f64b/1559054797862/Research+report+2019+-The+Next+Generation+of+Anti-Corruption+Tools+-+Big+Data%2C+Open+Data+%26+-+Artificial+Intelligence.pdf>.

30 See Larry Hardesty, *Explained: Neural networks*, MIT News, April 14, 2017, <http://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>.

and political factors. Using data from Spanish provinces in which actual corruption cases were reported by the media or actually went to court between 2000 and 2012, the study found “that the taxation of real estate, economic growth, the increase in real estate prices, the growing number of deposit institutions and non-financial firms, and the same political party remaining in power for long periods seem to induce public corruption.”³¹ Another study used a network-based technique to analyze bills-voting data reflecting the votes of Brazilian congressmen over a 28-year period, and found a high degree of accuracy in identifying politicians who were arrested for or convicted of corruption or other financial crimes.³²

Fourth, anti-corruption advocacy organizations and investigative journalists have made effective use of ANI. In 2018, the anti-corruption organization Global Witness reported that it used machine learning for review of large volumes of satellite imagery and detection of the presence of small-scale mines in the Democratic Republic of the Congo. The best model that Global Witness developed was able to identify 79.9 percent of the known mines in the region.³³

In addition, the International Consortium of Investigative Journalists (ICIJ) made use of machine learning to review the so-called “Luanda Leaks”—a cache of more than 700,000 leaked documents, equivalent to 356 gigabytes—and examine how Isabel dos Santos, the daughter of Angola’s former president José Eduardo dos Santos, reportedly took hundreds of millions of dollars in public funds out of Angola.³⁴

31 Félix J. López-Iturriaga and Iván Pastor Sanz, *Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces*, 140 SOCIAL INDICATORS RESEARCH 975 (2017), available at <https://link.springer.com/article/10.1007%2Fs11205-017-1802-2>.

32 See Tiago Colliri and Liang Zhao, *Analyzing the Bills-Voting Dynamics and Predicting Corruption-Convictions Among Brazilian Congressmen Through Temporal Networks*, 9 SCIENTIFIC REPORTS 16754 (2019), available at <https://www.nature.com/articles/s41598-019-53252-9>.

33 See Sam Leon, *How Can We Use Artificial Intelligence to Help Us Fight Corruption in the Mining Sector?*, Global Witness, November 8, 2018, <https://www.globalwitness.org/en/blog/how-can-we-use-artificial-intelligence-help-us-fight-corruption-mining-sector/>. It should be noted that Global Witness acknowledged that the model also had a high false-positive rate, identifying only 48.4 percent of the pixels that were not part of known mining sites. As this guidance will discuss below, any use of machine learning for anti-corruption purposes must reduce both false positives and false negatives to an acceptable level to be considered fit for purpose.

34 Jeremy B. Merrill, *How Quartz used AI to sort through the Luanda Leaks*, Quartz, January 19, 2020, <https://qz.com/1786896/ai-for-investigations-sorting-through-the-luanda-leaks/>.

All of these developments indicate that there can be genuine value in using machine learning systems for anti-corruption purposes.³⁵ At present, however, it appears—based on interviews and emails with more than 40 companies, financial institutions, law firms, and non-governmental organizations—very few private-sector companies are using or offering machine learning products and services specific to anti-corruption.³⁶

This guidance has four objectives:

- (1) Identify what types of machine learning warrant consideration for anti-corruption risk and compliance;
- (2) Discuss various current uses of machine learning for anti-corruption purposes, including corporate uses for anti-corruption risk and compliance;
- (3) State and discuss key considerations in evaluating whether and how to pursue the use of machine learning for anti-corruption risk and compliance; and
- (4) Present recommendations and conclusions about what companies should do, particularly in the current economic environment, in considering whether to incorporate anti-corruption machine learning into its governance, risk, and compliance functions.

This guidance, in short, is intended to assist companies in thinking about whether and how to approach the task of deciding whether some form of anti-corruption machine learning would make sense, in operational and economic terms, for their anti-corruption programs. While it will briefly identify a variety of issues relevant to that decisional process, such as ethical and legal issues, it does not purport to be an exhaustive or authoritative discussion of those issues. A company considering the development and deployment of an anti-corruption machine learning solution should treat this guidance as a basic frame of reference, but

35 As indicated elsewhere in this document, companies should bear in mind that machine learning systems can address more than one area of compliance. Identification of suspicious high-dollar transactions, for example, may be indicative of insider risk and money laundering risk as well as bribery and corruption risk.

36 While the questions varied to some degree from one entity to another, the interviews of companies sought to explore each company's general perspective on using machine learning for anti-corruption purposes. Each interview typically included open-ended questions about whether the company was currently using anti-corruption machine learning and if not, whether it was considering doing so or had explored doing so; and what factors were influencing the company's decision not to pursue anti-corruption machine learning at present (e.g., company structure, cost and time needed to develop and implement an anti-corruption solution).

will need to conduct its own research and analysis on the issues identified herein before it can be sure that it has identified and explored all issues relevant to its business and the purposes for which it would deploy that solution.

II. Executive Summary

- ▶ Before starting any serious consideration of developing or acquiring an anti-corruption machine learning solution, a company must first articulate a business case for doing so (Section IIIA)
- ▶ Once it has determined that there is a business case for upgrading its anti-corruption program with a machine learning solution, a company should work to develop a solution in a series of five steps:
 - Framing a machine learning problem and proposing a solution. This step has six elements: (1) Frame the problem, which involves determining what prediction task the solution needs to perform; (2) See what data it has (including whether it has any data already labeled) that could be used in that solution, including what types of structured and unstructured data it has; (3) Design the data for the model; (4) Determine where the data come from; (5) Determine and prioritize easily obtained inputs, and; (6) Determine quantifiable outputs (Section IIIB);
 - Constructing a dataset, for which the company needs to begin with a dataset of sufficient size to begin training the model (Section IIIC);
 - Transforming the data, which can include changing data types, handling missing data, removing nonalphanumeric characters, and converting categorical data to numerical data (Section IIID);
 - Training the model, which involves taking data from its raw or normalized source state and transforming it into data that is ready for analysis (Section IIIE); and
 - Making predictions and assessing performance (Section IIIF)
- ▶ Three companies that provide examples of anti-corruption-related machine learning are AB InBev, Microsoft, and Alexion Pharmaceuticals (Section IIIG).
- ▶ Anti-corruption machine learning, like other areas of computing and IT, raises a wide range of ethical issues. These concern the need for responsible design and use of AI systems and ethical use of AI. Anti-corruption machine learning also raises legal

issues (i.e., data privacy, cybersecurity, and the use of machine learning for lawful purposes), and governance issues (e.g., whether to have an anti-corruption machine learning solution reach across the enterprise to obtain a dataset of sufficient size to generate the most useful predictions (Section IV)).

III. Evaluating the Use of Machine Learning in Anti-Corruption Risk and Compliance Functions: Analysis and Recommendations

As stated in the Introduction, this paper identifies and discusses a number of key issues that companies need to consider in deciding whether they should develop or acquire some form of machine learning for incorporation into their anti-corruption compliance programs.

At the outset, it is important for companies to recognize that machine learning, in general, is not a panacea for all elements of an anti-corruption compliance program. Machine learning makes sense only when very large datasets must be reviewed and corruption-related risks identified in a more timely manner than human-only review or rule-based programming can provide.

For that reason, a number of the hallmarks of effective compliance programs—commitment from senior management and a clearly articulated policy against corruption; a code of conduct and compliance policies and procedures; adequate oversight, autonomy, and resources; training and continuing advice; and incentives and disciplinary measures³⁷—are generally not elements in which machine learning would have practical utility.³⁸ Machine learning's great strength is in efficiently reviewing and finding associations between data in very large datasets to improve identifying, monitoring, and acting on higher-risk transactions and relationships, both for general compliance oversight and for internal investigations.

Several hallmarks, however, could benefit from the application of machine learning in anti-corruption programs:

- ▶ **Risk Assessment:** Assessment of risk—including both enterprise risk and client/transactional risk—is fundamental to developing and maintaining a strong anti-

³⁷ See FCPA RESOURCE GUIDE, *supra* note 5, at 58-67.

³⁸ It should be noted that ANI can be useful with regard to certain aspects of training (e.g., training data such as test score data and time to complete training), and that some rule-based ANI may be useful in evaluating the program and contributing to effectiveness ratings.

corruption program.³⁹ For firms with higher volumes of transactions and relationships, machine learning can significantly assist in timely identification of higher-risk transactions and relationships, including contracts and transactions with third parties.⁴⁰ In order to calibrate the appropriate level of due diligence that it should conduct, any company needs to draw on a range of internal datasets reflective of specific industries and countries associated with particular transactions, as well as the size and nature of those transactions, gifts and entertainment expenses related to those transactions, and the methods and amounts of compensation to third parties.⁴¹ Machine learning and natural language processing (NLP) can incorporate large volumes of data from those diverse datasets to review, synthesize, and present data more efficiently than human beings.

- ▶ **Third-Party Due Diligence and Payments:** Companies that use large numbers of agents, distributors, and consultants need, as part of their third-party risk management process, to take account of their third parties' qualifications, associations, and behaviors, both in initial onboarding and ongoing monitoring of their third-party relationships.⁴² For those companies, machine learning and NLP can make use of internal and external datasets to maintain suitable levels of due diligence in both onboarding and ongoing monitoring, including continuous monitoring.
- ▶ **Continuous Improvement and Periodic Testing:** As businesses constantly evolve—due to changes in customers, markets, products, services, and relationships—enforcement agencies expect anti-corruption programs to evolve constantly as well.⁴³ Machine learning can help companies with large numbers of such customers, markets, products, services, and relationships in detecting new or changing risks,

39 See FCPA RESOURCE GUIDE, *supra* note 5, at 60.

40 See Stuart Brock, *Legal Contracts on the Frontline of Fighting Corruption with Artificial Intelligence*, International Banker, May 21, 2019, <https://internationalbanker.com/technology/legal-contracts-on-the-frontline-of-fighting-corruption-with-artificial-intelligence/>

41 See FCPA RESOURCE GUIDE, *supra* note 5, at 60. It should be noted that the extent to which data are accessible and managed in a way that analysis can be performed should be an integral part of due diligence.

42 See *Id.* at 62.

43 See *Id.* at 66.

testing their internal controls, and adjusting their compliance programs to keep pace with their businesses' evolution.

A. Stating the Business Case for Anti-Corruption Machine Learning

Before starting any serious consideration of developing or acquiring an anti-corruption machine learning solution, a company must first articulate a business case for doing so: i.e., determine that there is sufficient justification for adopting and implementing machine learning for anti-corruption purposes, based on its unique risk profile and a frank evaluation of the benefits, costs, and risks of that machine learning.⁴⁴ No law or regulation requires any company to adopt anti-corruption machine learning based on media hype or industry “buzz” about AI in the abstract.

As in other aspects of business decision-making, a company needs to satisfy itself that there is a genuine business case for adopting anti-corruption machine learning. That is especially true at a time when businesses in many countries continue to experience severe financial pressures because of the COVID-19 pandemic.

In developing the business case for anti-corruption machine learning, a company should bear in mind four general considerations.

1. Rule-Based ANI Versus Machine Learning

Many companies have relied on **rule-based ANI** for their anti-corruption programs. In simple terms, rule-based ANI “produces pre-defined outcomes that are based on a set of certain rules coded by humans. These systems are simple artificial intelligence models which utilize the rule of if-then coding statements.”⁴⁵ Rule-based ANI can be viewed as best suited to

44 See *What is a business case?*, Association for Project Management, <https://www.apm.org.uk/resources/what-is-project-management/what-is-a-business-case/>.

45 Robert Smith, *The Key Differences Between Rule-Based AI And Machine Learning*, *Becoming Human*, July 14, 2020, <https://becominghuman.ai/the-key-differences-between-rule-based-ai-and-machine-learning-8792e545e6>.

situations in which the company is reviewing lower volumes of data and the ANI rules are relatively simple. Many companies use such rule-based approaches for expense approvals, in which the company defines the dollar thresholds for which management approvals are required at various levels.⁴⁶

In contrast, machine learning does not rely on human beings to write the rules for inclusion in the programming, but uses a suitable algorithm that

analyzes the dataset and determines relationships within that data; logic is embedded in the algorithm and was not coded by a human. As referenced in the name, the model trains itself and learns from the data, creating a cohesive relationship between data inferences and future data outputs.⁴⁷

Machine learning is better suited for situations in which a company has a large dataset and a large number of factors (i.e., columns in the data set), because a machine learning model will be more likely to identify relationships in the data than rule-based programming, for which human beings must identify particular relationships of interest and manually write rules for each relationship.⁴⁸

A company currently using a rule-based AI system for its anti-corruption program should look closely at that system before actively exploring the use of machine learning. Among other considerations, it must determine whether that system is already accessing all relevant data sources from within the company (and possibly external data sources as well) and generating results from which its compliance team can make timely determinations about which transactions and relationships warrant further inquiry or even filing of Suspicious Activity Reports. A rule-based system, if it is not drawing on all relevant data sources, may not be an effective component of that program, and therefore may be a vulnerability of concern to regulatory agencies.⁴⁹ A company should therefore consider whether the problem

46 Joseph M. Carew, *How to choose between a rules-based vs. machine learning system*, Tech Target, July 23, 2020, <https://searchenterpriseai.techtarget.com/feature/How-to-choose-between-a-rules-based-vs-machine-learning-system>.

47 *Id.*

48 *Id.*

49 See Jim Richards, *Rules-Based Monitoring, Alert to SAR Ratios, and False Positive Rates—Are We Having The Right Conversations?*, RegTech Consulting, <https://regtechconsulting.net/uncategorized/rules-based-monitoring-alert-to-sar-ratios-and-false-positive-rates-are-we-having-the-right-conversations/>.

it seeks to solve would be difficult to solve with rule-based programming, which is likely to work better on problems for which there is only a small number of fixed outcomes (e.g., Yes/No) or for which the penalty of error is too great to risk false positives or negatives.⁵⁰

2. Staffing, Training, and Experience

Second, the company must have an adequate complement of risk and compliance professionals who would have the necessary training and expertise to make effective use of a machine learning solution's output.⁵¹ However quickly rule-based programming or machine learning may generate outputs, neither will meet regulatory expectations if the company has insufficient staffing to make timely and effective use of the output from either system.

3. Scope of the Solution

Third, the company will need to determine whether a machine-learning solution should address a broader range of its risks than anti-corruption. Any corporate third-party risk management program needs to factor in a variety of risks that may arise at the outset of and through the entire lifecycle of third-party relationships: strategic, operational, compliance, and information security risks, to name just a few. It would make little sense for a company to build a standalone solution addressing only third-party bribery and corruption risks, while using other solutions for other categories of third-party risks. At the same time, a company may need to consider starting with a smaller and more discrete solution and make sure that that solution is working effectively before expanding its capabilities.

4. Cost and Return on Investment of Machine Learning Solutions

Fourth, even if a company believes that it has the need, the staffing, the dataset sizes, and the scope to make machine learning (rather than rule-based) a potential solution for its anti-corruption program, it should expect that building that solution may involve significant costs.

⁵⁰ See Google, *Identifying Good Problems for ML*, in *Introduction to Machine Learning Problem Framing*, <https://developers.google.com/machine-learning/problem-framing/good>; Elana Krasner, *supra* note 15.

⁵¹ See Jim Richards, *supra* note 49.

In part because of the complexity of building a solution that is crafted to meet a particular company's unique risk and compliance needs, external vendors' pricing of such a solution (according to some companies contacted for this guidance) may range from the low six figures to several millions of dollars.

Companies, however, should consider both the potential cost and the potential return on investment that a machine learning solution could entail. In the end, such an expenditure may be warranted if a company concludes that (a) its current-state systems cost it multiple millions of dollars without reducing the relevant types of risk and (b) a machine learning solution may significantly reduce compliance costs while increasing the program's effectiveness.

B. Framing the Problem

Once it has determined that there is a business case for upgrading its anti-corruption program with machine learning, the company's development of that solution should proceed in a series of five major steps. The first of those steps is to frame (i.e., define) the problem.⁵² That step consists of six elements.⁵³

1. Articulate the Problem

This element requires the company to define the anti-corruption problem with reference to machine learning as precisely as it can. In particular, the company needs to determine what prediction task it needs to perform. In machine learning, "prediction" refers to the output an algorithm produces, after training of that algorithm on a historical dataset and application of the algorithm to current data, to forecast the likelihood of a particular outcome, such as

52 See, e.g., Paloma Cantero-Gomez, *How To Frame A Problem To Find The Right Solution*, Forbes, April 10, 2019, <https://www.forbes.com/sites/palomacanterogomez/2019/04/10/how-to-frame-a-problem-to-find-the-right-solution/?sh=2e7133dc5993>.

53 Google, *Introduction to Machine Learning Problem Framing*, <https://developers.google.com/machine-learning/problem-framing/>.

whether particular payments or transactions raise concerns about possible connections to bribery.⁵⁴

Some common machine learning approaches include:

- ▶ **Clustering:** Clustering involves grouping similar values, such as customers with similar characteristics.⁵⁵
- ▶ **Anomaly detection:** Anomaly detection (also known as outlier analysis) involves “searching for and identifying instances that do not conform to the typical data in a data set.”⁵⁶
- ▶ **Classification:** Classification involves selecting one of a defined number of labels (e.g., apple, peach, pear).
 - **Regression:** Regression involves predicting certain values.
 - **Association rule learning:** Association rule learning involves Inferring likely association patterns in data.⁵⁷

54 See W. James Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu, *Definitions, methods, and applications in interpretable machine learning*, 44 Proceedings of the National Academy of Sciences 22071 (October 16, 2019), <https://www.pnas.org/content/116/44/22071>; Prediction, Data Robot, <https://www.datarobot.com/wiki/prediction/#:~:text=%E2%80%9CPrediction%E2%80%9D%20refers%20to%20the%20output.will%20churn%20in%2030%20days.&text=The%20word%20%E2%80%9Cprediction%E2%80%9D%20can%20be%20misleading>.

55 JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 153.

56 *Id.* at 160. Rule-based approaches to anomaly detection have been used in a variety of contexts, such as fraud detection, for some time. But [t]he main drawback with a rule-based approach to anomaly detection is that defining rules in this way means that anomalous events can be identified only after they have occurred and have come to the company’s attention. . . . In some ways, anomaly detection is the opposite of clustering: the goal of clustering is to identify groups of similar instances, whereas the goal of anomaly detection is to find instances that are dissimilar to the rest of the data in the data set.

Id. 161-162. In the context of anti-corruption, clustering could assist in identifying high-risk third parties in specific jurisdictions, while anomaly detection could assist in identifying payments of unusual size to those high-risk third parties. Deep learning anomaly detection solutions may be particularly effective for such purposes. See Xuning (Mike) Tang and Yihua Astle, *The Impact of Deep Learning on Anomaly Detection*, Law.com, August 10, 2020, <https://www.law.com/legaltechnews/2020/08/10/the-impact-of-deep-learning-on-anomaly-detection/>.

57 See Google, *Common ML Problems*, in *Introduction to Machine Learning Problem Framing*, <https://developers.google.com/machine-learning/problem-framing/cases>. With regard to classification, there is a wide variety of classification algorithms that data scientists can use, such as decision trees, Naïve Bayes, and artificial neural networks. There is no single “correct” classification algorithm about which a company can decide in advance is best for its anti-corruption machine learning solution. That decision will depend on the application it intends for the solution and the nature of the dataset it proposes to use. See Sidath Asiri, *Machine Learning Classifiers*, Towards Data Science, June 11, 2018, <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623>.

Moreover, the company needs to consider what kinds of predictions it wants the anti-corruption solution to make. Among other possibilities, a company may be interested in identifying some or all of the following types of predictions along a scale of increasing corruption-related risk: (1) transactions and relationships that violate corporate policies, but that may or may not be corrupt; (2) transactions and relationships that are likely corrupt, whether or not they violate corporate policies; (3) transactions and relationships that are suspicious, but that require additional investigation; and (4) find patterns of behaviors that are suspicious or identify high-risk transactions or relationships (e.g., dealings with Specially Designated Nationals, or with foreign officials known to be involved in significant corruption, who are on sanctions lists⁵⁸).

Accordingly, as part of its effort to articulate the problem, the company should draw on its bribery and corruption risk assessment process to identify and rank its distinctive risks (e.g., insider risks and third-party relationships), and should make an initial judgment whether the relevant dataset is large enough to be a suitable candidate for the machine learning applications under consideration. It is important, however, for a company to know the problem it needs to solve before focusing on the data that could be used in a machine learning solution.⁵⁹

HYPOTHETICAL EXAMPLE: Nassau Systems, a U.S.-based developer, manufacturer, and seller of computers and related products and services, recently acquired FitzRandolph Technology, a United Kingdom-based manufacturer and seller of computers and related products in the European Union and Asia. Nassau Systems' senior leadership views FitzRandolph as a gateway to expanding its distribution of computer sales beyond the United States, but also has decided to undertake a substantial expansion of its existing network of third-party relationships.

From participation in meetings with Nassau Systems' key business leadership, Nassau Systems' chief compliance officer (CCO) is aware that the company strategy, after the FitzRandolph acquisition, is to increase substantially the number and variety of its third-party relationships in multiple regions. In particular, the strategy includes

58 See, e.g., U.S. Department of the Treasury, Specially Designated Nationals List (updated March 25, 2021), <https://home.treasury.gov/policy-issues/financial-sanctions/specially-designated-nationals-and-blocked-persons-list-sdn-human-readable-lists>.

59 See Google, *Introduction to Machine Learning Problem Framing*, *supra* note 53.

greater leveraging of its third-party resellers, allowing those resellers to enjoy greater autonomy and freedom to achieve sales goals and providing an incentive structure to do so. She also is aware that the strategy gives them only 24 months in which to make any necessary changes to the company's anti-bribery and corruption (ABC) compliance program.

Tasks: In this initial phase of framing the problem, the CCO must consider the combined effects of the FitzRandolph acquisition and the substantial expansion of Nassau Systems' third-party network on the current capacity of Nassau Systems' existing ABC program to monitor third-party relationships and take appropriate action on relationships and transactions with third parties that pose higher bribery and corruption risk. Because Nassau Systems to date has relied on risk-based programming to identify such risks, the CCO now needs to consider whether to pursue a machine learning solution that can be implemented within the 24-month timeframe for the new business strategy.

Among other considerations that can favor a machine learning solution for Nassau Systems, the CCO should consider –

- The sheer size and scale of the company's planned expansion of its third-party relationships, which will involve Nassau Systems' current third-party network and its plans to increase business through the FitzRandolph existing distribution and reseller network or newly designated Nassau Systems distributors and resellers;
- The number and size of jurisdictions in which Nassau Systems wants to do business through third parties;
- The degree to which those jurisdictions pose enhanced bribery and corruption risk; and

The volume and variety of data that potentially can be tapped within Nassau Systems' and FitzRandolph's systems.

2. See What Data It Has That Could Be Used in That Solution

The company should next see what data it has that could be used in that solution, including whether it has any data already labeled and what types of structured and unstructured data it has. One type of machine learning to be considered as part of a solution is supervised learning, in which human beings provide (or assist in providing) labels for instances in the relevant dataset. To do so may require substantial human effort, not only for collecting, curating, and labeling the data, but also for designing the architecture of a machine learning solution.⁶⁰ For example, a company may need to decide how to label very different types of data, such as accounts payable and contract terms, recognizing that the labels indicative of corruption in some contexts (e.g., terms such as “commitment fee” or “management fee”) may not be suitable in other contexts.

HYPOTHETICAL EXAMPLE:

Tasks: In this step, the CCO and her team, as well as others from the merged company, would need to take a complete inventory of all data that could potentially be useful to address the problem she has identified. The object, in this step, is not to make final decisions about which data to include in a machine learning solution, but simply to identify any and all categories of data that might help in constructing that solution. Nassau Systems should therefore cast its net broadly to include, from both its and the merged FitzRandolph systems, such data categories as spend data, sales data, and third-party transaction data, information from prior due diligence on third parties and investigations pertaining to third parties. It also should make initial determinations of the volume of data under each of the categories it has identified, and which parts of the business own those data.

60 MELANIE MITCHELL, *supra* note 13, at 97.

3. Design the Data for the Model

This element requires a company to anticipate how data for the model should be designed to make useful predictions.⁶¹ Among other challenges in this process, instances in the data set must be labeled with the value of the target **attribute**.

Often, however, the reason a target attribute is interesting is that it is not easy to directly measure, and therefore it is not possible to easily create a data set of labeled instances. In such scenarios, a great deal of time and effort is required to create a data set with the target values before a model can be trained using supervised learning.⁶²

HYPOTHETICAL EXAMPLE:

Tasks: In this step, Nassau Systems should be focusing on culling the excess data from the categories of data in its inventory. This process should include determining how useful the various data categories would be for its ABC compliance program, how easy or difficult it would be to access certain data categories, whether it needs the full volume of data in those categories (e.g., whether the company needs 10 or more years of prior data, rather than 5 or fewer years), and how much effort it would take to “clean” those data (i.e., to identify and correct errors in the dataset that may negatively impact a machine learning model). For example, because Nassau Systems will need to identify instances in which its managers have provided meals or travel expenses to third parties who have current or previous relationships with a foreign government, it should consider factors such as whether it needs to review such data in current data systems, or to access any FitzRandolph legacy data systems, for payments made more than five years ago.

61 See Google, *Formulate Your Problem as an ML Problem*, in *Introduction to Machine Learning Problem Framing*, <https://developers.google.com/machine-learning/problem-framing/formulate>.

62 JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 100 (emphasis supplied).

4. Determine Where the Data Come From

Identifying all sources of data, both structured and unstructured, that the company would potentially include in constructing and testing a possible model is critical. Among other considerations, if a model were to draw on data that have little or no relevance to the specific problem under consideration, it could potentially find patterns that human beings could not, but also could generate predictions that have a high number of false positives and negatives. Moreover, as discussed below, certain repositories of corporate data pertinent to an anti-corruption solution may be subject to legal constraints on the use of such data.

HYPOTHETICAL EXAMPLE:

Tasks: In this step, which is closely related to the preceding step, Nassau Systems should identify which specific parts of the business own the data in which it is interested for development of its machine learning solution. It should be noted that this step may require the company to consult with other components of the business, and potentially even with its current third parties.

5. Determine and Prioritize Easily Obtained Inputs

Especially at the outset of problem framing, no company needs to try to identify all potential inputs for the model it is developing. Its initial focus can be on one to three inputs that it can easily obtain (e.g., data relating to gifts and entertainment expenses and contracts with prospective and current third parties) and that it believes would produce a reasonable initial outcome.⁶³

HYPOTHETICAL EXAMPLE:

Tasks: This step requires Nassau Systems to decide which data categories and datasets are more important and more reliable for the problem to which it intends to direct the machine learning solution. As part of this step, the company will need to be clear about the potential limitations of certain datasets and whether it would be worth the

⁶³ See Google, *Formulate Your Problem as an ML Problem*, *supra* note 61.

investment to incorporate those data into its solution. For example, if Nassau finds that a FitzRandolph legacy data system contains corrupted data for expenditures more than three years old, it may decide that it is not worth the cost of separating corrupted from uncorrupted data.

6. Determine Quantifiable Outputs

Terms such as “artificial intelligence” and “deep learning” may prompt some to think that a machine learning solution can replicate many types of human judgment and decision making. Any machine learning model, however, is limited to generating quantifiable outputs. Algorithms and neural networks cannot produce qualitative outputs (e.g., “Should we maintain our relationship with Third-Party X?” or “Did our Vice President know that the payments he authorized were bribe payments?”). A company developing a machine learning model, therefore, needs to decide what kinds of quantifiable outputs (e.g., numbers, labels, or clusters) it needs for specific purposes such as risk scoring or ranking.⁶⁴

HYPOTHETICAL EXAMPLE:

Tasks: In this step, Nassau Systems would need to see that the machine learning solution it is considering would produce quantifiable outputs that are actionable in relation to the problem it identified at the outset. For example, if the proposed solution is directed at predicting bribery and corruption-related risk regarding third parties, Nassau Systems would want to make sure that its potential outputs—such as a confidence score indicating the likelihood of high risk, or an aggregate risk score for particular third parties—are actionable for particular functions of its compliance program. Those functions could include prospective identification of high-risk third parties that have reseller agreements with the company, or establishing different audit procedures that vary with the degree of risk that particular third parties pose.

64 See Google, *Deciding on ML*, in *Introduction to Machine Learning Problem Framing*, <https://developers.google.com/machine-learning/problem-framing/framing>.

C. Constructing the Dataset

One of the challenges for a company in developing a machine learning model for an anti-corruption solution is that there is no guarantee at the outset that the model will prove to be usable.⁶⁵ Even so, the company must begin with a dataset of sufficient size to begin training the model. The next step in developing a machine learning solution, then, is to construct the dataset.

Anti-corruption machine learning applications, however programmed, are unlikely to be successful if they do not have fairly large volumes of data (whether structured, unstructured, internal, or external) for its algorithms. As a general proposition, the larger the dataset, the more complex the patterns that an anti-corruption machine learning solution can detect.⁶⁶

Large multinational companies doing business in multiple markets are the most likely potential candidates for anti-corruption machine learning solutions, though other companies may be able to amass comparable volumes of relevant data. For example, in the case of Teva Pharmaceutical, the company had more than 500,000 vendors and customers — including agents, consultants, distributors, suppliers, vendors, and other entities and individual facilitating or involved in any transaction with government officials or agencies—for which it needed to evaluate risks pertaining to the U.S. Foreign Corrupt Practices Act (FCPA).⁶⁷

HYPOTHETICAL EXAMPLE:

Tasks: Once Nassau Systems has completed the six steps of the problem-framing process, it should turn to construction of the dataset (or datasets) on which its machine learning solution will draw. This phase of the machine learning process involves bringing together all of the data available from throughout the company that are (a) relevant to the problem, (b) accessible with reasonable effort, and (c) clean or capable of being cleaned within an acceptable timeframe and cost.

65 See Google, *The ML Mindset*, in *Introduction to Machine Learning Problem Framing*, <https://developers.google.com/machine-learning/problem-framing/big-questions>.

66 See Rodney Weidemann, *Using AI to uncover fraud and corruption*, ITWeb, February 21, 2019, <https://www.itweb.co.za/content/KBpdgvpPDzavLEew>.

67 EY, *Teva Pharmaceutical Reengineers Compliance With Data Analytics*, MIT Sloan Management Review, July 8, 2020, <https://sloanreview.mit.edu/sponsors-content/teva-pharmaceutical-reengineers-compliance-with-data-analytics/>.

D. Transforming the Data

Before a company can proceed to train an anti-corruption machine learning model, the next step in its development process is to engage in data transformation.⁶⁸ Data transformation is a process in which the company takes data from its raw or normalized source state “and transform[s] it into data that’s joined together, dimensionally modeled, de-normalized, and ready for analysis.”⁶⁹

Although a complete description of a data-transformation process is beyond the scope of this document, the steps involved in data transformation can include changing data types, handling missing data, removing nonalphanumeric characters, and converting categorical data to numerical data. The point of data transformation is to “ensure maximum data quality which is imperative to gaining accurate analysis, leading to valuable insights that will eventually empower data-driven decisions.”⁷⁰

HYPOTHETICAL EXAMPLE:

Tasks: This phase of the machine learning process involves “normalization” of the data in the relevant datasets. Normalization refers, in part, to bringing together the same types of data that are recorded differently in different corporate databases. For example, transaction dates in the United States typically follow a month/day/year format, while transaction dates in other regions of the world may follow a day/month/year format.

In this phase, it would be important for Nassau Systems’ compliance team to maintain regular dialogue with the data scientists working on the planned solution. In some situations, data scientists may need to consult with the project leads if they see that certain data proposed for use are less clean than expected, but that could be useful to the machine learning model if they could have additional time to conduct the necessary cleaning.

⁶⁸ See Google, *Introduction to Machine Learning Problem Framing*, *supra* note 53.

⁶⁹ Damian Chan, *Why You Need Data Transformation in Machine Learning*, Datanami, November 8, 2019, <https://www.datanami.com/2019/11/08/why-you-need-data-transformation-in-machine-learning/>.

⁷⁰ *Id.*

E. Training the Model

The next step in building an anti-corruption machine learning model is to train the model itself.⁷¹ While machine learning discussions often refer to “the” dataset to be used, a company should recognize that there are three distinct datasets that are part of the training process:

- ▶ **The training set:** This set, which one analyst has said constitutes the majority of the total data (around 60 percent), is also known as the historical data set. This set “is the one used to train an algorithm to understand how to apply concepts such as neural networks, to learn and produce results. It includes both input data and the expected output.”⁷²
- ▶ **The validation set:** This set is used “to select and tune the final Machine Learning model.”⁷³
- ▶ **The testing set:** This set is a subset to test the trained model.⁷⁴

The training phase is the critical point at which a company must ensure that the dataset on which it intends to draw for training, validating, and testing is sufficiently large to contribute to valid predictions. At least two significant problems can stem from a too-small dataset.

1. Class Imbalance

Class imbalance is a circumstance in which there is one overrepresented class and one heavily underrepresented class in the dataset, and the task that the company has set for a machine learning solution is to detect a rare event. In essence, class imbalance is a lack of data diversity—that is, there is insufficient breadth and variety in the data labels and related attributes for effective training of the model on the variety of scenarios that it is expected to

⁷¹ See Google, *Introduction to Machine Learning Problem Framing*, *supra* note 53.

⁷² Alexandre Gonfalonieri, *How to Build A Data Set For Your Machine Learning Project*, Towards Data Science, February 13, 2019, <https://towardsdatascience.com/how-to-build-a-data-set-for-your-machine-learning-project-5b3b871881ac>.

⁷³ *Id.*

⁷⁴ See Google, *Training and Test Sets: Splitting Data*, in *Machine Learning Crash Course*, <https://developers.google.com/machine-learning/crash-course/training-and-test-sets/splitting-data>.

analyze and understand. One data scientist has estimated that typically, an entity is facing class imbalance when the proportion of the two classes is lower than 10/90.⁷⁵

Detection of financial crimes is one area where class imbalance can arise:

It's handy to use AI for fraud detection, for example, a situation where it can weigh the subtleties of millions of online transactions and look for signs of suspicious activity. But suspicious activity is so rare compared to normal activity that people have to be careful that their AIs don't conclude that fraud never happens.⁷⁶

Bribery and corruption is a category of financial crime that may pose an even greater potential for class imbalance than fraud or money laundering. In contrast to money laundering, in which a criminal organization typically needs to process large volumes of criminal proceeds at a high velocity and frequency during the placement phase of laundering,⁷⁷ corporate payments of bribes typically take place infrequently at irregular intervals.

Class imbalance is not an insuperable problem.⁷⁸ Data scientists can use several approaches to fixing class imbalance. One such approach is a simple 2 x 2 matrix, known as a **confusion matrix**, that aids in measuring performance in a machine learning solution devoted to classification. A confusion matrix displays the numbers of true and false positives and true and false negatives.

Figure 1 - Example of Confusion Matrix⁷⁹

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FNP	TN

75 Anastasia Gorina, *Class imbalance problem in classification*, Towards Data Science, April 2, 2020, <https://towardsdatascience.com/class-imbalance-problem-in-classification-a2ddaba98f4a>.

76 JANELLE SHANE, *YOU LOOK LIKE A THING AND I LOVE YOU* 169 (2019).

77 See, e.g., United Nations Office on Drugs and Crime, *Introduction to money-laundering*, <https://www.unodc.org/unodc/en/money-laundering/overview.html>.

78 See Syed Sadat Nazrul, *Fraud Detection Under Extreme Class Imbalance*, Towards Data Science, April 12, 2018, <https://towardsdatascience.com/fraud-detection-under-extreme-class-imbalance-c241854e60c>.

79 Sarang Narkhede, *Understanding Confusion Matrix*, Towards Data Science, May 9, 2018, <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>.

Because it clearly displays the total numbers of true positives, false positives, true negatives, and false negatives, a confusion matrix can be especially useful in measuring accuracy, precision, and recall.⁸⁰

Another approach to dealing with class imbalance is to change the training data in some fashion that creates “roughly equal numbers of training examples in each category.”⁸¹ Using a technique known as *data augmentation*, programmers “make small changes to the data so that one bit of data becomes many slightly different bits.”⁸²

2. Overfitting and Underfitting

The second problem involves two related concepts, **overfitting** and **underfitting**. Although circumstances can vary widely from company to company, class imbalance could be a particular challenge for a company that has experienced very few provable instances of corruption-related transactions. In that situation, a machine learning solution could end up predicting few or no instances of such transactions simply because they were so rare in the training data.⁸³

Overfitting is a term in statistics that refers to “a modeling error that occurs when a function corresponds too closely to a dataset.” In the context of machine learning, it describes a situation in which the model cannot generalize or fit well on a dataset other than the training dataset. “A clear sign of machine learning overfitting is if its error on the testing or validation dataset is much greater than the error on training dataset.”⁸⁴ A machine learning model can also have the opposite problem, underfitting, when the model does not do well with either the training dataset or the test dataset.

Data scientists have a number of approaches they can use to address overfitting or underfitting. For overfitting, these approaches include cross-validation, training with more

80 *Id.*

81 JANELLE SHANE, *supra* note 76, at 169.

82 *Id.* at 116.

83 *Id.* at 76.

84 Mayank Tripathi, *Underfitting and Overfitting in Machine Learning*, Data Science Foundation, June 13, 2020, <https://datascience.foundation/sciencewhitepaper/underfitting-and-overfitting-in-machine-learning>.

data, reducing the complexity of the model, regularization (i.e., artificially forcing the model to be simpler), and ensembling (i.e., using “machine learning methods for combining predictions from multiple separate models”). For underfitting, if the model does not perform well on either the training or the testing set, data scientists need to try alternative machine learning algorithms.⁸⁵

3. Bias

In addition to problems that can stem from too-small datasets, a company developing an anti-corruption machine learning must be attentive to another major problem that can affect any machine learning solution, regardless of the size of the dataset. That problem is **bias**.

In one sense, data are “always partial and biased,” in that they “are never an objective description of the world”:

Data are generated through a process of abstraction, so any data are the result of human decisions and choices. For every abstraction, somebody (or some set of people) will have made choices with regard to what to abstract from and what categories or measurements to use in the abstracted representation.⁸⁶

A company developing an anti-corruption machine learning solution, however, must be attentive to at least two broad categories of bias that can result in unreliable predictions. The first is bias that has the effect of unjustifiably discriminating against individuals based on categories such as gender, race, and ethnicity. For example, in 2018 a leading technology company that had been pretrialing a machine learning solution for prescreening job applicants found that the solution was discriminating against women, by penalizing “resumes from candidates who had gone to all-female schools” as well as resumes “that mentioned the word women’s—as in, ‘women’s soccer team’.”⁸⁷

Bias of this kind could potentially arise in building an anti-corruption machine learning solution. For example, a solution dedicated to evaluating risk related to onboarding of

⁸⁵ *Id.*

⁸⁶ JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 46.

⁸⁷ JANELLE SHANE, *supra* note 76, at 178-79.

third parties in various countries, including countries with majority nonwhite populations, might draw from external data such as corruption indices and surveys that rank the relative degrees of corruption in those countries. If not carefully crafted, an algorithm for such a solution might generate predictions that certain company transactions in a foreign country generally regarded as a high risk for bribery and corruption are high risk simply because all individuals in that foreign country have the same racial or ethnic heritage (or that company representatives doing business in that country are suspect because they have that same heritage).

As one research scientist put it, “Since humans tend to be biased, the algorithms that learn from them will also tend to be biased unless humans take extra care to find and remove the bias.”⁸⁸ That problem may be compounded if human beings engage in bias laundering (i.e., treating machine learning predictions as impartial simply because they came from AI).⁸⁹

The second category of bias is a broad group of biases that are not driven or influenced by legally impermissible discriminatory judgments. These include:

- (1) Sample bias:** This term “describes how the process used to select a data set can introduce biases into later analysis, be it a statistical analysis or the generation of predictive models using [machine learning].”⁹⁰ Because sample bias directly affects the accuracy of a model’s predictions, it “is a bias that a data scientist will try to avoid.”⁹¹
- (2) Learning bias:** This type of bias refers to the type of generalization that an algorithm encodes. This type of bias is inevitable in any machine learning solution, because machine learning algorithms “are biased to look for different types of patterns, and because there is no one learning bias across all situations.”⁹² The company crafting a machine learning solution must therefore choose what kinds of generalizations

88 *Id.* at 180.

89 *Id.* at 181.

90 JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 143.

91 *Id.* at 144.

92 *Id.*

it wants the algorithm to produce. To choose desired generalizations may require the company to try different algorithms with its model, or even to “buil[d] multiple models using different algorithms and compar[e] the models to identify which algorithm generates the best model.”⁹³

4. Accuracy, Precision, Recall, and F1 Score

Once it anticipates how to address the problems of class imbalance and bias, a company should expect to evaluate the performance of its model with reference to various metrics. Four metrics commonly used for assessing machine learning performance are:

- (1) Accuracy.** Accuracy can be defined simply as the fraction of correct predictions by the model (i.e., all true positives and true negatives, divided by the total of all true positives, true negatives, false positives, and false negatives [i.e., the sum of all predictions, whether true or false]).⁹⁴

With accuracy, the smaller the fraction (i.e., the greater the number of false positives and false negatives in the denominator), the less accurate the model is, and the larger the fraction (i.e., the smaller the number of false positives and false negatives), the more accurate the model is.

Although accuracy does provide one perspective on the value of machine learning models in general, it can have limitations as a metric in the context of an anti-corruption machine learning model. For example, if a large company with substantial international business is considering one or more anti-corruption models focused on financial payments to third-party intermediaries (TPIs), the actual number of true positives (i.e., suspect large payments made to a TPI with high corruption risk) is likely to be extremely small relative to the total of all payments being reviewed. The consequences of a model’s falsely classifying any suspect payments as negative,

93 *Id.*

94 Google, *Classification: Accuracy*, in *Machine Learning Crash Course*, <https://developers.google.com/machine-learning/crash-course/classification/accuracy>. As a practical matter, a company is likely to have greater success in the training process with calculating true positives and false positives because it can validate against true positives.

however, can be substantial,⁹⁵ so a company should not rely solely on accuracy as a performance metric.

- (2) Precision.** Precision can be defined simply as the proportion of true positives (i.e., positive identifications that were, in fact, correct), or the total number of true positives divided by the total of true positives and false positives. In essence, precision indicates how good the model's predictions are within the context of what the model actually identified. With precision, the smaller the fraction (i.e., the greater the number of true and false positives in the denominator), the less precise the model is, and the larger the fraction the more precise the model is. Thus, a model that produces no false positives will have a precision of 1.0.⁹⁶

For an anti-corruption machine learning model, precision can provide a useful perspective on the model's effectiveness. In the machine learning model described above for detecting suspect payments to TPIs with high corruption risk, those evaluating the model can focus on how well the model does at predicting such suspect payments. What precision does not address, however, are the false negatives that may be highly important for a company in detecting potential corruption.

- (3) Recall.** Recall can be defined as the proportion of actual positives that were defined correctly (e.g., a high percentage of correctly identified high-risk third-party contracts), or the total number of true positives divided by the total of true positives and false negatives. In contrast to precision, which addresses how good the model's predictions are, recall addresses how complete the model's predictions are (i.e., how much of what the model should have identified was actually identified). In general, the higher a model's recall is, the lower its precision is likely to be; because recall is concerned with completeness, that broader focus can increase the risk of false positives.

95 See Koo Ping Shung, *Accuracy, Precision, Recall or F1?*, Towards Data Science, March 15, 2018, <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>.

96 Google, *Classification: Precision and Recall*, in *Machine Learning Crash Course*, <https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall>.

As with accuracy and precision, with recall the smaller the fraction the less effective the model is and the larger the fraction the more effective the model is. Thus, a model that produces no false negatives will have a recall of 1.0.⁹⁷ For the anti-corruption model focused on suspect payments to high-risk TPIs, recall can add value to evaluating the model because it takes into account both true positives and false negatives.

- (4) **F1 Score.** While accuracy, precision, and recall all are valid metrics in evaluating a machine learning model, in general precision and recall are better for that evaluation. A conundrum in using precision and recall, however, is that the two metrics “are often in tension”: i.e., “improving precision typically reduces recall and vice versa.”⁹⁸ To address the potential inverse relationship between precision and recall, data scientists can use a fourth metric known as the **F1 score**.

The F1 score, simply put, takes both precision and recall into account in measuring the accuracy of the model, by giving more weight to false negatives and false positives, while not letting large numbers of true negatives affect the score. The reasoning behind the F1 approach is that a company may regard false positives and false negatives as crucial to its evaluation, but regard TN as less important to the problem it is seeking to solve.⁹⁹

Ultimately, no company can completely “solve” problems such as class imbalance and bias in developing an anti-corruption machine learning solution.¹⁰⁰ What it needs to do is address any class imbalance problem it can identify, avoid the types of bias described above that pose legal and ethical problems, and devise a solution that incorporates the right kind of learning bias to generate meaningful output with suitably high levels of accuracy, precision, and recall.

97 *Id.*

98 *Id.*

99 Christopher Riggio, *What’s the deal with Accuracy, Precision, Recall and F1?*, Towards Data Science, November 1, 2019, <https://towardsdatascience.com/whats-the-deal-with-accuracy-precision-recall-and-f1-f5d8b4db1021>.

100 See Julia Powles, *The Seductive Diversion of “Solving” Bias in Artificial Intelligence*, One Zero, December 7, 2018, <https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53>.

HYPOTHETICAL EXAMPLE:

Tasks: The Nassau Systems project team should expect that this phase, which includes both constructing and testing the machine learning model or models, will be the most time-intensive. No machine learning solution, no matter how familiar the types of algorithms may be to data scientists, can be expected to produce immediate results with satisfactory degrees of accuracy, precision, and recall.

The Nassau Systems project leads should therefore set clear expectations with their senior management on the speed with which the model or models can be trained to yield actionable results. They also should make sure to engage both their subject matter experts and the data scientists on a regular basis, so that the solution ultimately produces results that are not only accurate but actionable. As data scientists continue to train the model, subject-matter experts can help the project leads to determine whether the outputs that the model is generating are the kinds of outputs the experts would expect to see (e.g., the number of high-risk third parties whom the compliance program was able to action).

F. Making Predictions and Assessing Performance

The final step in developing an anti-corruption machine learning solution is to use the model to make predictions.¹⁰¹ This step also requires a company, on a continuing basis, to analyze the predictions its model is making, to satisfy itself that there is sufficient goodness of fit (i.e., how closely a model's predicted values match the observed or true values)¹⁰² and that there are no overfitting, underfitting, or bias concerns.

At all stages of development and implementation of an anti-corruption machine learning solution, the company must bear in mind that no machine learning model is static, and that its model will require retraining as new data become available and as external conditions change.¹⁰³ The company should also expect that refining the model may take a number of months before it can be confident that the model continues to make correct predictions at

101 Google, *Introduction to Machine Learning Problem Framing*, *supra* note 53.

102 See *Overfitting in Machine Learning: What It Is and How to Prevent It*, Elite Data Science, <https://elitedatascience.com/overfitting-in-machine-learning>.

103 MARIYA YAO, MARLENE JIA, AND ADELYN ZHOU, *APPLIED ARTIFICIAL INTELLIGENCE: A HANDBOOK FOR BUSINESS LEADERS* 149 (2018).

an acceptable rate, especially as the model may acquire newer, more recent data than it first encountered during initial training, validation, and testing.

For example, as described below in the case of AB InBev, the algorithm that AB InBev deployed to detect anomalies in accounts payable transactions had an early failure rate of over 99 percent. As AB InBev improved its cataloging of data in a cloud-based repository, however, and incorporated more sophisticated machine learning and other strategies, the solution's accuracy improved.¹⁰⁴

The company also needs to be prepared to continue to monitor and update its solution after deployment. That may include not only building an updated model, but analyzing the differences between any initial and later models it uses, “and how the update will affect the overall system quality and user experience.”¹⁰⁵ A company should expect that this phase of the process may be particularly challenging for its culture and change-management process. It may also involve the incorporation of new data sets or other information, which in turn requires a reconfiguration of the harmonization, analysis, visualization, and potentially even the workflow elements of the solution.

The company should expect that employees in first-, second-, and third-line positions with anti-corruption risk and compliance responsibilities will also need training on use of the new anti-corruption solution.¹⁰⁶ As the solution will be performing certain functions that previously were left to manual review by human employees, employees in risk and compliance functions will need to understand how the solution operates, how it can identify new relationships between datasets that human employees could not, and how it can make them more efficient in identifying, triaging, and acting as appropriate on suspect transactions and relationships. Those employees should also be expected to play a significant role in providing timely and continuous feedback on the performance of the solution.

104 Joe Williams, *How \$132 billion brewery giant AB InBev is using AI to fight corruption and spot business fraud around the globe*, Business Insider, December 13, 2019, <https://www.businessinsider.com/ab-inbev-brewing-artificial-intelligence-to-spot-fraud>.

105 Google, Responsible AI practices, <https://ai.google/responsibilities/responsible-ai-practices/>.

106 See Jaclyn Jaeger, *Six steps for developing an AI ethics framework*, Compliance Week, November 20, 2019, <https://www.complianceweek.com/artificial-intelligence/six-steps-for-developing-an-ai-ethics-framework/28078.article>.

Finally, the company should expect that it will need to document each step of its development, implementation, and revision of the solution, so that it is prepared to demonstrate the operations and effectiveness of the solution for auditors and regulatory agencies as needed. As indicated above, regulators are increasingly expecting companies to have direct or indirect access to relevant sources of data for compliance monitoring and testing.¹⁰⁷ For that reason, a company should be prepared to explain to regulators how it considered, decided on, and implemented its anti-corruption solution, and to produce relevant documentation for each of those phases of the process that can help to demonstrate to auditors and regulators the effectiveness of the solution.

HYPOTHETICAL EXAMPLE:

Tasks: If all has gone well in the preceding phases of the process, in this final stage the Nassau Systems project leads should be confident that the solution is generating accurate and reliable outputs that are actionable for the needs of the ABC compliance program. It is important for Nassau Systems management, however, to recognize that the version of the machine learning solution that it can now deploy may require further revision, training, and testing. As the company establishes and expands its third-party relationships, those changes can significantly add to existing datasets and generate new datasets that must be incorporated into that solution. No machine solution, in short, should be expected to remain static if the company's business activities are dynamic and constantly changing.

Moreover, even at this stage Nassau Systems senior management should be mindful of the need to ensure the continuing reliability of the data essential to its machine learning solution. Even if data are clean, relevant, and accessible, those data may not be a suitable candidate for inclusion in the solution if they are not reliable. For example, the Nassau Systems machine learning team may find certain data sets unreliable if they are imprecise, based largely on inconsistent manual recordkeeping, estimates, or hunches, or contain too many errors or incorrect information.

¹⁰⁷ See Section III *supra*.

G. Examples

1. AB InBev

AB InBev is the largest brewer in the world, with approximately 170,000 employees in more than 80 countries. Co-headquartered in Belgium and New York, it operates approximately 200 breweries and has approximately 50 operations in nearly 50 countries.¹⁰⁸ Like other global beverage companies, AB InBev depends on a far-flung network of distributors and vendors to deliver its products to consumers.¹⁰⁹

AB InBev's approach to anti-corruption machine learning stemmed from its \$100 billion-plus acquisition of SABMiller in 2015. That acquisition required AB InBev to integrate SABMiller's compliance program, which covered that firm's operations in 25 countries. As a result of that acquisition, the combined AB InBev was employing people in 57 countries and vendor partnerships in nearly 150 additional countries. Because that expansion meant that AB InBev would be subject to numerous versions of anti-bribery and antitrust legislation, each with different requirements and obligations, AB InBev's leadership sought "to more uniformly manage risks that could emerge in different locations and focus on building a centralized compliance program"¹¹⁰

In the earlier phases of establishing its centralized program, AB In Bev issued a code of business conduct and an anti-corruption policy, and created a "compliance channel" website to facilitate pre-approvals by compliance and to provide a conduit for employee questions and concerns. It also created a mobile application—nicknamed "Brewing Right"—for compliance pre-approvals and compliance questions about policies. In addition, to address antitrust compliance risk, AB InBev conducted a risk assessment exercise that resulted in a list of red

108 See AB InBev, 2020 ANNUAL REPORT 7 (February 25, 2021), https://www.ab-inbev.com/content/dam/abinbev/news-media/press-releases/2021/02/AB%20InBev%202020%20Annual%20Report_FINALpdf.pdf; Cindy Moehring, *Season 3, Episode 7: Matt Galvin | How Technology Can Be Used in Business To Catch Potential Crimes Before They Happen*, Business Integrity Leadership Initiative, Sam M. Walton College of Business, University of Arkansas (March 4, 2021), <https://walton.uark.edu/business-integrity/blog/matt-galvin.php>.

109 *Id.* at 14.

110 Eugene Soltes, *Designing a Compliance Program at AB InBev*, Harvard Business School Paper No. 9-118-071 at 2 (revised April 30, 2018).

flags that it could independently evaluate, and that were transformed into Key Performance Indicators (KPIs) for an antitrust dashboard.¹¹¹

AB InBev also created a compliance dashboard to evaluate how compliance teams were performing on each of their initiatives. That dashboard had multiple areas of focus including training, touchpoint vendors, investigations, and license committees. Within each area, policies and controls linked to specific KPIs. For example, for investigations, compliance personnel had to ensure that cases were addressed within a specified time period, that closed cases have been signed off on by appropriate personnel, and that proper remediation plans have been implemented.¹¹²

With the SAB Miller acquisition, AB InBev chose to bypass the traditional process of post-transaction integration, which is heavily dependent on human beings reviewing masses of data to gauge the acquired firm's anti-corruption compliance program. Instead, it resolved to create a single enterprise-wide repository that would draw on data from across multiple systems, with the object of creating a solution "that was useful inside and outside the compliance department."¹¹³

The analytics platform that AB InBev envisioned "could input the immense amount of data available to identify risks . . . [and] would not only be useful for the compliance integration, but could help more effectively manage compliance risk going forward." That platform, eventually codenamed "Operation BrewRIGHT," was established to integrate data from finance, compliance, human resources, and other systems to improve the identification of transactions and third parties that posed a risk. The foundation for that platform was a risk-scoring approach, in which certain transactions or relationships could be deemed higher-risk based on the number of risk attributes (e.g., urgency of payment, payment to a political or state-owned entity, and high-risk vendor type) and greater weighting of higher-risk attributes (e.g., a politically connected entity).¹¹⁴

¹¹¹ *Id.* at 7-8.

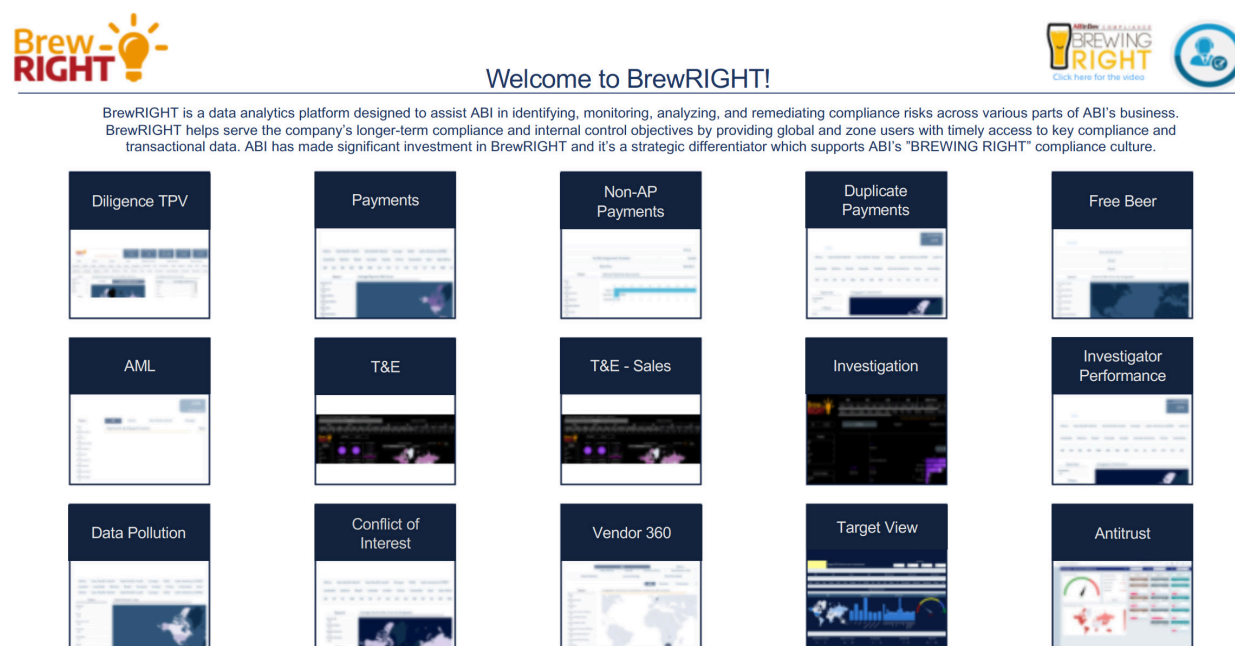
¹¹² *Id.* at 8.

¹¹³ Dylan Tokar, *Anheuser-Busch InBev's BrewRight: How It Works*, Risk & Compliance Journal, Wall Street Journal, January 17, 2020, https://www.wsj.com/articles/anheuser-busch-inbevs-brewright-how-it-works-11579257000?mod=article_inline.

¹¹⁴ Soltes, *supra* note 110, at 11.

AB InBev initially built its platform with assistance from external partners, but transitioned to an internally managed solution after one year. It ultimately expanded its tool to reach enterprise-wide,¹¹⁵ and can now access data from operations in dozens of countries.¹¹⁶ AB InBev's approach to anti-corruption now combines machine learning technology—including supervised learning¹¹⁷—with its BrewRIGHT analytics platform. BrewRIGHT provides users with dashboards that are built on third-party data-visualization software. A user who clicks on one of the icons on the dashboard is presented with a screen summarizing data about various processes within the company, such as third-party vendor due diligence and travel and entertainment, using lists, graphics, and maps.¹¹⁸

Figure 2—BrewRIGHT Platform Screenshot¹¹⁹



115 See Joe Williams, *supra* note 104.

116 See Dylan Tokar, *AB InBev Taps Machine Learning to Root Out Corruption*, Risk & Compliance Journal, Wall Street Journal, January 17, 2020, <https://www.wsj.com/articles/ab-inbev-taps-machine-learning-to-root-out-corruption-11579257001>.

117 See Matthew Galvin, *Overview of Data Analytics Network: AB InBev White Paper for Potential Tech Partner 2* (March 26, 2020).

118 See Dylan Tokar, *supra* note 113.

119 *Id.*

BrewRIGHT obtains those data “from more than a dozen enterprise resource planning systems—mainly related to accounts payable—as well as expense, compliance and investigations records, and external data from sources such as sanctions lists and Transparency International’s Corruption Perceptions Index.”¹²⁰

According to Matt Galvin, AB InBev’s global vice-president for ethics and compliance, each dashboard represents a workflow designed to solve a particular compliance problem, but many of the dashboards build upon each other and incorporate inputs from other dashboards. For example, a workflow might begin with a dashboard that deploys an algorithm to detect which suppliers are interfacing with the government. That workflow will end with a compliance review of the vendor and a profile and a risk score assigned to the vendor. That profile can be used to monitor whether certain economic activity is consistent with the purpose of the vendor as identified in the diligence process (i.e., was the vendor approved to perform activity A but ultimately paid to do activity B?). The risk score is then used to weigh and influence risk scoring of each transaction with that vendor.

As Figure 2 shows, BrewRIGHT is not limited to the broad topic of anti-corruption compliance, but includes anti-money laundering, antitrust, conflict of interest, payments, third-party vendors, travel and entertainment, and other compliance concerns. According to Galvin, these dashboards combine to provide aggregate risk scores for a wide range of risks across all AB InBev operating entities. This allows the AB InBev compliance team to direct resources in a risk-based way that is supported by its data, workflows, and ongoing assessments. Most importantly, all of these workflows produce structured data sets that are used to refresh and inform models through machine learning.

AB InBev’s solution reportedly has reduced the company’s costs associated with investigating suspect payments by millions of dollars. Galvin has stated that before BrewRIGHT, one AB InBev investigation into a certain type of third-party vendor in three countries cost AB InBev about \$1.8 million. In contrast, another investigation into the same type of vendor in six

¹²⁰ *Id.*

countries, using BrewRIGHT, cost about \$250,000.¹²¹ The continuous access to systems data alone, according to Galvin, helps AB InBev review and effectively manage compliance activities from the BrewRIGHT platform. It also has led the AB InBev compliance team to collaborate with other functions to expedite the identification and recovery of funds that would have otherwise been lost to fraud, graft, or lack of oversight.

Furthermore, AB InBev reports that after learning from experience with more than 9,000 vendors, its machine learning model has considerably improved. In some countries, the model can predict with a confidence level of approximately 80 percent whether a vendor has a proximity to a government official.¹²²

AB InBev is now advocating the idea of creating a data consortium “that enables companies to share insights from their data to boost the accuracy of their analytics models.”¹²³ At the World Economic Forum’s 2020 annual meeting, AB InBev’s Chief Executive Officer “called on other CEOs and world leaders to join the first anti-corruption data analytics consortium, based on the BrewRIGHT platform”, that aims to assist participating companies “better detect corruption and protect against it without revealing underlying company data.”¹²⁴

Galvin and others have described the proposal as “a framework for organizations to collaborate in vendor risk assessment, resulting in the cross-sharing of vendor risk scores.” The operation of the proposed framework would involve two phases. First, within each organization participating in the collaboration, there would be vendor risk assessment models (“organization-based models”) that use transactional data and other exogenous data to evaluate supplier risks.

These organization-based models would analyze the transactional data of the organization. Since organizations may have subsidiaries that run different systems, these organizational based models would perform not only a data integration role

¹²¹ Dylan Tokar, *supra* note 116.

¹²² See Matthew Galvin, *supra* note 117, at 2.

¹²³ *Id.*

¹²⁴ See AB InBev, *How BrewRIGHT is rooting out corruption at AB InBev and beyond*, January 31, 2020, <https://www.ab-inbev.com/news-media/innovation/how-brewright-is-rooting-out-corruption-at-ab-inbev-and-beyond.html>.

but also an aggregating function. The models would then perform risk scoring on both the transactional and aggregate data to generate risk scores for vendors.¹²⁵

Second, the organization-based models would continuously evaluate and score the risk of transactions within the organization, and calculate risk scores of supplier profiles based on specific predictive attributes of the suppliers. To validate the model's output, there would be a need for human input (e.g., business process owner/participant) to review the risk ratings generated by the models applying a consistent set of criteria and producing outputs following a specific manner. A human review would provide the necessary feedback for model updating.¹²⁶

Galvin and others have stated that those vendor risk scores would not only be evaluated within the organization, but also be pushed on a near real-time basis to a framework that would integrate the vendor risk assessment models that are running on the participating organizations' infrastructure. The framework would send the vendor risk scores "to a central core-algorithm that would aggregate the risk scores for all vendors and distribute these scores back to the collaborating organizations."¹²⁷

To address concerns about sharing of data that could contravene national data-privacy requirements, Galvin and others explain that the framework would use an algorithm that would maintain the privacy of transactional data, while enabling collaborating organizations to cross-share insights and experiences with vendors. At first, these various organization-based models would be static across the collaborating organizations, but later evolve as more industry-specific insight is gained.¹²⁸

This proposal would require participants to anticipate a variety of challenges, including maintaining the security and privacy of the data being analyzed and exchanged, determining

125 Matthew Galvin, Ivy Munoko, and Miklos Vasarelhi, *A Collaboration Framework for Democratizing Compliance Analytics*, 14 INTERNATIONAL REVIEW OF COMPLIANCE AND BUSINESS ETHICS 9 (October 2020).

126 *Id.* 8 and 10.

127 *Id.*

128 *Id.* at 9. As AB InBev explained elsewhere, the underlying concept would be to move AB InBev's algorithms onto a distributed platform that would allow other companies to borrow AB InBev's algorithms, apply them on their own datasets, and then give feedback to the consortium's central structure so that the algorithms learn from a collective of datasets. See Dan Clark, *How Anheuser-Busch Compliance Head Uses Analytics to Handle COVID-19 Challenges*, Corporate Counsel, July 1, 2020, <https://www.law.com/corpocounsel/2020/07/01/how-anheuser-busch-compliance-head-uses-analytics-to-handle-covid-19-challenges/>.

the level of transparency required for participants, establishing data collection and use limitations, treating vendors fairly in the risk-scoring process, establishing criteria for organizations to join the framework, and deciding how the framework would be funded and sustained.¹²⁹ The proposed framework nonetheless has the potential to allow the sharing of insights sufficient to improve models without the transfer or sharing of underlying commercial data. In theory, this could not only lead to substantial improvement of algorithms by allowing them access to greater scale, but even affect the underlying economic ecosystems within participating companies. A collective process for reviewing descriptive and behavioral characteristics of supply chains could allow members of a consortium to navigate their respective risks in a collective manner, without sharing personally identifiable characteristics of their supply chains.

Although this proposal would require companies to resolve a variety of legal and operational issues, including those mentioned above, before participating in such a consortium, it could also provide a means of directly addressing the class imbalance problem in anti-corruption machine learning, by providing larger datasets drawn from multiple consortium participants, which could lead to greater accuracy, precision, and recall in anti-corruption solutions.

2. Microsoft

Microsoft is a multinational company that sells technology services and products, with more than 168,000 employees worldwide. Headquartered in Redmond, Washington, Microsoft has subsidiaries in 120 countries and total revenues of more than \$43 billion for the last three months of 2020.¹³⁰

According to Microsoft, when Satya Nadella became Microsoft's Chief Executive Officer in 2014, he launched a series of internal and external initiatives to increase data-driven decision making and build a "data culture." Microsoft's Compliance and Ethics (C&E) team embraced this ethos and set out to build a data- and tech-driven early warning and monitoring

¹²⁹ Matthew Galvin, Ivy Munoko, and Miklos Vasarelhi, *supra* note 125, at 10-11.

¹³⁰ See Microsoft, *Facts About Microsoft*, <https://news.microsoft.com/facts-about-microsoft/#About>; Microsoft, *Earnings Release FY21 Q2*, <https://www.microsoft.com/en-us/Investor/earnings/FY-2021-Q2/income-statements>.

system for compliance risks. One of the C&E team's first steps was to write a compliance strategy memorandum that stated that the team wanted to "use big data to identify and predict compliance trends, and enhance monitoring and oversight—to identify a big data platform and define predictive models to identify compliance red flags and emerging areas of compliance risk." That statement was included in the memorandum that went to the company's senior leaders and the Audit Committee, and received a positive response.¹³¹

The Compliance Risk Management Solution that Microsoft sought to develop is directed at identifying and managing "high-risk" contracts and third parties. In developing that solution, however, Microsoft faced a number of challenges. One of the principal challenges involved the selection of data and models for its solution. As noted earlier, one challenge that a company may face in developing an anti-corruption solution is class imbalance.¹³² Microsoft was aware that machine learning

works better when it has a large number of examples from which it can identify patterns in the data and make better decisions in the future based on these examples. This means that it requires "labeled" training data to allow the system to learn automatically without human intervention or assistance. With only a relatively small number of contracts and third parties that have been historically reviewed out of millions of contracts and hundreds of thousands of third parties—meaning they have "unknown" risk profiles and no associated risk category training label—Microsoft was challenged to acquire sufficient labeled training data to create traditional ML classification models.¹³³

Moreover, only a small fraction of the relatively small number of contracts and third parties that Microsoft had reviewed received the highest risk label. Because only a small number of "high-risk" contracts and third parties were escalated to functions such as Legal, Compliance & Ethics, Finance, and Internal Audit, the vast majority of contracts and third parties did not receive a heightened level of review. Nor did Microsoft have a specific list of "good" or "low-risk" contracts and third parties identified within the unreviewed contracts and third parties.

131 See Valerie Charles, *Microsoft's Alan Gibson on the Power of Compliance Data*, GAN Integrity, May 11, 2020, <https://www.ganintegrity.com/blog/microsofts-alan-gibson-on-the-power-of-compliance-data/>.

132 See Section III E *supra*.

133 E-mail from Alan Gibson, Associate General Counsel, Microsoft, to Shruti Shah, President and CEO, Coalition for Integrity, and Jonathan J. Rusch, Adjunct Professor, Georgetown University Law Center (November 13, 2020) (hereinafter Gibson E-Mail).

Due to the smaller number of reviewed cases and even smaller number of associated risk category training labels, the Microsoft team “could not rely only on labels for training.”¹³⁴

Accordingly, instead of using a traditional labeling model, Microsoft used this partial information about known “high-risk” contracts and third parties to determine optimal weights for assembling individual risk model performance. For this approach, the foundation of the risk scores is derived from previously found compliance issues analyzed against three major risk areas: (1) contract and business relationship (i.e., sell-to, sell-with, buy-from, and “performance” data); (2) business environment (i.e., geography and local “performance” data); and (3) due diligence and vetting (i.e., third-party reports supplemented by continuous open, deep, and dark web searches).

Because risk category labels are available for only a small fraction of contracts and third parties, the Microsoft team recognized that restricting the dataset to include only investigated contracts and third parties would make the dataset too small for machine learning to produce reliable results. That view led to a conclusion that standard supervised learning methods that strictly classify a given contract or third party as “risky” or “not risky” cannot be used. Microsoft, however,

could use these known risk labels for the subset to assess the performance of individual component (risk attribute-level) models. If the individual component anomaly detection models are sensitive to contract or third-party level risk, Microsoft should observe a correlation between risk category and anomaly score. The High Risk Solution looks for the outliers and anomalies that have a heightened risk profile. Microsoft’s results bore this out—that, on average, the anomaly-based risk scores are useful for identifying the relative risk of contracts and third parties.¹³⁵

A second problem that Microsoft faced with its models involved the reality that compliance professionals, who are not data scientists, must make the final determination of whether controls are in place to effectively mitigate the identified risk. This means that human reviewers must be able fully to interpret the models and risk scores:

¹³⁴ *Id.*

¹³⁵ *Id.*

The goal[s] for the data models are to predict and identify the riskiest contracts and third parties for reviewers, and to reveal to the reviewers those risk attributes that require follow-up research, diligence, and mitigation. Because final assessments are ultimately carried out by human reviewers and because the goal of the program is to maximally facilitate these efforts, the models must contain strict interpretability. That is, no “black box” algorithms can be used to derive risk profiles, and the model output must be fully transparent and interpretable by human reviewers, who are subject matter experts in compliance, but non-machine-experts, via a series of dashboards and online tools (e.g., graphs, charts, workflows, and case management). Furthermore, a key component necessary for the success of the High Risk Solution was the buy-in from the compliance reviewer community; without trust and endorsement from the reviewers who are the intended users of the solution, some of whom are skeptical of “black box” machine learning scores, insights the High Risk Solution provides might go unactioned.¹³⁶

Ultimately, Microsoft’s approach to that solution is focused on deep learning based on neural networks.¹³⁷ Building compliance risk analytics solutions required close collaboration between the C&E team, with its knowledge of compliance risks, and the finance department, with its expertise in Microsoft’s IT systems, software engineering, data, and analytics. In addition, the C&E team gained sponsorship and support from the business groups impacted by the insights created by the solutions.

Microsoft also worked with several external partners to modernize its compliance program including a leading external consulting firm. A key consideration for Microsoft was to ensure that it could take a data-driven approach to manage “an effective compliance program and the risks associated with relying on a channel partner business model.”¹³⁸ Recognizing those risks, which could include both regulatory and reputation risks, it worked with the external firm, which incorporated knowledge about channel partner corruption and compliance risks into its analytics-based platform.¹³⁹

Microsoft’s Compliance Analytics Program Solution, as shown below in Figure 3, consists of three principal functions: (1) aggregating data from across the company; (2) using advanced

¹³⁶ *Id.*

¹³⁷ Alan Gibson, Microsoft, Compliance Analytics Program at 5-6 (May 11, 2020).

¹³⁸ *Id.*

¹³⁹ See PwC, *Transforming compliance into an asset*, <https://www.pwc.com/us/en/library/case-studies/microsoft-fighting-corruption-using-risk-command-platform-digital-technologies.html>.

statistics, machine learning, and AI to identify and predict transactions and third parties that create increased risk of noncompliance; and (3) operationalizing their review through dashboards and online tools to reduce compliance risks.

Figure 3—Microsoft’s Compliance Risk Management Solution

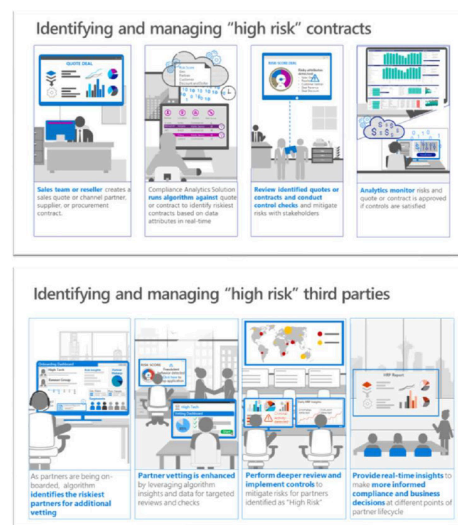
Microsoft’s Compliance Risk Management Solution

Data → Insight → Action

Aggregate data from across the company.

Use advanced statistics, machine learning and AI to **identify and predict** transactions and third-parties which create increased risk of non-compliance.

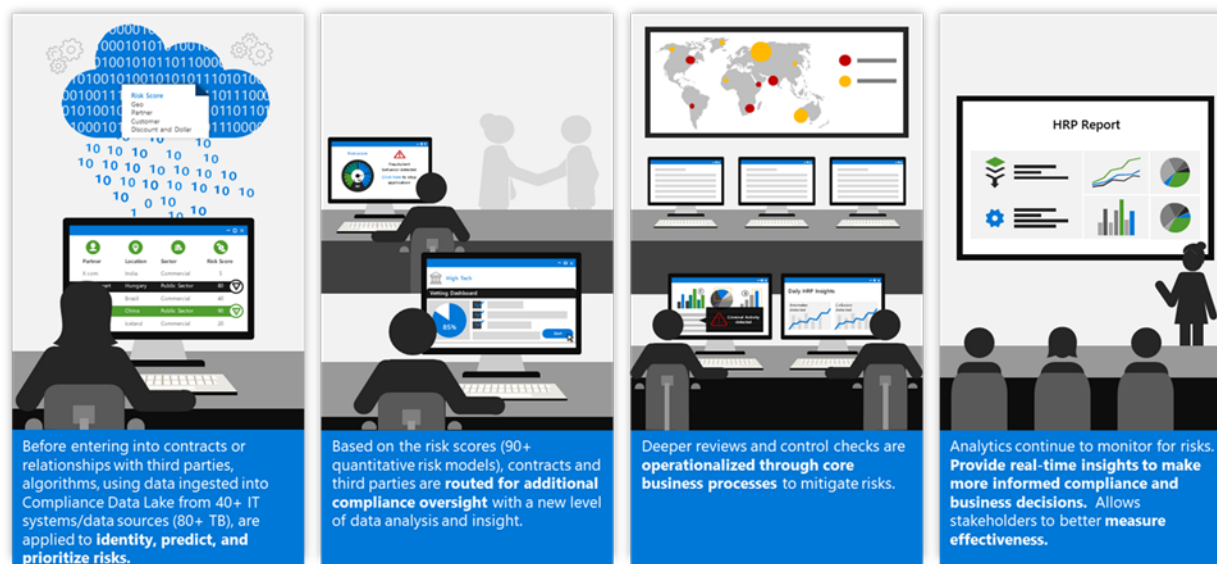
Operationalize their review through dashboards and online tools to **reduce compliance risks**. Solution identifies what is risky, why it’s risky, and what to do to mitigate.



The program’s approach involves “data ingestion” from more than 40 IT systems and/or data sources and more than 80 terabytes of data (both structured and unstructured), the use of more than 90 quantitative risk models, and more than 10 “risk scenarios” that provide a basis for action.¹⁴⁰ In essence, Microsoft proactively risk-scores its sales contracts and channel partners on a scale from zero to 100, and once a contract or partner exceeds a certain risk threshold or score, to assign a report to a compliance function embedded in the business for execution of controls.¹⁴¹

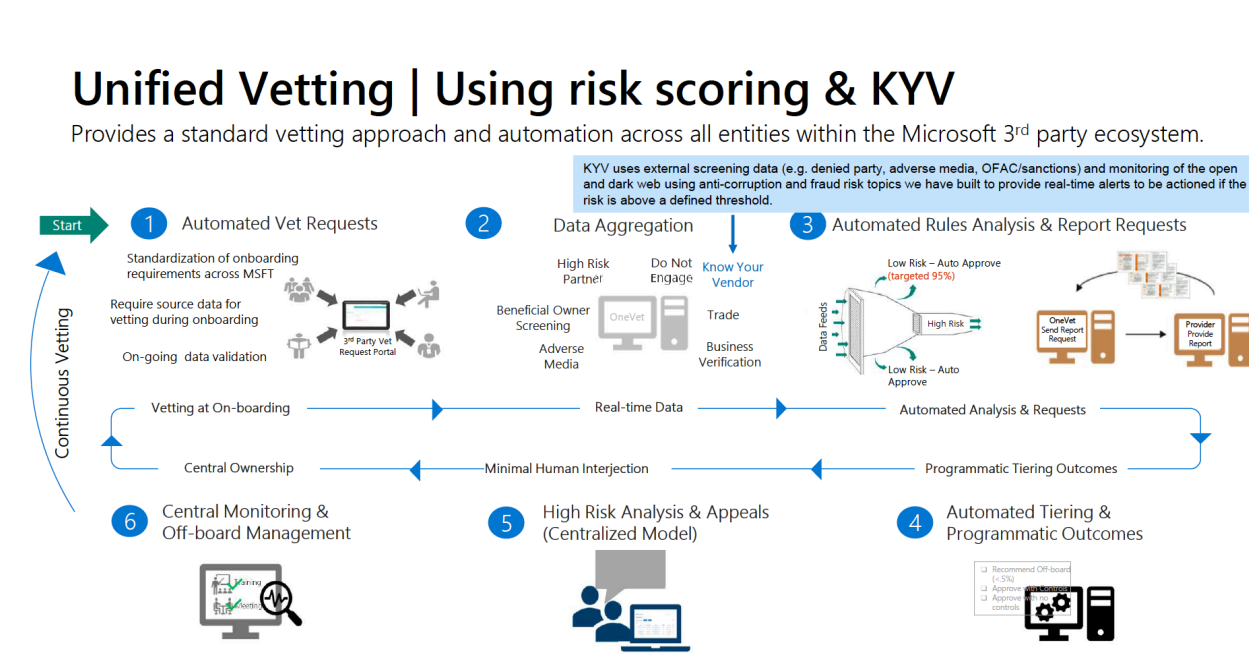
¹⁴⁰ See Alan Gibson, *supra* note 137, at 8-9.

¹⁴¹ Valerie Charles, *supra* note 131.

Figure 4—Overview of Microsoft Compliance Analytics Program

The Microsoft program provides a unified and continuing vetting process that consists of six steps: (1) an automated vetting process; (2) data aggregation; (3) automated rules analysis and report requests; (4) automated tiering and programmatic outcomes; (5) high risk analysis and appeals; and (6) central monitoring and off-board management.

Figure 5—Microsoft Compliance Program Vetting Approach¹⁴²



3. Alexion Pharmaceuticals

Alexion Pharmaceuticals is a global biopharmaceutical company that focuses on rare and devastating diseases such as myasthenia gravis. Headquartered in Boston, it had more than 3,000 employees at the start of 2020 and has a customer base primarily consisting of distributors, pharmacies, hospitals, hospital buying groups, and other healthcare providers.¹⁴³

Alexion began its pursuit of an anti-corruption machine learning solution while under investigation by the Securities and Exchange Commission (SEC), for foreign-official bribery by two of its subsidiaries seeking favorable regulatory treatment for its primary drug Soliris and approvals of Soliris prescriptions for individual patients.¹⁴⁴ The SEC alleged that Alexion,

¹⁴² Alan Gibson, *supra* note 137, at 13.

¹⁴³ See Alexion Pharmaceuticals, 2019 Annual Report 10 and 29 (February 4, 2020), <https://ir.alexion.com/static-files/8394a14e-2ae1-4b3d-aa27-96c5f4b2a4dc>.

¹⁴⁴ Dylan Tokar, *Corporate Compliance Programs Hit Refresh With Data-Analytics Tools*, Wall Street Journal, September 22, 2020, <https://www.wsj.com/articles/corporate-compliance-programs-hit-refresh-with-data-analytics-tools-11600767001>.

through subsidiaries, made payments to foreign officials in one country in order to influence them to provide favorable regulatory treatment, and to foreign officials in another country to influence the allocation of regional healthcare budgets for Soliris, increase the number of approved Soliris prescriptions, and favorably influence the regulatory treatment of Soliris. The SEC also stated that these payments were made in a variety of ways, including through the use of a third-party consultant, honoraria, and grants.¹⁴⁵

Because it serves patients in more than 50 countries, Alexion sought a solution to manage its relationship with doctors. Alexion worked with an external software provider to implement its machine learning solution, which addressed its need to manage its third-party relationships more effectively with streamlined approval and workflow.¹⁴⁶ That solution includes a cloud-based repository for some of Alexion's most high-value compliance data. According to Piyush Sharma, a deputy chief compliance officer at Alexion, it allows compliance team members to review data proactively in real time.¹⁴⁷ Through its solution, Alexion has a library of dozens of forensic risk analyses for different types of spend items, which leads to the calculation of an aggregate risk score for particular items.¹⁴⁸

Alexion's anti-corruption solution ultimately contributed to its reaching an agreement with the SEC to resolve charges that Alexion had violated the Foreign Corrupt Practices Act. In its cease-and-desist order to Alexion, the SEC noted that Alexion's remediation included "enhancing its policies and procedures regarding payments to third parties, including the implementation of a centralized system to track and monitor third-party payments."¹⁴⁹

Alexion's current solution incorporates supervised learning and transaction-level continuous monitoring that takes into account aggregate spend data from multiple sources (e.g., gifts, meals, and invoices) that are accumulated through the payment lifecycle, as well as

145 See Securities and Exchange Commission, Order Instituting Cease and Desist Proceedings Pursuant to Section 21C of the Securities Exchange Act of 1934, Making Findings, and Imposing a Case-and-Desist Order, In the Matter of Alexion Pharmaceuticals, No. 3-19852 (July 2, 2020), <https://www.sec.gov/litigation/admin/2020/34-89214.pdf>. Alexion agreed to this order without admitting or denying the SEC's findings.

146 See Ethisphere, *Using Data Analytics to Create a Next Generation Compliance Program: A Practical Roadmap*, YouTube, October 23, 2020, https://www.youtube.com/watch?v=SHCpifxYkNg&feature=emb_logo.

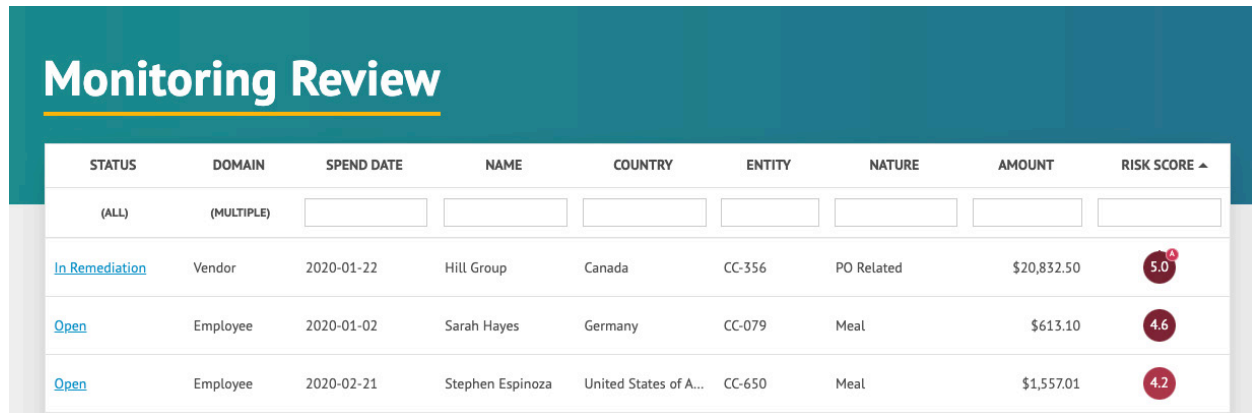
147 Dylan Tokar, *supra* note 144.

148 See Ethisphere, *supra* note 146 (remarks of Indrani Franchini, Executive Vice President and Chief Compliance Officer, Alexion).

149 Securities and Exchange Commission, *supra* note 145, at 7 ¶35, <https://www.sec.gov/litigation/admin/2020/34-89214.pdf>.

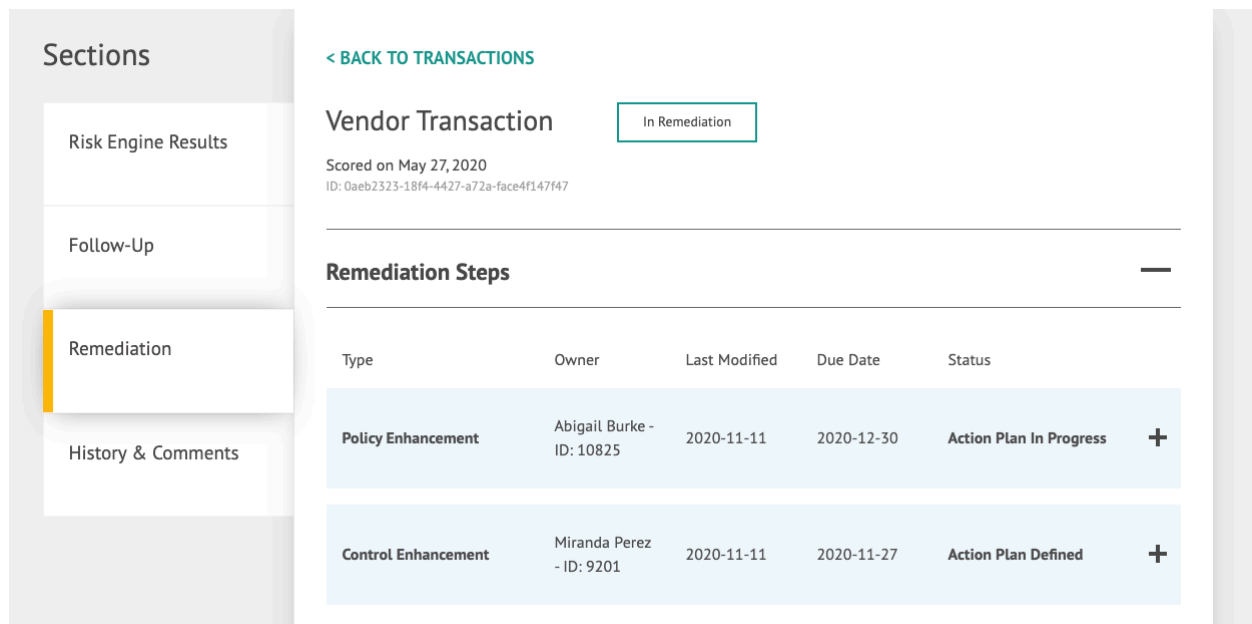
transactional risk scoring and the application of data analytics to those data.¹⁵⁰ The following two figures show examples of the type of data visualization that Alexion’s external software provider provides:

Figure 6—Transaction Review Screenshot



STATUS	DOMAIN	SPEND DATE	NAME	COUNTRY	ENTITY	NATURE	AMOUNT	RISK SCORE ▲
(ALL)	(MULTIPLE)							
In Remediation	Vendor	2020-01-22	Hill Group	Canada	CC-356	PO Related	\$20,832.50	5.0
Open	Employee	2020-01-02	Sarah Hayes	Germany	CC-079	Meal	\$613.10	4.6
Open	Employee	2020-02-21	Stephen Espinoza	United States of A...	CC-650	Meal	\$1,557.01	4.2

Figure 7—Transactional Risk Scoring Workflow Screenshot



Sections

- Risk Engine Results
- Follow-Up
- Remediation
- History & Comments

[< BACK TO TRANSACTIONS](#)

Vendor Transaction In Remediation

Scored on May 27, 2020
ID: 0aeb2323-18f4-4427-a72a-face4f147f47

Remediation Steps

Type	Owner	Last Modified	Due Date	Status
Policy Enhancement	Abigail Burke - ID: 10825	2020-11-11	2020-12-30	Action Plan In Progress +
Control Enhancement	Miranda Perez - ID: 9201	2020-11-11	2020-11-27	Action Plan Defined +

¹⁵⁰ Ethisphere, *supra* note 146.

Alexion's external software provider currently offers a product that integrates machine learning to improve its risk engine scoring over time. With regard to that product, the provider has stated that "as the system is trained by users resolving flagged expenses and disbursements, future transactions can be prioritized higher (i.e., receive higher scores) where past analyses or combinations of analysis have yielded confirmed results."¹⁵¹ The provider has also convened a network of customers and partner organizations to gather collective insights to improve the overall risk analyses in the platform.¹⁵²

151 See Lextegrity, *Integrity Gateway Monitoring*, www.lextegrity.com/monitoring.

152 See Lextegrity, *Integrity Analytics Collective*, www.lextegrity.com/collective.

IV. Ethical, Legal, and Governance Issues in Implementation and Operation of Anti-Corruption Machine Learning: Analysis and Recommendations

A. Ethical Issues

Anti-corruption machine learning, like other areas of computing and IT, raises a wide range of challenges that include ethical issues, data privacy, and data security.¹⁵³ While some ethical issues have legal ramifications (e.g., unethical conduct may cause harm to others that public authorities can sanction through the courts or administrative processes), this section will focus on broader ethical considerations in building and using anti-corruption ANI.

Codes of ethics are increasingly evident in larger companies around the world, and a number of companies are expanding the scope of those codes to address the development and use of various forms of AI. They typically discuss a broad range of issues associated with AI solutions in general, such as the need for fairness, transparency of solutions, security and transparency of use of customer data, and accountability.¹⁵⁴

Such codes of ethics are critical to the proper development and use of ANI. As a leading global technology company put it, “[e]thics must be embedded in the design and development process from the very beginning of AI creation.”¹⁵⁵

Moreover, there is already momentum in some quarters for government regulation of ANI that would address ethical concerns. In October 2020, the European Parliament adopted a legislative initiative urging the European Commission “to present a new legal framework

153 See, e.g., Dell Technologies, *Leveraging AI for Good—A Global Opportunity for Policy-Makers*, September 2, 2019, <https://www.delltechnologies.com/en-us/perspectives/leveraging-ai-for-good-a-global-opportunity-for-policy-makers/>.

154 See, e.g., AI Ethics Guidelines Global Inventory, Algorithm Watch, <https://inventory.algorithmwatch.org/>; Deutsche Telekom, *Digital Ethics Guidelines on AI* (July 19, 2018), available at <https://www.telekom.com/en/company/digital-responsibility/details/artificial-intelligence-ai-guideline-524366>; Microsoft, *Microsoft AI principles*, available at <https://www.microsoft.com/en-us/ai/responsible-ai>.

155 IBM, *Everyday Ethics for Artificial Intelligence* at 6 (2019 ed.), <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>.

outlining the ethical principles and legal obligations to be followed when developing, deploying and using artificial intelligence, robotics and related technologies in the EU including software, algorithms and data.”¹⁵⁶

Accordingly, any company that is considering using anti-corruption machine learning should review its codes of ethics to see whether and how those codes address the ethical dimensions of machine learning, with reference to the following categories of ethics-related issues.

1. Responsible Design and Use

The first category of ethical issues concerns the need for responsible design and use of AI systems, including machine learning.¹⁵⁷ This category consists of three subcategories of responsibility, all of which pertain to anti-corruption machine learning:

- ▶ **Safety and Security:** This responsibility concerns the importance of systems being designed, functioning, and being used “in a responsible, safe, and secure way throughout their life cycle.”¹⁵⁸
- ▶ **Ongoing Risk Management:** This responsibility concerns the importance of continually assessing, addressing, and managing potential risks that AI systems pose, commensurate with their expected impact.¹⁵⁹ While there is growing recognition of the utility of AI in enhancing an enterprise’s risk management program,¹⁶⁰ the enterprise also needs to plan for and execute a risk management program for any AI system that it deploys, including appropriate training for risk and compliance officers.

¹⁵⁶ European Parliament, Release: Parliament leads the way on first set of EU rules for Artificial Intelligence, October 20, 2020, <https://www.europarl.europa.eu/news/en/headlines/priorities/artificial-intelligence-in-the-eu/20201016IPR89544/parliament-leads-the-way-on-first-set-of-eu-rules-for-artificial-intelligence>.

¹⁵⁷ See Pamela Passman, *Artificial Intelligence: Evolving Risks and Responsibilities*, CIO Review, <https://storage.cioreview.com/cxinsight/artificial-intelligence-evolving-risks-and-responsibilities-nid-30215-cid-12.html>.

¹⁵⁸ *Id.*

¹⁵⁹ *Id.*

¹⁶⁰ See, e.g., *Artificial Intelligence and Risk Management*, ENTERPRISE RISK, <https://enterpriseriskmag.com/artificial-intelligence-risk-management/>; Deloitte, *Why artificial intelligence is a game changer for risk management*, <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/audit/us-ai-risk-powers-performance.pdf>.

- **Transparency:** This responsibility concerns the importance of transparency and responsible disclosure around AI systems, “to ensure that people understand AI-based outcomes and can challenge them.”¹⁶¹ Among other considerations, to maximize algorithmic transparency and accountability, companies should (1) be aware of the possible biases involved in ANI solutions’ design, implementation, and use; (2) seek to have their systems and institutions that use algorithmic decision-making produce explanations regarding both the procedures that the algorithm follows and the specific decisions that the solution makes; (3) have the builders of the algorithm maintain a description of the way in which the company collected the training data and an exploration of the solution’s potential biases; (4) record models, algorithms, data, and decisions for auditability purposes; and (5) “use rigorous methods to validate their models and document those methods and results.”¹⁶²

Because the spectrum of machine learning, from supervised learning to reinforcement learning, increasingly excludes human beings from direct roles in creating and modifying algorithms, a company that is interested in developing and deploying an anti-corruption machine learning solution must recognize that, as discussed earlier, it will need to use a variety of indirect approaches, including performance measures, to challenge that solution and see that it is generating an acceptably high level of accurate predictions.

2. Ethical Use

The second category of ethical issues addresses three responsibilities associated with the ethical use of AI, all of which pertain to anti-corruption machine learning:

- **Socially Useful.** In general, any type of ANI “should benefit people and the planet by driving inclusive growth, sustainable development, and well-being.”¹⁶³ As part of this responsibility, a company developing an anti-corruption machine learning

161 Pamela Passman, *supra* note 157.

162 ACM US Public Policy Council, Statement on Algorithmic Transparency and Accountability (January 12, 2017), https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf.

163 Pamela Passman, *supra* note 157.

solution should ensure that its planning and development is integrated into the company's more general commitments to supporting the United Nations' Sustainable Development Goals.¹⁶⁴

- ▶ **Human Values:** The use of any ANI “should not undermine but support human rights, dignity, liberties, fairness, and diversity, and avoid discrimination and bias.”¹⁶⁵
- ▶ **Human Control:** Human beings “should remain in control of choosing how and whether to delegate decisions to AI systems, and AI systems should be used to accomplish human-chosen objectives.”¹⁶⁶

As the Human Values and Human Control requirements indicate, the choices of how and whether to use an anti-corruption solution such as machine learning to generate predictions may have legal as well as ethical ramifications.

B. Legal Issues

This category addresses the need for anti-corruption machine learning to comply with all relevant and applicable laws. Under this category, there are three subcategories of legal requirements.¹⁶⁷

1. Data Privacy

This subcategory pertains to the need for the collection, processing, use, and transfer of personally identifiable data by AI systems should comply with relevant data privacy laws. The most widely applicable data privacy law that would be relevant to an anti-corruption machine

¹⁶⁴ See Department of Economic and Social Affairs, United Nations, Sustainable Development: The 17 Goals, <https://sdgs.un.org/goals>.

¹⁶⁵ Pamela Passman, *supra* note 157.

¹⁶⁶ *Id.*

¹⁶⁷ *Id.*

learning solution is the European Union General Data Protection Regulation (GDPR).¹⁶⁸ In general, section 5 of the GDPR sets out seven key principles with regard to personal data:

1. Data must be “processed lawfully, fairly and in a transparent manner in relation to individuals (‘lawfulness, fairness and transparency’)”;
2. Data must be “collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall not be considered to be incompatible with the initial purposes (‘purpose limitation’)”;
3. Data must be “adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (‘data ’)”;
4. Data must be “accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay (‘accuracy’)”;
5. Data must be “kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes subject to implementation of the appropriate technical and organisational measures required by the GDPR in order to safeguard the rights and freedoms of individuals (‘storage limitation’)”;
6. Data must be “processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing

168 European Parliament, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN>.

and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality'); and

7. The controller of the data controller "shall be responsible for, and be able to demonstrate compliance with, [each of the foregoing requirements] ('accountability')." ¹⁶⁹

Development and operation of any machine learning system necessarily involves drawing on large datasets whose data reside in a variety of locations, including (in many cases) data that are covered by the GDPR. For that reason, a company that is considering whether to develop or acquire a machine learning solution for anti-corruption purposes must carefully analyze whether and how it would need to comply with the seven GDPR principles listed above. That analysis may become more complex as the European Union is now considering draft guidelines that would require increased privacy safeguards for information transferred beyond the EU. ¹⁷⁰

In addition to these specifically stated principles, some have maintained that the GDPR confers a broad "right to explanation"—that is, that entities controlling personal data ensure "fair and transparent processing," which requires that those entities provide people with access to "meaningful information about the logic involved" in certain automated decision-making systems. ¹⁷¹

This "right" is controversial for two reasons. First, it is nowhere specifically stated in the text of the GDPR, which raises serious questions about whether it actually exists as an enforceable right and what the scope of its application might be. Second, it is not at all clear how such a "right" could be construed with respect to machine learning. As explained earlier, machine learning, unlike rule-based programming, does not involve human beings writing the code for specific algorithms. Even in supervised learning, human beings provide the labels for data to be analyzed, but do not write the algorithms themselves. And in other

169 Information Commissioner's Office, *Guide to the General Data Protection Regulation (GDPR)/The principle*, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/#the_principles.

170 See Catherine Stupp, *EU Restrictions Could Force Companies to Change Data Transfer Practices*, Wall Street Journal, November 17, 2020, <https://www.wsj.com/articles/eu-restrictions-could-force-companies-to-change-data-transfer-practices-11605609001>.

171 Bryan Casey, Ashkan Farhangi, and Roland Vogl, *Rethinking Explainable Machines: The GDPR's "Right to Explanation" and the Rise of Algorithmic Audits in Enterprise*, 34 Berkeley Technology Law Journal 145, 151 (2019).

forms of machine learning, such as semi-supervised, unsupervised, and reinforcement learning, human beings neither write the algorithms nor provide the labels for a machine learning application.

Thus, under all four categories of machine learning, there is no way in which programmers could directly show an external observer how and why a particular machine learning solution—especially one using a neural network with multiple layers of neurons—made the decisions it did. At best, programmers would not be able to articulate “the logic involved” in a particular machine learning solution if they have no way directly to observe the logic that the solution devised for itself. They could use indirect measures, however, to show that the results that the solution generated were free from impermissible biases and provided acceptable levels of accuracy, precision, and recall. This difficulty in showing how a machine learning solution made decisions affecting individuals covered by the GDPR may pose serious challenges to companies as they seek to accommodate individual rights under the GDPR, including access, deletion, correction, restriction, and not to be subject to automated decision making or profiling.

In addition, a number of individual nations have adopted, or are considering, legislation similar to the GDPR or consistent with the GDPR’s principles. These include Brazil, in which its *Lei Geral de Proteção de Dados Pessoais* (LGPD, or General Personal Data Protection Act)¹⁷² has recently come into force, as well as countries on five other continents.¹⁷³

The United States is also seeing growing support for data privacy legislation at the federal and state levels. While there is no general federal data privacy legislation yet, California has enacted the California Consumer Privacy Act (CCPA), which came into force in 2020,¹⁷⁴ and

172 Presidência da República, Lei Geral de Proteção de Dados Pessoais, Lei Nº 13.709, de 14 de Agosto de 2018, http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm.

173 See International Association of Privacy Professionals, The General Data Protection Regulation Matchup Series, <https://iapp.org/resources/article/the-general-data-protection-regulation-matchup-series/>.

174 Gary Grossman, *Why Businesses Should Adopt an AI Code of Ethics—Now*, Information Week, November 14, 2019, <https://www.ncsl.org/research/telecommunications-and-information-technology/2020-consumer-data-privacy-legislation637290470.aspx>.

Virginia recently enacted the Consumer Data Protection Act in 2021. In addition, at least 30 states and Puerto Rico considered data privacy legislation in 2020.¹⁷⁵

For those reasons, a company considering whether to adopt an anti-corruption machine-learning solution should be prepared to explain, for internal and external audiences, what kinds of indirect measures it could use to allow inferences about the proposed solution's logic and why those measures would be appropriate to warrant such inferences.

2. Cybersecurity

This requirement pertains to the need to ensure that the use of anti-corruption machine learning “should include effective cyber and physical security to mitigate the risk of data theft and to promote trust.”¹⁷⁶ Initially, a company planning to develop and deploy anti-corruption machine learning may not regard this as a significant concern as part of its planning process, because it already has a robust cybersecurity program. Companies, however, need to recognize that, as with any significant changes to its operational and information technology (IT) networks, developing an anti-corruption solution necessarily involves moving large amounts of data from across the enterprise and combining and storing those data in new ways and locations (e.g., the training, validation, and testing datasets for the solution). Both the company and any external consulting firm with which the company may partner to build the solution will need to factor those new data aggregations into their cybersecurity framework and closely coordinate on needed changes in those frameworks.

In addition, companies need to understand that machine learning can create new and different cybersecurity challenges that may not be obvious even to data scientists. One example is the phenomenon known as unintentional memorization. Researchers have found

175 SB1392ConsumerDataProtectionAct, LegislativeInformationSystem, <https://lis.virginia.gov/cgi-bin/legp604.exe?212+sum+SB1392>; National Conference of State Legislatures, 2020 Consumer Data Privacy Legislation (January 17, 2020), <https://www.ncsl.org/research/telecommunications-and-information-technology/2020-consumer-data-privacy-legislation637290470.aspx>. While it is beyond the scope of this guidance, the CCPA is expected to have significant effects on companies' use and retention of AI, including disclosures to consumers about its use of personal information in machine learning solutions. See Tom Taulli, *CCPA: What Does It Mean For AI (Artificial Intelligence)?*, Forbes, December 27, 2019, <https://www.forbes.com/sites/tomtaulli/2019/12/27/ccpa-what-does-it-mean-for-ai-artificial-intelligence/?sh=177694a46bb2>.

176 Pamela Passman, *supra* note 157.

that certain neural network models that classify or predict sequences of natural-language text have unintentionally “memorized” certain text sequences containing sensitive personal data in emails that were part of the training dataset. As a result, users of the model could enter a text prefix such as “My Social Security number is ...” and cause the model to auto-complete the text sequence to the Social Security number itself.¹⁷⁷ In one case, a leading technology company that had trained a neural network on a dataset of more than 100,000 emails containing sensitive employee information was

able to extract multiple Social Security numbers and credit card numbers from the neural net’s predictions. [The neural network] had memorized the information in such a way that it could be recovered by any user—even without access to the original dataset.¹⁷⁸

Unintentional memorization, though it can be persistent and hard to avoid,¹⁷⁹ is certainly a soluble problem. Options for researchers include selecting means of training models that minimize such behavior by the model,¹⁸⁰ and “keeping sensitive data out of a neural network’s training dataset in the first place.”¹⁸¹ The larger point is that companies need to anticipate that anti-corruption machine learning may raise cybersecurity issues that may not be adequately addressed in its cybersecurity program.

3. Use for Lawful Purposes

It should go without saying that machine learning “should not be used for dangerous purposes or activities that are otherwise illegal.”¹⁸² Nonetheless, a company considering the use of anti-corruption machine learning should bring together its ethics and compliance team, its legal department, and its data scientists to evaluate legal risks that have already been identified with regard to existing machine learning solutions and compliance programs, as well as potential legal risks that the proposed anti-corruption solution could trigger.

177 Nicholas Carlini, Chang Liu, Úlfar Erlingsson, Jernej Kos, and Dawn Song, The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks, <https://arxiv.org/abs/1802.08232> (revised July 16, 2019).

178 JANELLE SHANE, *supra* note 76, at 131.

179 Nicholas Carlini, Chang Liu, Úlfar Erlingsson, Jernej Kos, and Dawn Song, *supra* note 177, at 1.

180 *Id.*

181 JANELLE SHANE, *supra* note 76, at 131.

182 Pamela Passman, *supra* note 157.

Those legal risks can include machine learning solutions generating outputs that result in unjustifiably discriminating against individuals on the basis of categories such as gender, race, and ethnicity¹⁸³, as well as solutions that draw on data whose permissible uses may be limited under various jurisdictions' employment laws.

Some potential risks may not be obvious at the time of a company's consideration of an anti-corruption machine learning solution. For example, in 2020, a Dutch court ruled that the Risk Indication System (SyRI), a statutory data-processing tool that the Dutch Government used to combat fraud in areas such as benefits, surcharges, and taxes, did not comply with Article 8 of the European Convention on Human Rights (ECHR), which protects the right to respect for private life.¹⁸⁴ Noting that the tool had thus far been applied only to so-called "problem" (i.e., poorer) neighborhoods,¹⁸⁵ the court found that "there is a risk that the use of SyRI will inadvertently establish links based on bias, such as lower socioeconomic status or an immigration."¹⁸⁶ It was particularly critical of SyRI's lack of transparency, holding that under Article 8, the right to respect for private life "also means that a data subject must be reasonably enabled to follow his or her data."¹⁸⁷

Although this decision applies only within the Netherlands, its reasoning may influence courts in other European Union countries to adopt similar reasoning. In any event, it indicates that companies must be thorough in identifying and working to reduce potential legal risks that may arise in the design and operation of anti-corruption machine learning.¹⁸⁸

One other issue that a company implementing an anti-corruption learning solution should consider is the possibility, however remote, that if a company manager or executive involved in corrupt transactions knows about that solution, he may seek either to change his behavior

183 See Section III E *supra*.

184 See *Dutch Lawyers Committee for Human Rights v. State of the Netherlands*, No. C-09-550982-HA SAT 18-388 (Hague District Court, February 5, 2020), <https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:865>.

185 *Id.* at ¶6.93.

186 *Id.* at ¶6.93.

187 *Id.* at ¶6.90. The court also expressed concerns about whether SyRI violated various GDPR provisions, but did not conclude that it did so. See *id.* ¶6.107.

188 See Jeffrey P. Cunard, Luke Dembosky, Avi Gesser, Henry Lebowitz, Jim Pastore, Lisa Zornberg, Anna R. Gressel, and Steven Tegar, *Fifteen Ways to Reduce Regulatory and Reputational Risks for Your AI-Powered Applications—Lessons from Recent Court Decisions and Regulatory Activity*, Debevoise in Depth, February 20, 2020, available at <https://www.debevoise.com/insights/publications/2020/02/fifteen-ways-to-reduce-regulatory-and>.

to reduce the risk of being detected or to seek in some way to interfere with the operation of the solution. Because no machine learning model is static and will undergo retraining to address new data and changing external conditions,¹⁸⁹ it should prove difficult for an executive involved in corruption to conceal his conduct completely from an anti-corruption compliance team's scrutiny. As for the possibility of the executive overtly attempting to interfere with the solution's operations, the company should provide specific directions to both its compliance team members and data scientists working on the solution that they should immediately report any such attempt to a designated senior executive (e.g., the Chief Compliance Officer or General Counsel) and be prepared to escalate the issue immediately if circumstances warrant.

Finally, a company should note that a potential category of legal risk can arise if it fails to maintain and retrain its machine learning model as its business activities and bribery and corruption risks change, or if it assumes that deployment of an anti-corruption solution allows it to eliminate or reduce other anti-corruption measures. A company should always regard an anti-corruption machine learning solution as one component of its overall compliance program, but not as a substitute for that program. Regulators and prosecutors are likely to consider an anti-corruption program ineffective if the company's anti-corruption solution proves ineffective at identifying higher-risk transactions and relationships, or if the company fails to maintain sufficient staff with the necessary training and expertise to make effective use of the solution's output.

C. Governance Issues

A further step that a company considering an anti-corruption machine learning solution needs to take is to examine and revise as necessary its existing governance and compliance structures and processes with reference to that solution.¹⁹⁰ Among other considerations, to provide the maximum benefit to the company, the company should consider having the anti-

¹⁸⁹ See Section III F *supra*.

¹⁹⁰ See, e.g., Paul Huck, Aaron Johnson, Nicholas Kiritz, and C. Erik Larson, *Why AI Governance Matters*, RMA Journal, May 2020, at 18, https://www.promontory.com:3000/static/pdf/1588624225_title.pdf.

corruption machine learning solution reach across the enterprise, including all operating divisions and lines of business, to obtain a dataset of sufficient size to generate the most useful predictions. For companies that have expanded through multiple acquisitions and have multiple divisions with different structures and different enterprise risk management approaches and software, that may pose a considerable challenge.

Consequently, a company may have to make substantial governance and structural changes to make possible an enterprise-wide anti-corruption solution. In the case of AB InBev, its decision-making process was spurred by a \$100 billion-plus acquisition that required AB InBev to integrate the acquired firm's compliance program, which covered that firm's operations in 25 countries. AB InBev chose to bypass the traditional process of post-transaction integration, which is heavily dependent on human beings reviewing masses of data to gauge the acquired firm's anti-corruption compliance program. Instead, it resolved to create a single enterprise-wide repository that would draw on data from across multiple systems, with the object of creating a solution "that was useful inside and outside the compliance department."¹⁹¹

But a company also needs to think more broadly about how to integrate its anti-corruption machine learning solution into its larger governance framework. That can mean bringing that solution within the company's larger ANI governance plan, including ensuring that the solution's consistency with the plan's objectives (e.g., ethics, fairness, explainability, and transparency).¹⁹²

191 Dylan Tokar, *supra* note 116.

192 See Tom Taulli, *AI (Artificial Intelligence) Governance: How To Get It Right*, Forbes, October 10, 2020, <https://www.forbes.com/sites/tomtaulli/2020/10/10/ai-artificial-intelligence-governance-how-to-get-it-right/#7d44e512745f>.

Conclusions

This document provides companies in multiple sectors with guidance on whether and how they should pursue developing or acquiring anti-corruption ANI. To date, only a few companies have actually developed and deployed machine learning solutions pertaining to anti-corruption risk and compliance programs. The experiences of the three companies described above provide significant proof of concept, with two significant caveats.

First, the size, business models, and financial resources of two of these companies (AB InBev and Microsoft) likely made it more feasible for them to undertake the extended process of developing and implementing anti-corruption machine learning solutions tailored to their specific risk and compliance needs. Continuing development of anti-corruption machine learning solutions, however, may substantially reduce the time and cost for other companies to implement such solutions.

Second, the volume of these companies' data associated with key compliance areas, such as due diligence on third-party contracts and relationships, made anti-corruption machine learning more cost-effective for them, and made it more likely that machine learning algorithms could ultimately generate an acceptably high rate of predictions. Many small to moderate-size companies that may otherwise be willing to consider using anti-corruption machine learning may have challenges in identifying enough internal data to overcome the class imbalance problem. On this point, such companies could benefit from consulting with companies now offering anti-corruption machine learning solutions, after reviewing the guidance in this document, to see whether they have enough categories of relevant data to make development of an anti-corruption solution feasible.

For those reasons, it would be inappropriate to say that all companies, regardless of size, business model, and financial resources, need to adopt anti-corruption machine learning, or that regulators will expect all companies to incorporate anti-corruption machine learning into their compliance programs. But experience to date indicates that anti-corruption machine learning holds considerable promise, and that companies should take that into account in deciding how to improve their anti-corruption and related compliance programs.

Companies, then, should take this guidance not as a recommendation that they immediately pursue anti-corruption machine learning, but rather as a template to assist them in internal and external discussions about possible deployment of anti-corruption machine learning. As the examples in this guidance indicate, there can be substantial costs and time commitments associated with establishing and maintaining an effective anti-corruption machine learning system. But depending in part on the financial resources of the company and in part on the size of the dataset on which the company can draw for training, validation, and testing, the cost in the end may be significantly less than if the company were to use full-time employees to do similar data review and analysis.¹⁹³

Based on the examples in this document, companies should also expect that, depending on their business and compliance needs, it may make sense for them to adopt machine learning solutions that are not narrowly tailored just to anti-corruption compliance, but that can address a broader range of compliance needs. As data from multiple sources within a company may be useful in addressing multiple compliance risks, both for compliance oversight and for investigations, and integrating machine learning solutions that can draw on the full range of those data into a broader compliance platform, rather than building multiple standalone compliance solutions, may be more efficient and cost-effective over the long term.

In addition, companies should anticipate that, as they are considering adoption of anti-corruption machine learning solutions, they should look for opportunities to collaborate with other companies in making those solutions more effective. Such opportunities may include the anti-corruption data analytics consortium that AB InBev is promoting,¹⁹⁴ as well as an Integrity Analytics Collective on which eight firms have partnered to promote continuous improvement of data analytics based on best practices and to assist companies in responding more quickly to evolving risks and schemes.¹⁹⁵

193 See *Mini-Roundtable: Data Analytics and AI for Anti-Corruption Compliance*, Risk & Compliance Magazine, January–March 2019 (remarks of John H. Loesch, Navigant), <https://guidehouse.com/-/media/www/site/insights/gic/2019/corporate-disputes-magazine-data-analytics-and-ai.pdf>.

194 See Section III G *supra*.

195 See Lextegrity, *supra* note 152.

In short, anti-corruption machine learning may constitute a substantial improvement for companies over their current rule-based programming and data analytics. What matters most in the end is not whether a company can genuinely “predict” corruption before it begins¹⁹⁶, but whether it can satisfy itself that it has a genuine business case, on a risk-based basis, for adopting and implementing anti-corruption machine learning, based on its own risk profile and a frank evaluation of the benefits, costs, and risks of that machine learning.

196 See Cindy Moehring, *supra* note 108.

APPENDICES

Appendix 1: Glossary

Artificial Intelligence

Artificial intelligence can be divided into two main categories. **Artificial general intelligence (AGI)** can be defined as a machine that can understand the world as well as a human being, and with a human being's capacity to learn how to carry out a huge range of tasks. **Artificial narrow intelligence (ANI)** can be defined as a machine or system that can perform only one narrowly defined task or a small set of related tasks. Because there is no such thing as AGI that can function as a human being does, including exercising judgment and discretion, in carrying out anti-corruption duties, this guidance will confine its discussion to specific concepts and categories within ANI.

ANI is closely related to **data science**, a broader field that “encompasses a set of principles, problem definitions, algorithms, and processes for extracting non-obvious and useful patterns [i.e., actionable insight] from large data sets.”¹⁹⁷ An **algorithm**, in computer science, is simply a set of instructions that tell a computer how to perform a specific task.¹⁹⁸

A **dataset**, in the context of machine learning, “is a collection of data [that] corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question.”¹⁹⁹

¹⁹⁷ JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 1.

¹⁹⁸ See *Algorithm*, TechTerms (updated August 2, 2013), <https://techterms.com/definition/algorithm#:~:text=An%20algorithm%20is%20a%20set,playing%20a%20compressed%20video%20file.&text=In%20computer%20programming%2C%20algorithms%20are%20often%20created%20as%20functions>.

¹⁹⁹ [Alexandre Gonfalonieri](#), *supra* note 72.

Rule-Based Programming

Rule-based programming can be considered a form of ANI. In rule-based programming, human programmers write code that establishes rules defining all aspects of a specific task (e.g., “If A occurs then do X, if something other than A occurs, then do Y”) and install them in a computer system.²⁰⁰

Machine Learning

Machine learning is a subset of AI. It “focuses on developing and evaluating algorithms that can extract useful patterns from data sets. A machine-learning algorithm takes a data set as input and returns a model that encodes the patterns the algorithm extracted from the data.”²⁰¹ In functional terms, machine learning can be characterized as a process in which algorithms use statistics to find patterns in very large amounts of various types of data, such as numbers, words, and even batches of text.²⁰²

A machine learning **model** “is a representation of a pattern extracted using machine learning from a data set. Consequently, models are trained, fitted to a data set, or created by running a machine learning algorithm on a data set.”²⁰³

Deep Learning

Deep learning is a subfield of machine learning that depends on the concept of a **neural network**. A neural network is “[a] type of machine learning model that is implemented as a network of simple processing units called neurons. . . . A neuron takes a number of input values (or activations) as input and maps these values to a single output activation.”²⁰⁴ (An **activation function** is a mathematical equation that determines the output of a neural network. The function is attached to each neuron in the network, and determines whether

200 Elana Krasner, *supra* note 15.

201 JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 243 (emphasis removed).

202 See Karen Hao, *What is machine learning?*, MIT Technology Review, November 17, 2018, <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/>.

203 JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 243.

204 *Id.* at 244.

it should be activated (“fired”) or not, based on whether each neuron’s input is relevant for the model’s prediction.”²⁰⁵)

A deep-learning model is therefore

a neural network that has multiple (more than two) layers of [neurons]. Deep networks are deep in terms of the number of layers of neurons in the network. Today many deep networks have tens to hundreds of layers. The power of deep-learning models comes from the ability of the neurons in the later layers to learn useful attributes derived from attributes that were themselves learned by the neurons in the earlier layers.²⁰⁶

An **attribute** captures one piece of information relating to an instance (i.e., an example) in a data set.²⁰⁷ An **instance** or example can have multiple attributes (e.g., characteristics pertaining to transactions or relationships, as well as a label indicating the class to which the instance belongs).²⁰⁸

In the field of anti-corruption, researchers in one study used a neural-network approach to examine the votes of Brazilian legislators over a 28-year period and predict subsequent arrests and convictions of those legislators for corruption or other financial crimes.²⁰⁹

Supervised Learning

Supervised learning is one of four types of machine learning. In simple terms, supervised learning can be defined as a process in which human beings select certain classifications of data in a data set “to tell the machine exactly what patterns it should look for.”²¹⁰ The object

205 Z2 Little, *Activation Functions (Linear/Non-linear) in Deep Learning*, Medium.com, May 17, 2020, <https://medium.com/@xzz201920/activation-functions-linear-non-linear-in-deep-learning-relu-sigmoid-softmax-swish-leaky-relu-a6333be712ea>. In the context of machine learning, the concept of “prediction patterns” refers not to prediction of future events, but prediction of the missing value of an attribute. One type of prediction patterns “identify strange or abnormal events, a process known as *anomaly* or *outlier detection*.” JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 2.

206 JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 241-242. One writer termed deep learning “machine learning on steroids” because “it uses a technique that gives machines an enhanced ability to find—and amplify—even the smallest patterns.” Karen Hao, *supra* note 199.

207 JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 239.

208 DIFFUSE—Machine Learning, Centre for Advanced Internet Infrastructures, Swinburne University of Technology, [http://caia.swin.edu.au/urp/diffuse/ml.html#:~:text=Instance%3A%20An%20instance%20is%20an.by%20a%20number%20of%20attributes.&text=Attributes%20are%20often%20called%20features.\(required%20for%20supervised%20learning\)](http://caia.swin.edu.au/urp/diffuse/ml.html#:~:text=Instance%3A%20An%20instance%20is%20an.by%20a%20number%20of%20attributes.&text=Attributes%20are%20often%20called%20features.(required%20for%20supervised%20learning)).

209 See Tiago Colliri and Liang Zhao, *supra* note 32.

210 Karen Hao, *supra* note 202.

of selecting particular classes to test against a data set is to perform **clustering**: i.e., “[i]dentifying groups of similar instances in a data set.”²¹¹

In more formal terms, the goal of supervised learning is “to learn a function that maps from a set of input attribute values for an instance to an estimate of the missing value for the target attribute of the same instance.”²¹²

For instance, the primary purpose of a [machine learning] spam-filter model is to label new emails as either spam or not spam rather than to reveal the defining attributes of spam email. . . . [W]hen supervised learning is used to train a spam filter, the algorithm attempts to learn a function that maps from the attributes describing an email to a value (spam/not spam) for the target attribute; the function the learns is the spam-filter model returned by the algorithm.²¹³

Unsupervised Learning

Unsupervised learning is the second type of machine learning. While unsupervised learning also involves the use of clustering (i.e., identifying groups of similar instances), it does not involve human classification of data *a priori*²¹⁴—in other words, it does not define a target attribute in the data set.²¹⁵ Instead, the algorithm in an unsupervised-learning process

has the more general task of looking for regularities in the data. The most common type of unsupervised learning is cluster analysis, where the algorithm looks for clusters that are more similar to each other than they are to other instances in the data. These clustering algorithms often begin by guessing a set of clusters and then iteratively updating the clusters (dropping instances from one cluster and then adding them to another) so as to increase both the within-cluster similarity and the diversity across clusters.²¹⁶

211 JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 240.

212 *Id.* at 245.

213 *Id.* at 98-99.

214 See DIFFUSE—Machine Learning, *supra* note 205.

215 JOHN D. KELLEHER AND BRENDAN TIERNEY, *supra* note 17, at 100 and 246.

216 *Id.* at 102.

Semi-Supervised Learning

Semi-supervised learning is the third type of machine learning. As the term indicates, semi-supervised learning has elements of both supervised and unsupervised learning, in that human beings classify some but not all of the data *a priori*. In some situations, a semi-supervised learning may combine advantages “of working with a small labeled dataset to guide the learning process and a larger unlabeled dataset to increase the generalizability of the found solution”²¹⁷ It also can provide higher classification accuracy than purely unsupervised learning.²¹⁸

Reinforcement Learning

Reinforcement learning is the fourth type of machine learning. It is a type of dynamic programming that trains algorithms using a system of reward and punishment. A reinforcement learning algorithm, or agent, learns by interacting with its environment. The agent receives rewards for performing correctly and penalties for performing incorrectly. The agent learns without intervention from a human by maximizing its reward and minimizing its penalty.²¹⁹

Natural Language Processing

Natural language processing “is a branch of artificial intelligence that deals with the interaction between computers and humans using the natural [human] language.”²²⁰ Its ultimate objective “is to read, decipher, understand, and make sense of the human languages in a manner that is valuable. Most NLP techniques rely on machine learning to derive meaning from human languages.”²²¹

217 Mohammad Peikari, Sherine Salama, Sharon Nofech-Mozes, and Anne L. Martel, *A Cluster-then-label Semi-supervised Learning Approach for Pathology Image Classification*, 8 Scientific Reports, No. 7193 (May 8, 2018), <https://www.nature.com/articles/s41598-018-24876-0#article-info>.

218 See DIFFUSE—Machine Learning, *supra* note 205.

219 *Reinforcement Learning (RL)*, Techopedia (updated July 2, 2020), <https://www.techopedia.com/definition/32055/reinforcement-learning-rl>.

220 Dr. Michael J. Garbade, *supra* note 19.

221 *Id.*

NLP can include review of **structured data** (i.e., clearly defined types of data, such as addresses, that are easily searchable in relational databases) and **unstructured data** (i.e., less defined types of data, such as text files and email, that are more difficult to search).²²²

NLP can use both machine learning and deep learning methodologies to review and process unstructured speech and text datasets.²²³

Rule-Based Programming

Because most companies are not yet using machine learning in their current compliance programs, it is important to distinguish between machine learning and **rule-based programming**. In rule-based programming, programmers create rules that define all aspects of a particular task, typically in the form of “if/then” statements), and that are stored in a knowledge base. Rule-based programming can be used in ANI, and rules-based systems “are generally lower effort, more cost-effective, and less risky since these rules won’t change or update on their own. However, rules can limit AI capabilities with rigid intelligence that can only do what they’ve been written to do.”²²⁴

Because AI, for all of its undeniable advantages in extracting information from large data sets, can be subject to substantial hype,²²⁵ it should be noted that rule-based programming may be appropriate in certain situations—for example, when there is only a small or fixed number of outcomes needed from the programming (e.g., to click “Purchase” or not), when a 100 percent accuracy rate is essential because “[t]he penalty of error is too high to risk false positives,” when a business does not have plans (or finances) to source for machine learning,²²⁶ or when the data sets from which a business could extract meaningful information are too small to warrant the time and expense necessary for development and implementation of a machine learning solution.

²²² See, e.g., Christine Taylor, *supra* note 20.

²²³ Stephanie Overby, *Artificial intelligence (AI) vs. natural language processing (NLP): What are the differences?*, The Enterprisers Project, February 26, 2020, <https://enterpriseproject.com/article/2020/2/artificial-intelligence-ai-vs-natural-language-processing-nlp-differences>.

²²⁴ Elana Krasner, *supra* note 15.

²²⁵ Kathleen Walch, *Is AI Overhyped?*, Forbes, June 4, 2020, <https://www.forbes.com/sites/cognitiveworld/2020/06/04/is-ai-overhyped/#541db32a63ee>.

²²⁶ Elana Krasner, *supra* note 15.

On the other hand, machine learning may be appropriate in situations when data sets are so large that it is inefficient for programmers to write and modify algorithms for effective data extraction, when “there is no easily definable way to solve a task using simple rules,” when “situations, scenarios, and data are changing faster than the ability to continually write new rules,” or when tasks require NLP to review and identify meaningful words or phrases in masses of unstructured text data.²²⁷

Performance Measures for a Machine Learning Model

Four measures to judge the performance of a machine learning model are **accuracy**, **precision**, **recall**, and **the F1 score**. **Accuracy** can be defined as the fraction of correct predictions by the model (i.e., true positives (TP) and true negatives (TN)), divided by the total of all TPs, TN, false positives (FP), and false negatives (FN), or

$$\frac{TP + TN}{TP + TN + FP + FN}^{228}$$

Precision can be defined as the proportion of positive identifications that were in fact correct, or

$$\frac{TP}{TP + FP}^{229}$$

The third metric is **recall**, which can be defined as the proportion of actual positives that were defined correctly, or

$$\frac{TP}{TP + FN}^{230}$$

The **F1 score** takes precision and recall into account in measuring the accuracy of the model, by giving more weight to FN and FP while not letting large numbers of TN affect the score.²³¹

²²⁷ *Id.*

²²⁸ Google, *Classification: Accuracy*, *supra* note 94.

²²⁹ Google, *Classification: Precision and Recall*, *supra* note 96.

²³⁰ *Id.*

²³¹ See Christopher Riggio, *supra* note 99.

Appendix 2: Sources

I. Primary Sources

Interviews

This project involved interviews with representatives of 42 companies, financial institutions, associations, and non-governmental organizations, and email exchanges with one financial institution and one pharmaceutical firm. Those 42 entities fell into the following categories of industries and other organizations: airline entertainment and communications (1); associations (1); beverages (1); consulting (7); entertainment (1); financial institutions (8); information and computer technology and cybersecurity (7); law firms (1); manufacturing (7); non-governmental organizations and university schools (3); oil and gas drilling (1); pharma (2); telecom (1); and transportation (1). In certain cases, one or more company representatives participated in interviews or email exchanges, but either requested that their companies' identities not be disclosed or did not affirmatively state that they were willing to have their companies' identities disclosed.

The entities contacted are as follows:

1. 3M
2. AB InBev
3. Arm Ltd.
4. Barclays
5. BNY Mellon
6. Booz Allen Hamilton
7. Center for International Private Enterprise
8. China Merchants Bank

9. Columbia Journalism School
10. Cook Group
11. Dell
12. Fiat Chrysler
13. Fox Corporation
14. Goldman Sachs
15. Google
16. Hewlett Packard Enterprise
17. International Consortium of Investigative Journalists
18. Lextegrity
19. Lockheed Martin
20. Microsoft
21. Oxford Insights
22. Panasonic Avionics
23. PwC
24. Ropes & Gray Insights Lab
25. Uber
- 26.–42.: Anonymous

Note: Inclusion of any named entity in this list does not constitute that entity's approval or endorsement of this guidance.

II. Secondary Sources

Books

1. Stephen Finlay. *Artificial Intelligence and Machine Learning for Business*. Relativistic, Third Edition, 2018.
2. Guidehouse, *Using Machine learning to thwart financial crime* (2020), <https://guidehouse.com/-/media/www/site/insights/financial-services/2020/ai-financial-crime-final.pdf>
3. John D. Kelleher and Brendan Tierney, *Data Science*. MIT Press, Cambridge, 2018.
4. Melanie Mitchell. *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux, New York, 2019.
5. Janelle Shane. *You Look Like a Thing and I Love You*. Voracious/Little, Brown, New York, 2019.
6. Mariya Yao, Marlene Jia, and Adelyn Zhou. *Applied Artificial Intelligence: A Handbook for Business Leaders*. TOPBOTS, 2018.

Government Documents

1. Roger Alford, Deputy Assistant Attorney General, Antitrust Division, U.S. Department of Justice, *Antitrust Enforcement and the Fight Against Corruption*, Remarks for the Conference on Rule of Law and Anti-Corruption Challenges, University of Notre Dame and University of Sao Paulo (October 3, 2017), <https://www.justice.gov/opa/speech/file/1001076/download>
2. Criminal Division, U.S. Department of Justice and Enforcement Division, Securities and Exchange Commission, *A Resource Guide to the U.S. Foreign Corrupt Practices Act 59-65* (2d ed. 2020), <https://www.justice.gov/criminal-fraud/file/1292051/download>
3. Criminal Division, U.S. Department of Justice, *Evaluation of Corporate Compliance Programs* (updated June 2020), <https://www.justice.gov/criminal-fraud/page/file/937501/download>
4. Deferred Prosecution Agreement, *United States v. Commonwealth Edison Co.* (N.D. Ill., signed July 17, 2020), available at <https://www.justice.gov/usao-ndil/press-release/file/1295241/download>

5. Deferred Prosecution Agreement, *United States v. Goldman Sachs Group*, Criminal No. 20-437 (MKB), ¶¶4(f) at 5 and 54 at 19 (E.D.N.Y., filed October 22, 2020), <https://www.justice.gov/usao-edny/press-release/file/1329961/download>
6. Department of Economic and Social Affairs, United Nations, Sustainable Development: The 17 Goals, <https://sdgs.un.org/goals>
7. European Parliament, Release: Parliament leads the way on first set of EU rules for Artificial Intelligence, October 20, 2020, <https://www.europarl.europa.eu/news/en/headlines/priorities/artificial-intelligence-in-the-eu/20201016IPR89544/parliament-leads-the-way-on-first-set-of-eu-rules-for-artificial-intelligence>
8. European Parliament, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN>
9. *First National Bank of Boston v. Bellotti*, 435 U.S. 789 (1978)
10. Information Commissioner's Office, *Guide to the General Data Protection Regulation (GDPR)/The principle* (accessed October 1, 2020), https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/#the_principles
11. Non-Prosecution Agreement and Statement of Facts, Microsoft Magyarország Számítástechnikai Szolgáltató és Kereskedelmi Kft. (July 22, 2019), <https://www.justice.gov/opa/press-release/file/1185686/download>
12. Presidência da República, Lei Geral de Proteção de Dados Pessoais, Lei Nº 13.709, de 14 de Agosto de 2018, http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm
13. SB 1392 Consumer Data Protection Act, Legislative Information System, <https://lis.virginia.gov/cgi-bin/legp604.exe?212+sum+SB1392>
14. Securities and Exchange Commission, Order Instituting Cease-and-Desist Proceedings Pursuant to Section 21C of the Securities Exchange Act of 1934, Making Findings, and Imposing a Cease-and-Desist Order, *In re Alexion Pharmaceuticals, Inc.*, No. 3-19852 (July 2, 2020), <https://www.sec.gov/litigation/admin/2020/34-89214.pdf>

15. United Kingdom Financial Conduct Authority, Global AML and Financial Crime TechSprint, <https://www.fca.org.uk/events/techsprints/aml-financial-crime-international-techsprint>
16. *United States v. Stanford*, 805 F.3d 557 (5th Cir. 2015), cert. denied, 137 S. Ct. 491 (2016)

Corporate Documents

1. AB InBev, 2020 Annual Report (February 25, 2021), https://www.ab-inbev.com/content/dam/abinbev/news-media/press-releases/2021/02/AB%20InBev%202020%20Annual%20Report_FINAL.pdf
2. Alexion Pharmaceuticals, 2019 Annual Report (February 4, 2020), <https://ir.alexion.com/static-files/8394a14e-2ae1-4b3d-aa27-96c5f4b2a4dc>
3. AI Ethics Guidelines Global Inventory, Algorithm Watch (updated April 2020), <https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/>
4. Association for Computing Machinery, US Public Policy Council, Statement on Algorithmic Transparency and Accountability (January 12, 2017), https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf
5. Deutsche Telekom, Digital Ethics Guidelines on AI (July 19, 2018), available at <https://www.telekom.com/en/company/digital-responsibility/details/artificial-intelligence-ai-guideline-524366>
6. IBM, Everyday Ethics for Artificial Intelligence (2019 ed.), <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>
7. Microsoft, Earnings Release FY21 Q2, <https://www.microsoft.com/en-us/Investor/earnings/FY-2021-Q2/income-statements>
8. Microsoft, Facts About Microsoft, <https://news.microsoft.com/facts-about-microsoft/#About>
9. Microsoft, Microsoft AI principles, available at <https://www.microsoft.com/en-us/ai/responsible-ai>
10. United Nations Office on Drugs and Crime, Introduction to money-laundering, <https://www.unodc.org/unodc/en/money-laundering/introduction.html>

11. Christiane Wild-Raidt, Release, AI code of ethics: Bosch sets company guidelines for the use of artificial intelligence, bosch-presse.de. February 19, 2020, <https://www.bosch-presse.de/pressportal/de/en/ai-code-of-practice-bosch-sets-company-guidelines-for-the-use-of-artificial-intelligence-208384.html>

Reports

1. Per Aarvik, *Artificial Intelligence—a promising anti-corruption tool in development settings?*, U4 Report 2019:1, <https://www.u4.no/publications/artificial-intelligence-a-promising-anti-corruption-tool-in-development-settings>
2. Finextra with Feedzai, *Utilising AI to Prevent Financial Crime* (May 2019), <https://feedzai.com/wp-content/uploads/2019/04/q2-19-utilizing-ai-to-fight-financial-crime.pdf>
3. One Hundred Year Study on Artificial Intelligence, Stanford University, Artificial Intelligence and Life in 2030: Report of the 2015 Study Panel 12 (September 2016), https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf
4. Oxford Insights, *The Next Generation of Anti-Corruption Tools: Big Data, Open Data & Artificial Intelligence 3* (May 2019), https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/5ced49ccc8302518cb27f64b/1559054797862/Research+report+2019_+The+Next+Generation+of+Anti-Corruption+Tools_++Big+Data%2C+Open+Data+%26++Artificial+Intelligence.pdf
5. SAP AI Ethics Steering Committee, *SAP's Guiding Principles for Artificial Intelligence* (September 2018), <https://www.sap.com/products/intelligent-technologies/artificial-intelligence/ai-ethics.html?pdf-asset=940c6047-1c7d-0010-87a3-c30de2ffd8ff&page=1>

Peer-Reviewed and Other Research Articles and Papers

1. Bryan Casey, Ashkan Farhangi. and Roland Vogl, *Rethinking Explainable Machines: The GDPR's "Right to Explanation" and the Rise of Algorithmic Audits in Enterprise*, 34 Berkeley Technology Law Journal 145, 151 (2019)
2. Tiago Colliri and Liang Zhao, *Analyzing the Bills-Voting Dynamics and Predicting Corruption-Convictions Among Brazilian Congressmen Through Temporal Networks*, 9 Scientific Reports 16754 (2019), <https://www.nature.com/articles/s41598-019-53252-9>

3. Matthew Galvin, Ivy Munoko, and Miklos Vasarelhi, *A Collaboration Framework for Democratizing Compliance Analytics*, 14 *International Review of Compliance and Business Ethics* 8, 9 (October 2020)
4. Matthew Galvin, *Overview of Data Analytics Network: AB InBev White Paper for Potential Tech Partner 2* (March 26, 2020)
5. Félix J. López-Iturriaga, Iván Pastor Sanz, *Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces*, 140 *Social Indicators Research* 975-998 (2017), available at <https://link.springer.com/article/10.1007%2Fs11205-017-1802-2>
6. W. James Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu, *Definitions, methods, and applications in interpretable machine learning*, 44 *Proceedings of the National Academy of Sciences* 22071 (October 16, 2019), <https://www.pnas.org/content/116/44/22071>
7. Mohammad Peikari, Sherine Salama, Sharon Nofech-Mozes, and Anne L. Martel, *A Cluster-then-label Semi-supervised Learning Approach for Pathology Image Classification*, 8 *Scientific Reports*, No. 7193 (May 8, 2018), <https://www.nature.com/articles/s41598-018-24876-0#article-info>
8. Eugene Soltes, *Designing a Compliance Program at AB InBev*, Harvard Business School Paper No. 9-118-071 at 2 (revised April 30, 2018).
9. Mayank Tripathi, *Underfitting and Overfitting in Machine Learning*, Data Science Foundation, June 13, 2020, <https://datascience.foundation/sciencewhitepaper/underfitting-and-overfitting-in-machine-learning>
10. Xiaojin Zhu and Zoubin Ghahramani, *Learning from Labeled and Unlabeled Data with Label Propagation* (June 2002), available at <http://reports-archive.adm.cs.cmu.edu/anon/cald/CMU-CALD-02-107.pdf>

Blog Posts and Websites

1. AB InBev, *How BrewRIGHT is rooting out corruption at AB InBev and beyond*, January 31, 2020, <https://www.ab-inbev.com/news-media/innovation/how-brewright-is-rooting-out-corruption-at-ab-inbev-and-beyond.html>
2. Valerie Charles, *Microsoft's Alan Gibson on the Power of Compliance Data*, GAN Integrity, May 11, 2020, <https://www.ganintegrity.com/blog/microsofts-alan-gibson-on-the-power-of-compliance-data/>

3. Chartis, *AI in RegTech: A quiet upheaval* (2018), available at <https://www.ibm.com/downloads/cas/NAJXEKE6>
4. Damian Chen, *Why You Need Data Transformation in Machine Learning*, Datanami, November 8, 2019, <https://www.datanami.com/2019/11/08/why-you-need-data-transformation-in-machine-learning/>
5. Jeffrey P. Cunard, Luke Dembosky, Avi Gesser, Henry Lebowitz, Jim Pastore, Lisa Zornberg, Anna R. Gressel, and Steven Tegrar, *Fifteen Ways to Reduce Regulatory and Reputational Risks for Your AI-Powered Applications—Lessons from Recent Court Decisions and Regulatory Activity*, Debevoise in Depth, February 20, 2020, available at <https://www.debevoise.com/insights/publications/2020/02/fifteen-ways-to-reduce-regulatory-and>
6. Dell Technologies, *Leveraging AI for Good—A Global Opportunity for Policy-Makers*, September 2, 2019, <https://www.delltechnologies.com/en-us/perspectives/leveraging-ai-for-good-a-global-opportunity-for-policy-makers/>
7. Deloitte, *Why artificial intelligence is a game changer for risk management*, <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/audit/us-ai-risk-powers-performance.pdf>
8. EY, *How Teva Pharmaceutical is managing third-party risk better*, EY, https://www.ey.com/en_us/forensic-integrity-services/how-teva-pharmaceutical-is-managing-third-party-risk-better
9. Alan Gibson, Microsoft, *Compliance Analytics Program* [PowerPoint], May 11, 2020
10. Google, *Introduction to Machine Learning Problem Framing*, <https://developers.google.com/machine-learning/problem-framing>
11. Google, *Machine Learning Crash Course*, <https://developers.google.com/machine-learning/crash-course>
12. Alexandre Gonfalonieri, *How to Build A Data Set For Your Machine Learning Project*, Towards Data Science, February 13, 2019, <https://towardsdatascience.com/how-to-build-a-data-set-for-your-machine-learning-project-5b3b871881ac>
13. Google, *Learn from ML experts at Google*, <https://ai.google/education/>
14. International Association of Privacy Professionals, *The General Data Protection Regulation Matchup Series*, <https://iapp.org/resources/article/the-general-data-protection-regulation-matchup-series/>

15. Elana Krasner, *How to choose between rule-based AI and machine learning*, TechTalks.com, June 5, 2020, <https://bdtechtalks.com/2020/06/05/rule-based-ai-vs-machine-learning/>
16. Lextegrity, *Integrity Analytics Collective*, <https://www.lextegrity.com/collective>
17. Lextegrity, *Integrity Gateway Monitoring*, www.lextegrity.com/monitoring
18. National Conference of State Legislatures, *2020 Consumer Data Privacy Legislation* (October 9, 2020), <https://www.ncsl.org/research/telecommunications-and-information-technology/2020-consumer-data-privacy-legislation637290470.aspx>
19. Stephanie Overby, *Artificial intelligence (AI) vs. natural language processing (NLP): What are the differences?*, The Enterprisers Project, February 26, 2020, <https://enterpriseproject.com/article/2020/2/artificial-intelligence-ai-vs-natural-language-processing-nlp-differences>
20. *Overfitting in Machine Learning: What It Is and How to Prevent It*, Elite Data Science, <https://elitedatascience.com/overfitting-in-machine-learning>
21. Pamela Passman, *Artificial Intelligence: Evolving Risks and Responsibilities*, CIO Review, <https://storage.cioreview.com/cxoinight/artificial-intelligence-evolving-risks-and-responsibilities-nid-30215-cid-12.html>
22. André Petheram and Isak Nti Asare, *From open data to artificial intelligence: the next frontier in anti-corruption*, Oxford Insights, July 27, 2018, <https://www.oxfordinsights.com/insights/aiforanticorruption>
23. *Prediction*, Data Robot, <https://www.datarobot.com/wiki/prediction/#:~:text=%E2%80%9CPrediction%E2%80%9D%20refers%20to%20the%20output,will%20churn%20in%2030%20days.&text=The%20word%20%E2%80%9Cprediction%E2%80%9D%20can%20be%20misleading>
24. PwC, *Risk Command*, <https://www.pwc.com/us/en/products/risk-command.html>
25. PwC, *Transforming compliance into an asset*, <https://www.pwc.com/us/en/library/case-studies/microsoft-fighting-corruption-using-risk-command-platform-digital-technologies.html>
26. *Reinforcement Learning (RL)*, Techopedia (updated July 2, 2020), <https://www.techopedia.com/definition/32055/reinforcement-learning-rl>

27. Vinay Sharma, *Can artificial intelligence stop corruption in its tracks?*, Governance for Development, November 15, 2018, <http://blogs.worldbank.org/governance/can-artificial-intelligence-stop-corruption-its-tracks>
28. Koo Ping Shung, *Accuracy, Precision, Recall or F1?*, Towards Data Science, March 15, 2018, <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>
29. Robert Smith, *The Key Differences Between Rule-Based AI And Machine Learning, Becoming Human*, July 14, 2020, <https://becominghuman.ai/the-key-differences-between-rule-based-ai-and-machine-learning-8792e545e6>
30. Kathleen Walch, *Is AI Overhyped?*, Forbes, June 4, 2020, <https://www.forbes.com/sites/cognitiveworld/2020/06/04/is-ai-overhyped/#541db32a63ee>
31. Z2 Little, *Activation Functions (Linear/Non-linear) in Deep Learning*, Medium.com, May 17, 2020, <https://medium.com/@xzz201920/activation-functions-linear-non-linear-in-deep-learning-relu-sigmoid-softmax-swish-leaky-relu-a6333be712ea>

General Media Articles and Commentaries

1. Ajay Agrawal, Joshua Gans, and Avi Goldfarb, *How to Win with Machine Learning*, Harvard Business Review, September–October 2020, <https://hbr.org/2020/09/how-to-win-with-machine-learning>
2. *Artificial Intelligence and Risk Management*, Enterprise Risk, <https://enterpriseriskmag.com/artificial-intelligence-risk-management/>
3. Sidath Asiri, *Machine Learning Classifiers*, Towards Data Science, June 11, 2018, <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623>
4. *Enlisting AI And Biometrics In The Fight Against Digital Identity Theft*, PYMNTS, May 18, 2020, <https://www.pymnts.com/authentication/2020/enlisting-artificial-intelligence-biometrics-fight-against-digital-identity-theft/>
5. *Mini-Roundtable: Data Analytics and AI for Anti-Corruption Compliance*, Risk & Compliance Magazine, January–March 2019, available at <https://guidehouse.com/-/media/www/site/insights/gic/2019/corporate-disputes-magazine--data-analytics-and-ai.pdf>
6. *What is a business case?*, Association for Project Management, <https://www.apm.org.uk/resources/what-is-project-management/what-is-a-business-case/>

7. Stuart Brock, *Legal Contracts on the Frontline of Fighting Corruption with Artificial Intelligence*, International Banker, May 21, 2019, <https://internationalbanker.com/technology/legal-contracts-on-the-frontline-of-fighting-corruption-with-artificial-intelligence/>
8. Paloma Cantero-Gomez, *How To Frame A Problem To Find The Right Solution*, Forbes, April 10, 2019, <https://www.forbes.com/sites/palomacanterogomez/2019/04/10/how-to-frame-a-problem-to-find-the-right-solution/?sh=2e7133dc5993>
9. Joseph M. Carew, *How to choose between a rules-based vs. machine learning system*, Tech Target, July 23, 2020, <https://searchenterpriseai.techtarget.com/feature/How-to-choose-between-a-rules-based-vs-machine-learning-system>
10. Dan Clark, *How Anheuser-Busch Compliance Head Uses Analytics to Handle COVID-19 Challenges*, Corporate Counsel, July 1, 2020, <https://www.law.com/corpocounsel/2020/07/01/how-anheuser-busch-compliance-head-uses-analytics-to-handle-covid-19-challenges/>
11. Louis Columbus, *Top 9 Ways Artificial Intelligence Prevents Fraud*, Forbes, July 9, 2019, <https://www.forbes.com/sites/louiscolumbus/2019/07/09/top-9-ways-artificial-intelligence-prevents-fraud/#44af29d514b4>
12. EY, *Teva Pharmaceutical Reengineers Compliance With Data Analytics*, MIT Sloan Management Review, July 8, 2020, <https://sloanreview.mit.edu/sponsors-content/teva-pharmaceutical-reengineers-compliance-with-data-analytics/>
13. Ragnar Fjelland, *Why general artificial intelligence will not be realized*, 7 Humanities and Social Sciences Communications 10 (June 17, 2020), available at <https://www.nature.com/articles/s41599-020-0494-4>
14. Dr. Michael J. Garbade, *A Simple Introduction to Natural Language Processing*, Becoming Human, October 15, 2018, <https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32>
15. Anastasia Gorina, *Class imbalance problem in classification*, Towards Data Science, April 2, 2020, <https://towardsdatascience.com/class-imbalance-problem-in-classification-a2ddaba98f4a>
16. Gary Grossman, *Why Businesses Should Adopt an AI Code of Ethics—Now*, Information Week, November 14, 2019, <https://www.informationweek.com/big-data/ai-machine-learning/why-businesses-should-adopt-an-ai-code-of-ethics---now-/a/d-id/1336207>

17. Karen Hao, *Establishing an AI code of ethics will be harder than people think*, MIT Technology Review, October 21, 2018, <https://www.technologyreview.com/s/612318/establishing-an-ai-code-of-ethics-will-be-harder-than-people-think/>
18. Karen Hao, *What is machine learning?*, MIT Technology Review, November 17, 2018, <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/>
19. Larry Hardesty, *Explained: Neural networks*, MIT News, April 14, 2017, <http://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>
20. Nick Heath, *What is artificial general intelligence?*, ZDNet, August 22, 2018, <https://www.zdnet.com/article/what-is-artificial-general-intelligence/>
21. Paul Huck, Aaron Johnson, Nicholas Kiritz, and C. Erik Larson, *Why AI Governance Matters*, RMA Journal, May 2020, at 18, https://www.promontory.com:3000/static/pdf/1588624225_title.pdf
22. Jaclyn Jaeger, *Six steps for developing an AI ethics framework*, Compliance Week, November 20, 2019, <https://www.complianceweek.com/artificial-intelligence/six-steps-for-developing-an-ai-ethics-framework/28078.article>
23. Sam Leon, *How Can We Use Artificial Intelligence to Help Us Fight Corruption in the Mining Sector?*, Global Witness, November 8, 2018, https://www.globalwitness.org/en/blog/how-can-we-use-artificial-intelligence-help-us-fight-corruption-mining-sector/?gclid=Cj0KCQiApt_xBRDxARIsAAMUMu8uG_ER9G4UwpavxXsROO_gp2RW4SLgqng1i4VvAWxLBilpoNoc_pUaAkJBEALw_wcB
24. Jeremy B. Merrill, *How Quartz used AI to sort through the Luanda Leaks*, Quartz, January 19, 2020, <https://qz.com/1786896/ai-for-investigations-sorting-through-the-luanda-leaks/>
25. Cindy Moehring, *Season 3, Episode 7: Matt Galvin | How Technology Can Be Used in Business To Catch Potential Crimes Before They Happen*, Business Integrity Leadership Initiative, Sam M. Walton College of Business, University of Arkansas (March 4, 2021), <https://walton.uark.edu/business-integrity/blog/matt-galvin.php>
26. Sarang Narkhede, *Understanding Confusion Matrix*, Towards Data Science, May 9, 2018, <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>

27. Syed Sadat Nazrul, *Fraud Detection Under Extreme Class Imbalance*, Towards Data Science, April 12, 2018, <https://towardsdatascience.com/fraud-detection-under-extreme-class-imbalance-c241854e60c>
28. Pamela Passman, *Artificial Intelligence: Evolving Risks and Responsibilities*, CIO Review, <https://storage.cioreview.com/cxinsight/artificial-intelligence-evolving-risks-and-responsibilities-nid-30215-cid-12.html>
29. Tibi Puiu, *Artificial intelligence for fraud detection is bound to save billions*, ZME Science, March 23, 2020, <https://www.zmescience.com/science/ai-fraud-detection-0942323/>
30. Christopher Riggio, *What's the deal with Accuracy, Precision, Recall and F1?*, Towards Data Science, November 1, 2019, <https://towardsdatascience.com/whats-the-deal-with-accuracy-precision-recall-and-f1-f5d8b4db1021>
31. Isha Salian, *SuperVize Me: What's the Difference Between Supervised, Unsupervised, Semi-Supervised and Reinforcement Learning?*, NVIDIA, August 2, 2018, <https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/#:~:text=In%20a%20supervised%20learning%20model,and%20patterns%20on%20its%20own>
32. Catherine Stupp, *EU Restrictions Could Force Companies to Change Data Transfer Practices*, Wall Street Journal, November 17, 2020, <https://www.wsj.com/articles/eu-restrictions-could-force-companies-to-change-data-transfer-practices-11605609001>
33. Xuning (Mike) Tang and Yihua Astle, *The Impact of Deep Learning on Anomaly Detection*, Law.com, August 10, 2020, <https://www.law.com/legaltechnews/2020/08/10/the-impact-of-deep-learning-on-anomaly-detection/>
34. Tom Taulli, *AI (Artificial Intelligence) Governance: How To Get It Right*, Forbes, October 10, 2020, <https://www.forbes.com/sites/tomtaulli/2020/10/10/ai-artificial-intelligence-governance-how-to-get-it-right/#7d44e512745f>
35. [Tom Taulli](#), *CCPA: What Does It Mean For AI (Artificial Intelligence)?*, Forbes, December 27, 2019, <https://www.forbes.com/sites/tomtaulli/2019/12/27/ccpa--what-does-it-mean-for-ai-artificial-intelligence/?sh=177694a46bb2>
36. Christine Taylor, *Structured vs. Unstructured Data*, Datamation, March 28, 2018, <https://www.datamation.com/big-data/structured-vs-unstructured-data.html>

37. Dylan Tokar, *AB InBev Taps Machine Learning to Root Out Corruption*, Risk & Compliance Journal, Wall Street Journal, January 17, 2020, <https://www.wsj.com/articles/ab-inbev-taps-machine-learning-to-root-out-corruption-11579257001>
38. Dylan Tokar, *Anheuser-Busch InBev's BrewRight: How It Works*, Risk & Compliance Journal, Wall Street Journal, January 17, 2020, https://www.wsj.com/articles/anheuser-busch-inbevs-brewright-how-it-works-11579257000?mod=article_inline
39. Dylan Tokar, *Corporate Compliance Programs Hit Refresh With Data-Analytics Tools*, Wall Street Journal, September 22, 2020, <https://www.wsj.com/articles/corporate-compliance-programs-hit-refresh-with-data-analytics-tools-11600767001>
40. Rodney Weidemann, *Using AI to uncover fraud and corruption*, IT Web, February 21, 2019, <https://www.itweb.co.za/content/KBpdgvpPDzavLEew>
41. Joe Williams, *How \$132 billion brewery giant AB InBev is using AI to fight corruption and spot business fraud around the globe*, Business Insider, December 13, 2019, <https://www.businessinsider.com/ab-inbev-brewing-artificial-intelligence-to-spot-fraud>
42. Ellen Zimiles and Tim Mueller, *How AI is transforming the fight against money laundering*, World Economic Forum, January 17, 2019, <https://www.weforum.org/agenda/2019/01/how-ai-can-knock-the-starch-out-of-money-laundering/>

Acknowledgements

This guidance was prepared by the Coalition for Integrity and edited by Shruti Shah, President, and CEO of the Coalition for Integrity. The research and drafting of the guidance was conducted by Jonathan J. Rusch, Adjunct Professor at Georgetown University Law Center and American University Washington College of Law.

The funding for this report was provided by CREATE.org. While Coalition for Integrity benefited greatly from the funding, and from the work provided by the foregoing persons, and from the meetings and conversations that they had with many representatives of companies, financial institutions, law and consulting firms, and non-governmental organizations, this guidance, including its analysis and conclusions, represents the views of the Coalition for Integrity and does not necessarily reflect the views of those who provided funding, information, advice, and services to the guidance.



About the Coalition for Integrity

The Coalition for Integrity (C4I) is a 501(c)(3) research organization that focuses its leadership and advocacy primarily on ending impunity throughout the world, promoting transparency and accountability in U.S. government and elections, and fostering greater integrity in the private sector.

Corruption, bribery, lack of accountability, and subterfuge by governmental agencies, non-governmental organizations, and private-sector entities breed distrust. Corruption diverts resources from communities—often those in the most need. It also erodes credibility, distorts competition, rewards dishonest actors, and undermines development. Perhaps most troubling, corruption and impunity shake citizens' faith in institutions that are responsible for protecting human rights, keeping nations secure, and maintaining functioning societies.

The Coalition for Integrity brings together individuals and organizations in government, business, labor, academia, and civil society that are committed to encouraging openness and accountability in all sectors. Through deeply researched reports, detailed guides, and other resources, C4I disseminates information and tools that assist institutions that wish to foster greater transparency, and hold accountable those that are resistant to upholding honesty, integrity, and openness in their practices.

©2021

Coalition for Integrity





Coalition for Integrity

