



OXFORD INSIGHTS

The Next Generation of Anti-Corruption Tools: Big Data, Open Data & Artificial Intelligence

Research Report, May 2019

André Petheram

Walter Pasquarelli

Richard Stirling



Index

Introduction	3
Using open data for anti-corruption	4
What data makes a difference?	5
Open procurement data: tendering and contracting	5
Ukraine	6
Big data, the European Union, and Digiwhist	6
Artificial intelligence: the next step	8
The theory	8
Implementation and benefits	9
Conclusion: ongoing projects and final thoughts	11
Bibliography	12



Introduction

Corruption is a serious global problem, transcending borders, political affiliations and systems of government. As more governments digitise their operations, there is an increased opportunity to develop practical anti-corruption tools based on big data, open data and artificial intelligence. Here, we consider the viability of such tools, in the hope that their implementation would eventually lead to fewer public funds lost to the private accounts of corrupt officials and politicians.

Specifically, we suggest that the conditions are right to test artificial intelligence tools for anti-corruption in a particular group of countries: Argentina, Brazil, Bulgaria, Colombia, Mexico, Paraguay, Romania, Slovakia, Russia, and Ukraine. They score highly in open data rankings, but have high levels of perceived corruption. In these places, therefore, there is a potentially large amount of data available for developers to use to train an AI anti-corruption tool.

The International Monetary Fund (IMF) estimates that US\$1 trillion of global government revenues is lost to corruption each year.¹ This would be enough to pay for the annual UK NHS budget nine times over. In Afghanistan, a country with one of the highest levels of corruption in the world, it would cover all annual government expenditure nearly 200 times.²

Yet corruption remains intractable. The proceeds of corrupt activities are often moved to accounts in offshore jurisdictions, making them hard to monitor and control without concerted collaboration between governments. Neither activists nor authorities are making much progress to reduce corruption.

¹ IMF (2019a). *Fiscal Monitor: Curbing Corruption*. <https://www.imf.org/en/Publications/FM/Issues/2019/03/18/fiscal-monitor-april-2019> [accessed 17 April 2019].

² *Full Fact* (2018). Spending on the NHS in England. <https://fullfact.org/health/spending-english-nhs/>. [Accessed 21 May 2019]; Islamic Republic of Afghanistan (2019). *National Budget, Fiscal Year 1398*. https://www.budgetmof.gov.af/images/stories/DGB/BPRD/National%20Budget/Fiscal_Year_1398/1398%20National%20Budget%20English%20Version.pdf [accessed 21 May 2019].



Following their 2018 Corruption Perceptions Index, which measures the extent to which people believe corruption is prevalent in their countries, Transparency International reported that ‘most countries are failing to make serious inroads against corruption’.³

Corruption is a complex and varied phenomenon, defined broadly by Transparency International as ‘the abuse of entrusted power for private gain’. It is common to distinguish between ‘petty’ and ‘grand’ corruption. The ‘petty’ form occurs when officials withhold everyday public services from citizens unless personally recompensed (usually through bribes). ‘Grand’ corruption, meanwhile, is the misuse of high office for individual benefit.⁴

Especially within grand corruption, corruption might involve unfairly directing public money to private accounts, meaning that citizens do not receive full value from the system of government taxation and expenditure. This happens, for example, when a government official directs a public contract to a particular company; that company then secretly returns part of the money received from the government back to the original official. In cases such as these, corruption can be thought of as a fiscal ‘leakage’.⁵

This is an incomplete definition of corruption. However, it remains a useful concept because it focuses attention on an obvious area of corruption risk: the point at which governments spend money. The WTO suggests that public procurement comprises 10-15% of world GDP;⁶ for OECD countries the figure is 12%;⁷ for the EU, it is 16%.⁸ Accessing just a portion of this is a tempting

³ *Transparency International* (2019a). How corruption weakens democracy.

https://www.transparency.org/news/feature/cpi_2018_global_analysis [accessed 25 April 2019].

⁴ *Transparency International* (2019d). What is corruption?

<https://www.transparency.org/what-is-corruption#define> [accessed 6 May 2019]

⁵ IMF (2019a). *Fiscal Monitor: Curbing Corruption*. See also IMF (2019b). The trillion-dollar question. The IMF and anti-corruption one year on.

https://www.transparency.org/news/feature/the_trillion_dollar_question_the_imf_and_anti_corruption_one_year_on [accessed 25 April 2019].

⁶ Djankov, S., Islam, A., Saliola, F. (2016). How Large Is Public Procurement in Developing Countries? *PIIE*, 7 November 2016.

<https://piie.com/blogs/realtime-economic-issues-watch/how-large-public-procurement-developing-countries> [accessed 17 April 2019].

⁷ *OECD* (2016). Preventing Corruption in Public Procurement.

<http://www.oecd.org/gov/ethics/Corruption-Public-Procurement-Brochure.pdf> [accessed 17 April 2019].



prospect for corrupt officials and politicians. In the EU, for example, RAND Corporation estimates that corruption connected to public procurement totals about €5 billion annually.⁹ It is reasonable to assume that the amount is higher in regions with more widespread corruption than the EU.

Moreover, with an apparent trend towards digitising procurement, often by running government purchasing through online portals, there is a large amount of data that relates to public tenders, contracts and suppliers. This is frequently available openly, providing a valuable resource for anyone seeking to understand the dynamics of corruption through large-scale data analysis. This explains the broad emphasis on procurement in this report.

Big and open data might tell us more about the nature of corruption - including its causal relationships and scale - than we already knew. Such data potentially provides the basis for automated monitoring and enforcement tools with previously unparalleled power. Moreover, with big data comes the potential to use machine learning to radically deepen our ability to know where corruption begins in government. Crucially, activists, anti-corruption authorities and international watchdogs might identify new cases of corruption and reduce the fiscal loss that corruption frequently causes.

With huge thanks to the Thom Townsend and Warren Smith at the UK Government, and Ágnes Czibik at the Government Transparency Institute, for their comments and advice.

⁸ *European Commission* (2017). Public procurement.

http://ec.europa.eu/trade/policy/accessing-markets/public-procurement/index_en.htm [accessed 17 April 2019].

⁹ *RAND Corporation* (2016). The Cost of Corruption in Europe - up to €990 Billion (£781.64 Billion) Lost Annually. <https://www.rand.org/news/press/2016/03/22.html> [accessed 17 April 2019].

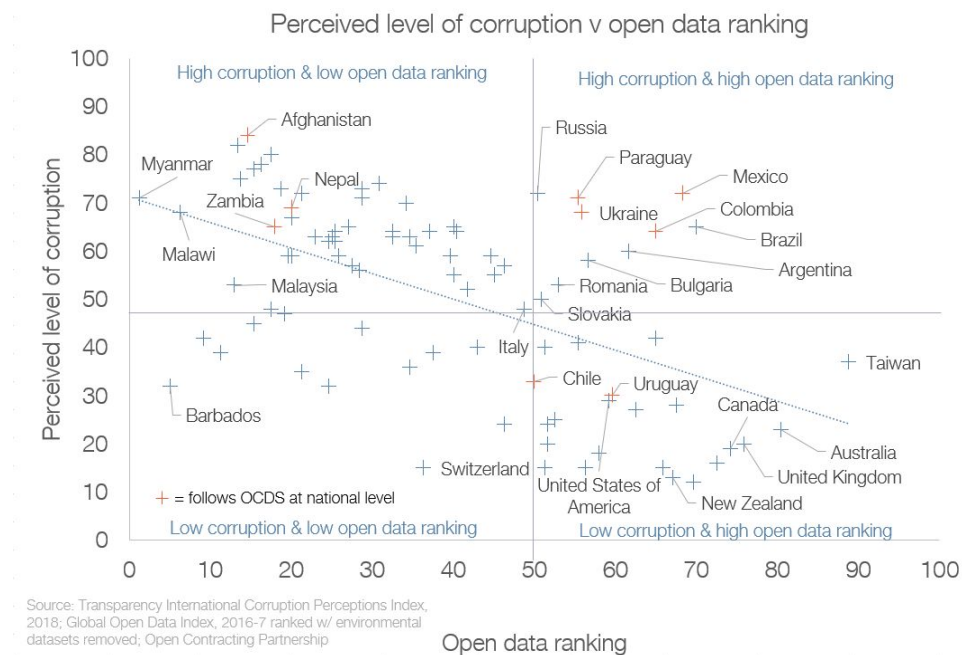


Using open data for anti-corruption

Alongside a well-functioning system of sanctions and prosecutions, and a strong free press, transparency in government is crucial to reducing corruption. The IMF's recent research identified transparency about government finances and expenditure as the most important factor in increasing control over corruption in countries with high corruption.¹⁰

In recent years open data has become one of the most important methods of enabling transparency. With the increased digitisation of public services and government processes, governments have opportunities to release datasets of progressively higher quality, extent and variety. And, as Chart 1 shows, there is a negative correlation between the perceived level of corruption in a country and the amount of open data available:

Chart 1: the lower a country scores on open data, the higher its level of perceived corruption.



¹⁰ IMF (2019). *Fiscal Monitor: Curbing Corruption*.



Recent years, therefore, have seen an increase in the number of guides and standards designed to help governments publish data in a way that allows for easy and efficient scrutiny. These include the *Open Up Guide: Using Open Data to Combat Corruption*, which has been introduced in Mexico as an ‘official standard’ in its open data policy,¹¹ and the Open Contracting Data Standard (OCDS), a model for releasing contracting data in a structured format which is used in Mexico, Ukraine, Colombia, Canada and the UK, among other countries.¹² However, returning to Chart 1 above, there remain countries with high open data rankings and high levels of corruption that are yet to adopt these approaches.

What data makes a difference?

For those investigating and seeking to reduce corruption, access to certain datasets is particularly useful for interrogating a government’s activities. Investigative journalists and civil society organisations will likely rely heavily on data that is openly available, such as procurement data, and, in some cases, beneficial ownership data. In addition, anti-corruption authorities may have access to further data that is not publicly available for security, legal or political reasons: government payments data, for example. Of these three examples, open procurement data is probably most easily accessible and useful in its current form, and is examined in depth here.

Certainly, information on beneficial ownership, or who *ultimately* owns a company, is potentially vital: money accessed through corruption is likely to be moved through accounts registered with a series of companies, each owned by another, until the final owner becomes hard to detect.¹³ The

¹¹ Zapata, E. (2017). Implementation of the Anticorruption Open Up Guide in Mexico. *Open Data Charter*. <https://medium.com/opendatacharter/implementation-of-the-anticorruption-open-up-guide-in-mexico-a16015fe3b3d> [accessed 18 April 2019].

¹² OCP (2019a). Worldwide. <https://www.open-contracting.org/why-open-contracting/worldwide/#/> [accessed 18 April 2019].

¹³ see Bullough, O. (2018). *Moneyland: Why Thieves & Crooks Now Rule The World & How To Take It Back*. London: Profile Books.



OpenOwnership Register is an ongoing project seeking to collect and open global beneficial ownership information. It currently shows beneficial ownership data from the UK, Slovakia, Denmark and Ukraine.¹⁴

Data that *directly* and *systematically* shows individual government payments, taken from accounts payable systems, is even harder to find. Government departments in the UK must publish details of all spending above £25,000, and show the supplier, category, and transaction number (see data.gov.uk), but few other governments do the same. Even in the UK, the data is not available in one place from a central portal, making it harder to combine for a unified overview of government spending, and data is sometimes published irregularly. Some other countries make expenditure data available that can be broken down to particular account codes (Paraguay, for example), but it is not clear that the data is available on the level of the individual *transaction*.

Governments are, of course, sensitive about exposing the fine details of their expenditure. In many cases, they may not collect payments data within a central financial management system. However, where available, payments data should be an important target for transparency activists, potentially providing irreplaceably granular information on where public funds go.

Open procurement data: tendering and contracting

Open data proponents argue that by systematically publishing information on contract amounts, awards and the number of bidders, for example, those contemplating corrupt activities will be discouraged. Their activities will be detectable by the institutions that hold governments and officials to account. Alternatively, where contracts are found to have been awarded to close associates of officials, or project costs have been clearly inflated, anti-corruption authorities will have the basis for a prosecution.

¹⁴ *OpenOwnership* (2019). FAQ. <https://register.openownership.org/faq> [accessed 25 April 2019].



Ukraine and the European Union provide powerful examples of open procurement data in use.

The Open Contracting Data Standard

The OCDS provides a step-by-step guide helping governments publish procurement data across 'all stages of the contracting process'.¹⁵ This process consists of 5 stages: planning (pre-procurement), the initial tender, the contract award, contract finalisation, and implementation (post-procurement). The guide is intended to help compliant governments prevent corruption, reduce costs and increase competition in procurement.

In discussions relating to this study, one senior manager within the UK's Government Digital Service (GDS) emphasised to us the importance of the 'planning' and 'implementation' stages for anti-corruption, calling them the 'bookends' of public procurement.

The 'planning' stage (pre-procurement) is the first step in the OCDS guide to open contracting. During this stage budgets, project plans, procurements and public hearing data are released.¹⁶ Here, it is possible to isolate and mitigate the risk of bias towards particular suppliers, while also ensuring that cumbersome contract requirements do not close down the possibility of innovation.

Following the planning of the open procurement, the initial tender, award and final contract, the open procurement process terminates with the 'implementation' (post-procurement) stage. During implementation data about payments made, progress updates, and completion/termination is released. This is where it is possible for observers to check whether the

¹⁵ OCDS (2016a). Documentation. <http://standard.open-contracting.org/latest/en/> [accessed 10 May 2019].

¹⁶ OCDS (2016b). The Contracting Process. http://standard.open-contracting.org/latest/en/getting_started/contracting_process/ [accessed 10 May 2019].



project or purchase is properly meeting the requirements of the contract, and that money is being paid out on the right basis.

As our GDS contact told us, the OCDS allows contracting and delivery to be considered as a single process, with transparency stressed at every stage. The five-stage schema thereby helps officials embedded within the process to understand how they contribute to the overall goal of transparent government.

Ukraine

Ukraine's ProZorro e-procurement platform went live in February 2015, one year after the climax of the 2014 Maidan revolution, which saw the corrupt Viktor Yanukovich ousted from the country's presidency.¹⁷ Before its introduction, Ukraine's paper-based procurement system lost an estimated \$2 billion annually to corruption and other inefficiencies.¹⁸ In spring 2019, the Ukrainian Prime Minister Volodymyr Groysman claimed that since its launch, ProZorro had cut a total of \$2.35 billion from procurement.¹⁹

The system was initially conceived of by an informal group of activists and volunteers. It developed with the assistance of international NGOs including Transparency International and the Open Contracting Partnership, who helped ensure that ProZorro followed the OCDS.²⁰ All Ukrainian public procurement now runs through the platform, meaning that its scope is wide.

¹⁷ ProZorro was heavily influenced by procurement reforms in Georgia. For further information, see Nozadze, S. (2018). Improving public procurement in Georgia: what's the magic recipe? *The World Bank*.

<http://blogs.worldbank.org/europeandcentralasia/improving-public-procurement-georgia-what-s-magic-recipe> [accessed 25 April 2019].

¹⁸ Kyiv School of Economics (2017). Co-Creation of ProZorro. *Transparency International*.

https://www.transparency.org/whatwedo/publication/co_creation_of_prozorro_an_account_of_the_process_and_actors [accessed 18 April 2019].

¹⁹ *Ukrinform* (2019). Ukraine has saved UAH 63 billion in public procurement due to the ProZorro system. 17 April 2019.

<https://www.ukrinform.net/rubric-economy/2682860-ukraine-already-saved-uah-63-bln-due-to-prozorro-system-groysman.html> [accessed 18 April 2019].

²⁰ Kyiv School of Economics (2017). Co-Creation of ProZorro. *Transparency International*.



All data relating to successful suppliers, contract values, and tender notices and outcomes is accessible through an online console.²¹ Between the beginning of 2018 and mid-April 2019, 1.67 million tenders ran through the system, with a total value of \$50 billion, and with 17,000 bidders.²² About 80% of ProZorro's users are SMEs, a very high proportion compared to other countries.²³

ProZorro's impacts appear to support the rationale behind open procurement data. In a 2016 USAID survey of 300 Ukrainian entrepreneurs, 29% of respondents said that they had faced corruption within the ProZorro system as against 54% for the old paper-based system.²⁴ While still scoring relatively poorly, the country has risen by ten places since 2017 in Transparency International Corruption Perceptions Index, to 120th out of 180 countries in 2018. ProZorro was cited as a possible contributor to an increase in business confidence in this measure.²⁵ Finally, civil society groups have used the portal to find contracts that seem deliberately overvalued.²⁶

Big data, the European Union, and Digiwhist

Analysis of large and open datasets to identify possible instances of corruption in European Union tendering anticipates machine learning's role in anti-corruption. The EU-funded Digiwhist - 'Digital Whistleblower' project was a collaboration between six European universities and anti-corruption institutions.

²¹ OCP (2019b). Ukraine.

<https://www.open-contracting.org/why-open-contracting/showcase-projects/ukraine/> [accessed 18 April 2019].

²² ProZorro (2019). Public monitoring. <https://prozorro.gov.ua/en/monitoring> [accessed 18 April 2019].

²³ Granickas, K. (2018). Learning insights: The latest impacts emerging from Ukraine's ProZorro reforms. *Open Contracting Partnership*. <https://www.open-contracting.org/2018/01/12/learning-insights-latest-impacts-emerging-ukraines-prozorro-reforms/> [accessed 18 April 2019].

²⁴ Brown, S. (2016). 'Everyone sees everything': Overhauling Ukraine's corrupt contracting sector'. *Open Contracting Partnership*, 28 November 2016. <https://medium.com/open-contracting-stories/everyone-sees-everything-fa6df0d00335> [accessed 18 April 2019].

²⁵ (2019c). Ukraine in CPI 2018: Second Worst Score of All Neighbours After Russia. <https://ti-ukraine.org/en/news/cpi-2018-second-worst-score-of-all-neighbors-after-russia/> [accessed 18 April 2019]

²⁶ Brown, S. (2016). 'Everyone sees everything': Overhauling Ukraine's corrupt contracting sector'. *Open Contracting Partnership*.



Through large-scale data collection, cleaning and analysis, it sought to ‘simultaneously increase trust in governments and increase the efficiency of public spending across Europe’.²⁷

In January 2018, Digiwhist launched *opentender.eu*, a platform which allows users to easily examine details of at least 17.5 million EU tenders, dating back to 2003 and with a total value of above €27 billion.²⁸ Updated twice a year, the platform includes a tender ‘integrity’ filter function, so that tenders with comparatively high corruption risks can be identified.²⁹

Opentender’s central innovation is to use data crawling algorithms that systematically work through websites containing open tendering information (such as the EU’s Tenders Electronic Daily), download and structure the data included, and combine this with information taken from other sources about a company’s history and location, and its political connections.

Ultimately, this means that information on a single tender sourced from procurement portals is expanded to include whether the contract winner is, for example, registered in a tax haven or is a newly-registered company, potential red flags of a corrupt contract award. A tender’s ultimate ‘integrity’ score is the average of seven such indicators.³⁰ Research into the utility of ‘corruption risk indicators’ has been dominant within Digiwhist’s output since the project began.³¹

²⁷ Digiwhist (2019). About the project. <http://digiwhist.eu/about-digiwhist/> [accessed 21 April 2019].

²⁸ University of Cambridge (2018). 17.5 million of public tenders are being released on Opentender.eu today, covering 2003-2017 for selected countries. <https://www.sociology.cam.ac.uk/news/follow-the-money-european-digital-whistleblowers-no-longer-just-a-slogan-17-5-million-of-public-tenders-are-being-released-on-opentender-eu-today-covering-2003-2017-for-selected-countries> [accessed 21 April 2019].

²⁹ Opentender (2019). Integrity Indicators. <https://opentender.eu/all/dashboards/integrity> [accessed 21 April 2019]. Similar tools include Transparency International’s *redflags.eu*, which highlights risky Hungarian contract notices as they are published, and ARACHNE, which flags fraud risks in EU projects.

³⁰ Digiwhist (2017). D2.8 Methods Paper. <http://digiwhist.eu/wp-content/uploads/2018/03/D2.8-revised-version-FINAL.pdf> [accessed 22 April 2019].

³¹ Fazekas, M., Cingolani, L., Tóth, B. (2016a). A comprehensive review of objective corruption proxies in public procurement: risky actors, transactions, and vehicles of rent extraction. *Government Transparency Institute*.



The most efficient use of the Opentender portal would arguably be to halt projects and purchases linked to suspicious procurements, meaning that public funds do not leave the system. This is dependent on the sanctions and powers available to institutions responsible for procurement, major projects, or anti-corruption within individual EU countries.

That is, both Opentender and ProZorro demonstrate how big and open data can be translated into user-friendly corruption monitoring tools. To be fully effective, however, such tools must be embedded within a wider system of scrutiny, anti-corruption legislation and effective enforcement on the part of authorities. This way, it will be possible to establish the use cases that prove the utility of open data within anti-corruption. The same is true of any anti-corruption tool based on artificial intelligence.



Artificial intelligence: the next step

The theory

In the extent and usability of their data, big and open data portals like Opentender provide starting points for developing anti-corruption tools based on artificial intelligence. They do not rely on machine learning, however, in which a computer programme learns relationships between inputs and outputs without being instructed at every step of the way.³²

The Digiwhist project described above developed an extensive understanding of corruption ‘red flags’ in public procurement. ‘Red flags’ refer to states of affairs, or inputs, within the procurement process that are correlated with corruption, measured as ‘recurrently awarding contracts to a pre-selected company.’³³ Drawing on expert interviews and a literature review - that is, humans postulating input-output relationships - the Digiwhist team isolated 14 inputs as ‘significant and substantial predictors’ of corruption. These included: procurements with only one bidder, contracts being changed once a project has started, and the price of the tender documents.³⁴

The promise of machine learning lies in its ability to describe input-output relationships too complex for humans to detect. Theoretically, a programmer teaching an anti-corruption algorithm would begin by collecting and cleaning a wide range of data; in addition to the datasets mostly used in Opentender and Digiwhist project (open procurement data, with some company ownership and registration information), they could add further data on beneficial ownership and, potentially, data showing individual government transactions.

³² Maini, V. and Sabri, S. (2017). *Machine Learning for Humans*. https://www.dropbox.com/s/e38nil1dn17481q/machine_learning.pdf?dl=0 [accessed 24 April 2019].

³³ Fazekas, M., Tóth, I.J., King, L.P. (2016b). An Objective Corruption Risk Index Using Public Procurement Data. *European Journal of Criminal Policy Research*, 22: 369-397. <https://link.springer.com/article/10.1007/s10610-016-9308-z> [accessed 24 April 2019].

³⁴ *ibid.*



The programmer would then label outputs (contract awards, for example) as corrupt or non-corrupt. This is an important stage, requiring a clear and measurable definition of corruption.³⁵ With data and labelled outputs (i.e. training data) in place, the programmer can begin training a machine by directing it to follow a particular machine learning algorithm. The goal is for the machine to learn the complex function that determines the relationship between multiple states of affairs (inputs, for example, including payments made to a company registered offshore, or similar payments being made to individuals under unusual account codes) and outputs that satisfy the programmer's definition of corruption.

If the model was shown to operate to a high level of accuracy (i.e., it performs satisfactorily on test data), and presuming access to real-time data from the same sources as the training data, it could be put into use as a predictive or monitoring tool. That is, where conditions emerge that will likely result in corrupt outcomes according to the function learned by the machine, it would flag the transactions or contract in question for a human to investigate further.

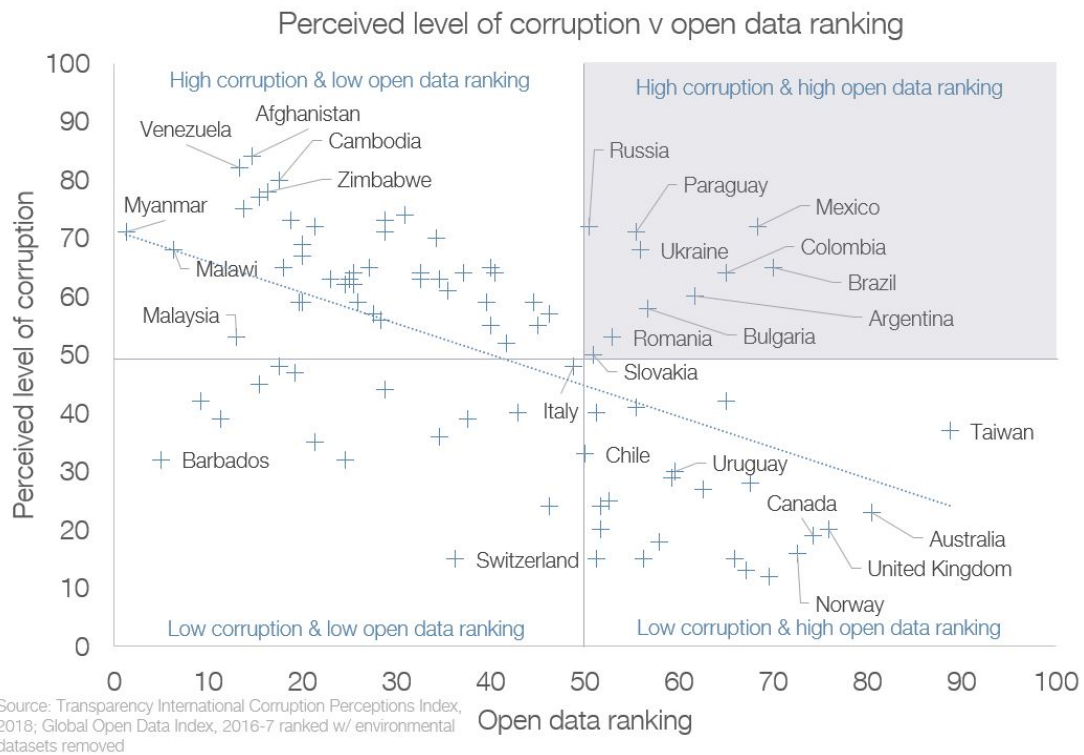
Implementation and benefits

The use case above describes a hypothetical or ideal scenario. There are, of course, many possible barriers to successful development and implementation of an anti-corruption machine learning model, including: the absence or non-availability of extensive and high-quality data; authorities' refusal to sanction such widespread monitoring; and, a poor enforcement environment.

³⁵ Hypothetically, this stage could see contracts or payments labelled as 'corrupt' if an actual prosecution had been linked to it. This, however, would assume that the set of corruption prosecutions is largely equivalent to the set of actually corrupt contracts, and would therefore depend on the efficacy of anti-corruption enforcement in the region or country in question. A method similar to that used by Digiwhist, where corruption is defined in a way that can be clearly measured within the data at hand is likely to be more helpful.



Chart 2: there are certain countries with high perceptions of corruption and high open data scores (highlighted in top right)



As this suggests, not every country will have the data available to make use of AI in this way. We might use the Global Open Data Index to indicate how much data is potentially available to AI developers working within anti-corruption.³⁶ This would suggest that countries in the developing world with high perception of corruption and a low open data ranking, such as Myanmar, Malawi and Zimbabwe, are unlikely to be ready to introduce AI into their anti-corruption efforts. However, there are ten countries, all of which are in Latin America or are ex-Communist countries, which have relatively high perceptions of corruption, but still have an apparently high volume of open data (see Chart 2).

These countries are: Argentina, Brazil, Bulgaria, Colombia, Mexico, Paraguay, Romania, Slovakia, Russia, and Ukraine. An in-depth and systematic assessment of the volume and quality of available data in these countries is

³⁶ For this research, the Global Open Data Index was adapted to exclude indices relating to environmental data, which are less relevant to measuring corruption.



required. Pending such research, however, we put forward this hypothesis: in these places, implementing machine learning within anti-corruption mechanisms will have the greatest effect in terms of capturing funds that risk leaving the fiscal system. Beyond the expanded analytical power that AI provides, the benefits would likely derive from the automated tool freeing up labour in what are often underfunded anti-corruption authorities for more detailed investigation and case-building. This, too, will require empirical verification.



Conclusion

Artificial intelligence is nascent within anti-corruption efforts, but early use cases are emerging.

Reflecting its position in Chart 2 as a high-corruption, high-open-data nation, Ukraine provides one that is especially promising. DoZorro AI was released in beta in November 2018. Mirroring the theoretical model outlined above, experts labelled 3,500 tenders as risky or non-risky to train the programme to detect possibly corrupt inputs. They continue to provide feedback to improve DoZorro's accuracy. The system does not rely on a stable set of risk indicators, meaning that it is able to follow the changing behaviours of corrupt officials. The early results were good, with many more risky tenders being identified than before.³⁷

Elsewhere, the World Bank is working with Microsoft to develop a machine learning tool that detects anomalies in procurement, combining tendering data with beneficial ownership information. The results are yet to be seen.³⁸ Another example comes from Spain, where researchers developed a neural network with the capacity to predict that corruption would emerge within certain regions three years ahead of the actual case. Their model suggested that sustained economic growth and 'an increase in real estate prices', among

³⁷ *Transparency International* (2018). DoZorro Artificial Intelligence To Find Violations In ProZorro: Here's How It Works. <https://ti-ukraine.org/en/news/dozorro-artificial-intelligence-to-find-violations-in-prozorro-how-it-works/> [accessed 25 April 2019]; Kucherenko, A. (2019). AI-watchdog for public procurement. *Medium*.

<https://medium.com/@andriykucherenko/ai-watchdog-for-public-procurement-730411ad66fe> [accessed 25 April 2019].

³⁸ Sharma, V. (2018). Can artificial intelligence stop corruption in its tracks? *The World Bank*. <http://blogs.worldbank.org/governance/can-artificial-intelligence-stop-corruption-its-tracks> [accessed 25 April 2019].



other things, predicted cases of corruption.³⁹ However, the neural network does not seem to be in current use within anti-corruption enforcement.

As these examples demonstrate, and building on substantial developments in both big and open data, using machine learning tools to combat corruption is not only viable, but highly likely to enter regular use in multiple countries in coming years. As our initial research suggests, it may have the largest effects in countries in Latin America and former Communist countries. These appear to have a good amount of available data and yet still have high levels of corruption.

Further research into data, artificial intelligence and anti-corruption should move beyond issues of what datasets and algorithms are most useful. The success of AI in fighting corruption will depend on solving problems that have long occupied government transparency advocates.⁴⁰ Who is empowered to introduce and sustain reform in government? How do you encourage civil servants to properly value the collection and release of relevant data? These are problems of authority, communication, relationships and power. Answering them will require a deft understanding of government and of human nature, but the rewards will be large.

³⁹ López-Iturriaga, F.J. and Sanz, I.P. (2017). Predicting Public Corruption with Neural Networks: an Analysis of Spanish Provinces. *Social Indicators Research*, forthcoming. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3075828 [accessed 25 April 2019].

⁴⁰ Darabi, A. (2017). Open data is stalling - new research says elections might be the answer. *Apolitical*. https://apolitical.co/solution_article/reforming-fire-secure-open-data-elections-loom/ [accessed 25 April 2019].