# Linked Democracy 3.0 -  Global Machine Translated Legislation and Compliance in the Age of Artificial Intelligence

Dr. Sean Goltz[1]

University of Waikato, Hamilton, New Zealand

sean.goltz@waikato.ac.nz

## Abstract

This paper outlines the efforts made by Global-Regulation, a world legislation search engine, to engage artificial intelligence in two ways: (i) employing machine translation to translate the world's legislation to English and, (ii) creating an automated system to identify compliance clauses and extract penalties from legislation. This paper describes Global-Regulation's vision and technology in the context of linked democracy and the democratization of artificial intelligence.

## Keywords

Legislation, Machine Translation, Compliance, Penalties, Linked Democracy, Democratization AI

## 1      Introduction

There is a strong relationship between democracy and transparency.[1] At the same time, some argue that big data will enable citizens to be governed by a data-empowered "wise king", who would be able to produce desired economic and social outcomes almost as if with a digital magic wand. [2] These trends bring to the front a term recently used by Microsoft's CEO Satya Nadella, of democratizing AI. By making AI available to everyone, it can move from a centralized tool, to one which can be used in fields such as healthcare, education, manufacturing, retail and more. The ultimate aim, is sharing AI's power with the masses, allowing anyone and everyone to use the AI systems they need.[3]

This paper outlines the efforts made by Global-Regulation (www.global-regulation.com), an online search engine of the world's legislation, to engage artificial intelligence in two ways: (i) employing machine translation to translate the world's legislation to English and, (ii) creating an automated system to identify compliance clauses and extract penalties from this legislation. These means are intended to foster democracy and improve regulation by enabling lessons drawing from one jurisdiction to the other.[4] As stated by Lloyd: "The internet collapses geography and expands our concept of community, yet geographic community is a cornerstone of our structures for democratic participation".[5]

---

[1] The author is a co-founder of Global-Regulation

The paper will start with a background section underpinning the problem that Global-Regulation came to solve, how the project was born, preliminary steps that were taken and the underlying motivation that drove this project's co-founders to build it. The second section deals with the method in which artificial intelligence was used in engaging machine translation on a massive scale and in the creation of the PenaltyAI system, designed to identify compliance clauses and extract penalties from the legislation. Finally, I will conclude with what have been learned and what can and should be done next.

## 2       Background

According to Monson, "Government services can and should be delivered as efficiently and effectively as the technology you use to get a ride or order dinner".[6] In a nutshell, Monson is capturing both the problem and the underlying motivation behind Global-Regulation. Before meeting by chance, both co-founders were running novel legal websites. Addison Cameron-Huff was running a website that tracks and provide alerts on new provincial legislation and the author was running a website that provides summaries of case studies in regulation. Joining forces enabled the co-founders to embark on this ambitious project creating a search engine of the world's legislation. The co-founders did not realized at the time the scale and magnitude of this project nor the challenges lying ahead.

Very quickly it dawned on the co-founders that when one is offering a source that was never available before (e.g., Canadian academics can now read Italian legislation in English) one need to convince users that it is valuable. More than once we heard American or Australian regulators pondering why should they look at regulations from Denmark, for example? 'We are looking only at comparable jurisdictions', they told us. 'What lessons can we possibly draw from remote parts of the world?!'. It was academics, that immediately realized the potential and started exploring the database with enthusiasm. And surprisingly enough, it was the tech giants Microsoft, Google and Amazon, that generously supported our vision.

Global-Regulation is now the largest search engine of legislation from around the world, enabling comparative search of 1.6 million laws and regulations from 88 countries. Global-Regulation has employed Microsoft and Google's machine translation on a massive scale translating 750,000 laws and regulations from 26 languages into English. By providing this information, Global-Regulation unlocks the global village vision in law by automating database translation. To support its vision, it employs cloud based technology powered by Amazon to gather, index and standardize legislation from different countries across the globe. The challenge Global-Regulation faced was twofold: how to deal with laws in different languages that are coming from different legal systems.

## 3       Machine Translation

Initially, Global-Regulation connected to each country's official government website and uploaded the legislation to its database. This process did not enable the inclusion of legislation in foreign languages. Dealing with this challenge has bearing not only on commercial aspects, but also on Global-Regulation's founders' vision, to have all the world's legislation, searchable, in English, in one place.

The importance of this vision cannot be overstated, mainly for developing economies with unique regulatory structure interested in drawing in external investment on the one hand, and making their legal system transparent to its citizens, on the other. Making legislation transparent, accessible and searchable, especially on a comparative basis, is one of the cornerstones of democracy and a task made possible on this scale only due to recently mature technology, advances in artificial intelligence, and governments making laws available online.

The process of machine translation for laws is as follows:

1. Index the laws in the original language and track which language the law is in (in some countries laws are published in several languages)
2. Download the laws in the original language.
3. Convert laws to "plaintext" (from HTML, XML, PDFs, etc.), where plaintext means UTF-8-encoded plain text files.
4. Format the plaintext so that items like headers, footers, and extra non-legal information is removed. Attempt to normalize line endings (especially important for PDF conversions which have odd formatting issues).
5. Break the plaintext into pieces that can be handled by machine translation systems (which generally have a size limit) using logical break points such as line endings. Also translate the title of the law.
6. Convert each piece into English then stitch the English version together using the breakpoint identified in the previous step.
7. Store the translated law and the original law in the database.
8. As machine translation models for languages improve, periodically re-translate the laws and store them in the database.
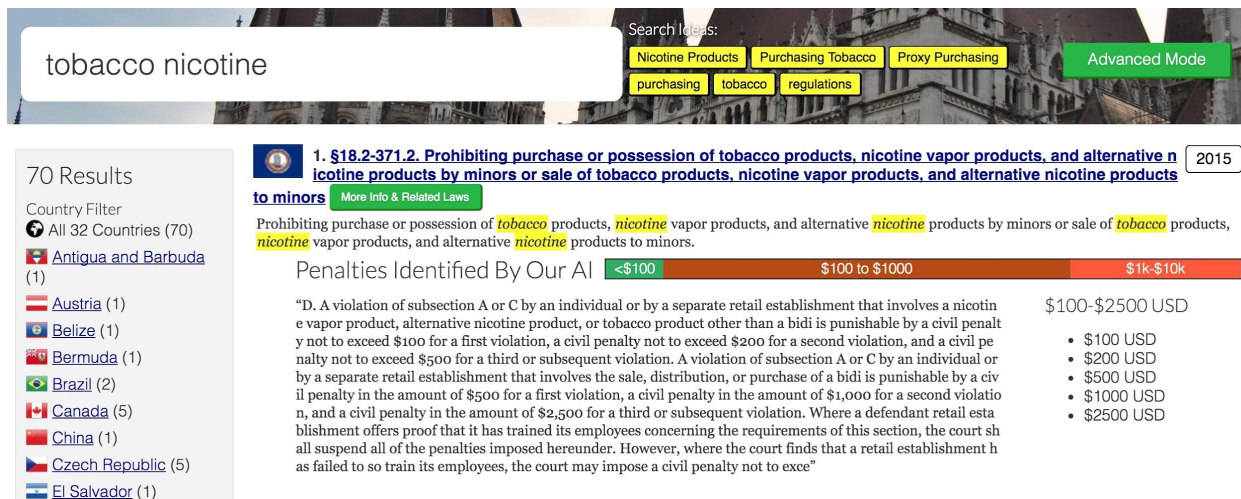
## 4 PenaltyAI

Following the use of machine translation, Global-Regulation have taken a step further in order to use its huge database of world laws along with the advanced capabilities of artificial intelligence. This step involved the development of a system (called 'PenaltyAI') to identify compliance clauses in legislation and extract the actual penalties from these clauses, and convert it to US dollars (if needed). This ambition to create the ultimate risk and compliance system came into life when Global-Regulation's founders realized that penalties are the kind of information that can be identified with a high degree of certainty by an artificially intelligent system.[7]

After seemingly endless testing, experimenting, coding, consulting[2] and hard work, Global-Regulation presented its PenaltyAI Search (https://www.global-regulation.com/penalty_ai.php) – the first and only AI system that identify compliance clauses in legislation on a global scale, extracts the actual penalties amount and serve it all to the user in US dollars. See two examples in Figure 1 and Figure 2.

---

[2] thank you to Kyle Gorman from Google for the words to numbers converter recommendation) https://research.google.com/pubs/KyleGorman.html

**Fig. 1.** PenaltyAI search results for 'tobacco nicotine' (Source: www.Global-Regulation.com)

This as a great achievement on several levels:

1. an AI system that can read legal text and produce useful meaning; and,
2. enabling risk and compliance professionals to explore real and relevant data on a global scale, in English; and,
3. allowing governments and businesses to assess and enhance their compliance efforts; and,
4. for researchers to compare and contrast risk and compliance data globally; and finally and perhaps most importantly,
5. A first step in enabling the public to have a transparent and informed access to regulatory compliance hence
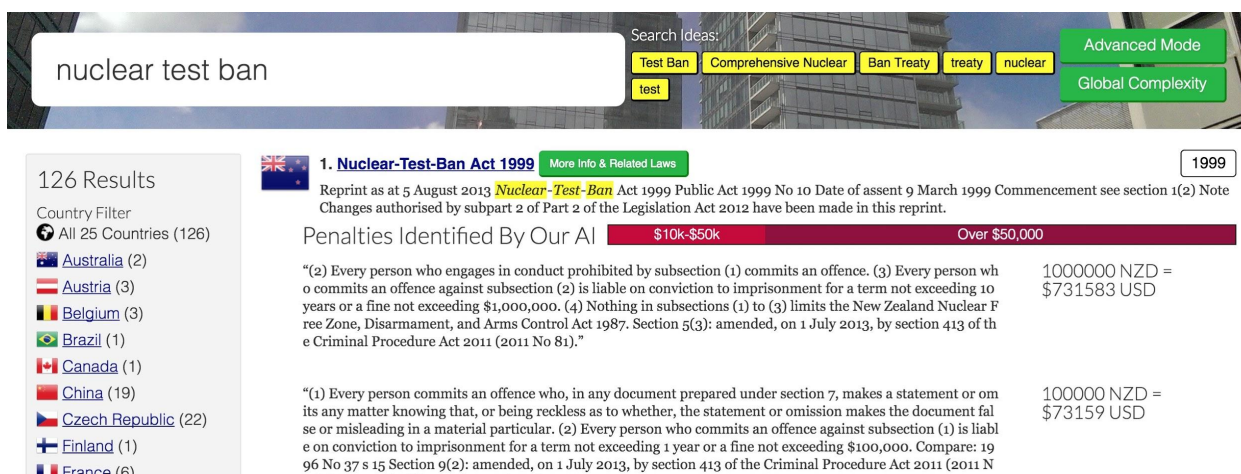


an enhancement of democracy.

**Fig. 2.** PenaltyAI search results for 'nuclear test ban' (Source: www.Global-Regulation.com)

The "PenaltyAI Search" can answer questions like "What would I pay for violating money laundering laws in Jamaica?" or "How much would a smuggler who warehouses stolen goods in China pay if they're caught?".

The penalties are extracted by an offline algorithm that runs on an Azure Virtual Machine that does the following steps:

1. Find laws that mention keywords associated with civil penalties (as a first pass);
2. Convert all word numbers (like "one million") into international number format ("1,000,000.00");
3. Identify the paragraphs that likely contain civil penalties based on words and numbers;
4. Merge several penalties into one, whether they related to the same "clause" (section) of a law;
5. Extract all the clauses and penalties;
6. Exclude certain classes of text that are almost never penalties but look like penalties (such as laws about gold coins and section references in laws that have to do with money);
7. Recognize currencies in text, and combine this data with our table of national currencies, and convert penalties into USD using Yahoo! Finance rates (through the XML API call);

| | ID | law_ID | excerpt_md5 md5 hash of excerpt. Note: could be several w/ sam... | excerpt | inserted_ts |
|---|---|---|---|---|---|
| ⋮ Copy ⊖ Delete | 98961 | 2894980 | 1aae5b55b8406794c5b442d0614daa1a | § 7 the noncompliance of the preceding paragraph s... | 2017-02-06 18:16:06 |
| ⋮ Copy ⊖ Delete | 98962 | 2894980 | f12ca13cab754c0dddabbfd3cd0a5101 | Sole paragraph. The deduction referred to in the c... | 2017-02-06 18:16:06 |
| ⋮ Copy ⊖ Delete | 98963 | 2895138 | 1aae5b55b8406794c5b442d0614daa1a | § 7 the noncompliance of the preceding paragraph s... | 2017-02-06 18:16:06 |
| ⋮ Copy ⊖ Delete | 98964 | 2895138 | f12ca13cab754c0dddabbfd3cd0a5101 | Sole paragraph. The deduction referred to in the c... | 2017-02-06 18:16:06 |
| ⋮ Copy ⊖ Delete | 98965 | 2895175 | f44a461e472793f004074fa42659e828 | CHAPTER II of INFRACTIONS AGAINST the GENETIC HERI... | 2017-02-06 18:16:06 |
| ⋮ Copy ⊖ Delete | 98966 | 2895175 | 430f88843fd8c4930965f98528dc02b3 | Art.   16. Access component of genetic heritage fo... | 2017-02-06 18:16:06 |
| ⋮ Copy ⊖ Delete | 98967 | 2895175 | 924b4e25a50641213fee865279946714 | Art.   17. Refer to the outer component sample of ... | 2017-02-06 18:16:06 |

8. Store the penalties and clauses in a MySQL database (RDS) - see Figure 3.

**Fig. 3.** Screenshot of one of the MySQL tables for penalties (Source: www.Global-Regulation.com)

After following the said steps, the system then note in its search instance whether or not a law has penalties attached to it, so that the search instance can filter by laws that have penalties (as opposed to Global-Regulation's regular search that includes laws that don't have explicit fines attached to them). This process is run as a batch job offline because 1.6 million laws takes several hours to process.

When a user does a search, the search is first sent to Global-Regulation's Elasticsearch instance, and then the penalties are looked up from the MySQL database afterwards. This allows full-text search of laws to be combined with penalties, and in a way that results in much less strain on Global-Regulation's relational database (because penalties are looked up by IDs rather than a JOIN). Storing the penalties separately allows to reduce the amount of data in the in-memory search instance, and decouples the services (since Global-Regulation have other types of search like technical standards and law analytics). See Figure 4 for the overall global penalties summary.
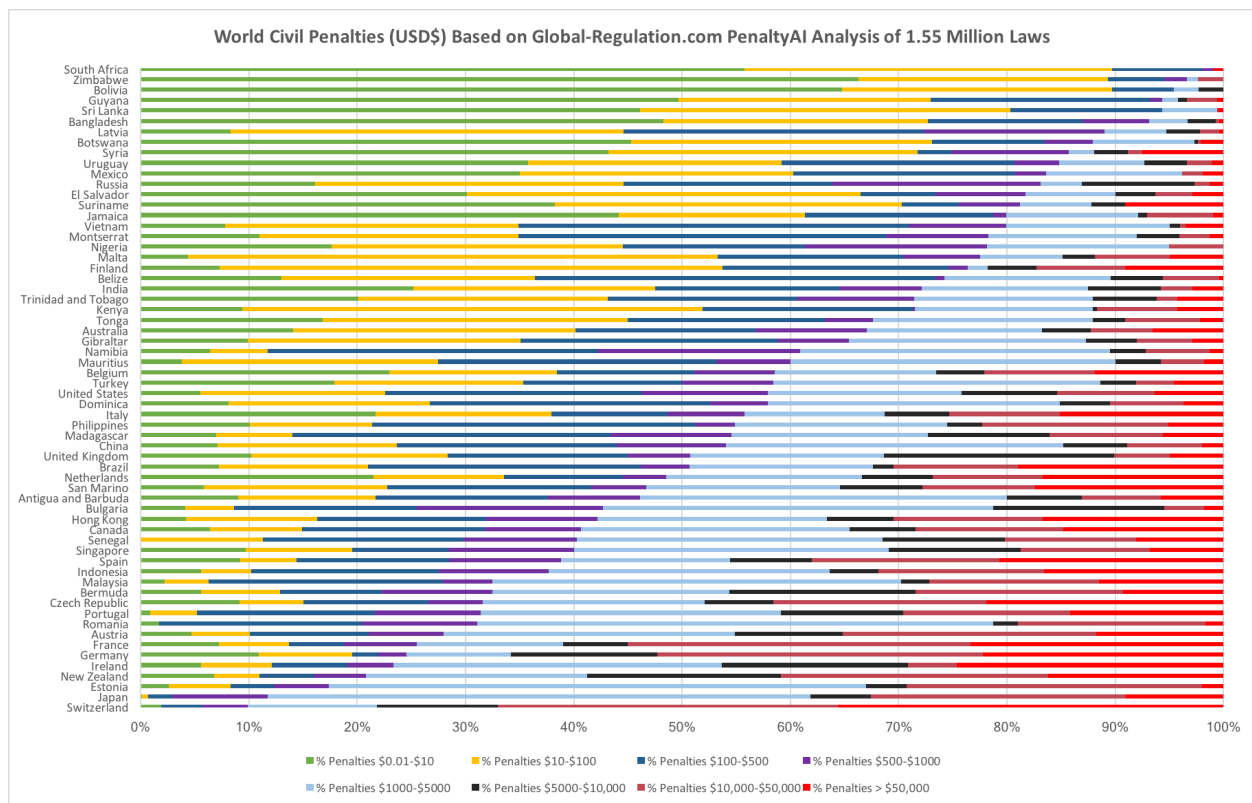
**Fig. 4.** Visualization of penalties for non-compliance (Source: www.Global-Regulation.com)

## 5 Conclusion

Global-Regulation enhances linked democracy by using state of the art artificial intelligence technology to provide the world's legislation in English. Furthermore, Global-Regulation provides an advance system to extract compliance clauses and penalties from this legislation. Furthermore, Global-Regulation builds on the democratization of artificial intelligence and Microsoft's generous support to bring the benefits of technology in general and AI in specific, for the advantage of the public.

What could only be a dream few years ago, has turned with Global-Regulation's vision and advanced technology into an innovative tool of global democracy, now used by leading academic institutions and governments around the world. Going forward we look to expand this democratising tool in a way that will enable every citizen of the world to ask for her legal rights and responsibilities in her country of choice, based on this database of world laws and receive a reply, completely automated, with a click of a mouse.

We have learned that combining big data - the world laws, with advanced artificial intelligence, creates fertile ground for endless opportunities in the realm of bringing the law to the people and bridging the gap between society's bricks (laws) and its citizens. Yet the main challenge going ahead is twofold: how to use the technology in a way that will be understandable, intuitive and friendly to people; and, perhaps more importantly, how to explain to users around the world, that Global-Regulation is a key to linked democracy.

## Bibliography

[1] Molly Schwartz, Democracy and open data: are the two linked?, May 22, 2014, Congressional Data Coalition, http://congressionaldata.org/democracy-and-open-data-are-the-two-linked/

[2] Dirk Helbing, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V. Zicari, Andrej Zwitter, Will Democracy Survive Big Data and Artificial Intelligence?, Scientific American, February 25, 2017, https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/

[3] Democratizing AI: Satya Nadella on AI vision and societal impact at DLD, 17/01/2017, https://news.microsoft.com/europe/2017/01/17/democratizing-ai-satya-nadella-shares-vision-at-dld/#sm.00001d5vpu2vg5dm8tucb72g11vbl#iFfMr1YvJHXYZtg1.97

[4] Sean Goltz & A. Nikolic, Global-Regulation – Drawing Future Regulatory Tools from the Experience of the Past, 4(3) The European Journal of Risk Regulation 391-398 (2013), http://ejrr.lexxion.eu/list/articles/author/Nikolic,%20Aleksandar

[5] Alexis Lloyd, Disentangling Democracy From Geography, The Atlantic, May 9, 2017, https://www.theatlantic.com/technology/archive/2017/05/disentangling-democracy-from-geography/524124/

[6] Rebekah Monson, Freeing Technology From the Pace of Bureaucracy, The Atlantic, May 16, 2017, https://www.theatlantic.com/technology/archive/2017/05/freeing-technology-from-the-pace-of-bureaucracy/524034/

[7] Sean Goltz & M. Mayo, Enhancing Regulatory Compliance by Using Artificial Intelligence Text Mining to Identify Penalty Clauses in Legislation, MIREL 2017 - Workshop on `MIning and REasoning with Legal texts', held at the 16th International Conference on Artificial Intelligence and Law, King's College, London, UK, June 12 - 16, 2017 [forthcoming]