

Back to the future

Citation for published version (APA):

Goanta, C., van Dijck, G., & Spanakis, G. (2020). Back to the future: Waves of legal scholarship on artificial intelligence. In S. Ranchordas, & Y. Roznai (Eds.), *Time, Law, and Change: An Interdisciplinary Study* Hart Publishing.

Document status and date:

Published: 16/04/2020

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Back to the Future: Waves of Legal Scholarship on Artificial Intelligence

Catalina Goanta, Gijs van Dijck, Gerasimos Spanakis*

Forthcoming in Sofia Ranchordás and Yaniv Roznai, *Time, Law and Change* (Oxford, Hart Publishing, 2019)

Introduction

In the past years, artificial intelligence has received increased attention in legal scholarship. Topics such as self-driving cars, predictive policing or discriminatory profiling are only a few examples of trending legal scholarship, mirroring similar academic developments in other scientific disciplines – especially in computer science research. Interestingly, and perhaps not surprisingly, several of the same research questions asked in relation to artificial intelligence have been posed in earlier research decades ago, albeit with more emphasis on normativity. As legal research does not often make use of self-standing literature reviews, and given the ever-growing body of legal literature available on artificial intelligence, it is getting more and more difficult to have an overview of existing work on the subject, which may lead to repetitive research.

This chapter presents the first part of a project aimed at making such a research overview. It does so by using a dataset of 3931 academic journal articles obtained from HeinOnline referencing the topic of artificial intelligence (hereinafter Corpus). HeinOnline, one of the largest legal databases in the world, indexes sufficient publications to shape a reliable knowledge map of existing research that refers to artificial intelligence: this leaves the possibility for additional observations to be added from other sources (e.g. other databases, other publishers, etc.). In the first section, we describe the Corpus, including the methodology used in obtaining it and the characteristics of the publications therein. The second section comprises visualisations of the Corpus using descriptive statistics, as well as examples of thematic clusters, which are subsequently discussed. Additionally, it briefly explains the statistical model (Latent Dirichlet Allocation Topic Modelling) to be deployed in the second part of the project. The third section reflects upon the meaning and future of legal research on artificial intelligence, and is followed by the conclusion of the chapter.

Doing research on research is a remarkable opportunity for legal scholars to zoom out of a given field of law and observe it not only spatially but also temporally. Time plays an important and often overlooked role, especially in research that focuses on the socio-legal implications of complex technologies. By placing legal debates in a broader timeline, it is possible to identify patterns (allowing for a faster understanding of the socio-legal phenomena central to legal research on artificial intelligence), and to provide a checklist for questions both that have and have not yet been answered.

I. Time and Technology: The Rise of Artificial Intelligence as a Topic of Legal Inquiry

That nowadays more and more attention is focused on digitalisation is well-known. Even national research policies include, at the moment, a preference for themes such as Digital Legal Studies.¹

* Catalina Goanta is Assistant Professor in Private Law at the Faculty of Law, Maastricht University. Gijs van Dijck is Professor of Private Law at the Faculty of Law, Maastricht University. Gerasimos Spanakis is Assistant Professor of Computer Science at the Department of Data Science and Knowledge Engineering, Maastricht University.

¹ For instance, in the Netherlands, ‘Digital Legal Studies’ is one of the themes law faculties can choose to focus on in order to benefit from an extra investment in the social sciences and humanities made by the Ministry of Education, Culture and Science. See Netherlands Organisation for Scientific Research, ‘Samen sterker. Beeld van het SSH Domein’ (*Sectorplan SSH*, November 2018) <<http://www.sectorplan-ssh.nl/artikelen/documenten>> accessed 9 June 2019.

Inspired by the diversification of societal phenomena resulting from technological innovation, the past decade has seen a newfound interest in the emerging field of law and technology. Legal research – both doctrinal and multidisciplinary – on topics such as reputation systems,² fake news and democracy,³ content moderation,⁴ platform governance,⁵ algorithmic gatekeeping,⁶ or distributed ledger technologies,⁷ has been on the rise. One of the topics enjoying increased popularity, partly due to its media hype,⁸ is artificial intelligence. Both praised and feared, the technological progress allowing machines to mimic cognitive functions that are inherently human (e.g. learning) has a much richer history than popular analyses often acknowledge, and is developing at an astounding pace.⁹

In the remainder of this section, we will first outline a brief history of artificial intelligence as a field of computer science research. Legal audiences – especially those interested in artificial intelligence – can benefit from such an expose, which is important but frequently missing in legal scholarship.¹⁰ Subsequently, this information will be overlapped with the Corpus of legal literature, consisting of 3931 academic journal articles obtained from HeinOnline, which will be explored using descriptive statistics.

A. A Brief History of Artificial Intelligence

In the words of Bruce Buchanan, a computer scientist from the University of Pittsburgh, ‘the history of AI is a history of fantasies, possibilities, demonstrations and promise’.¹¹ The fantasy part of this history has left its traces in literary fiction.¹² The scientific path – which led to the consolidation of artificial intelligence as a field of research – lies in the history of computer science.

In 1955, mathematicians John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon initiated the Dartmouth summer research project on artificial intelligence for the following

² Sofia Ranchordás, ‘Public Values, Private Regulators: Between Regulation and Reputation in the Sharing Economy’ (2019) 13 *Law & Ethics of Human Rights* (forthcoming).

³ Alberto Allemano, ‘How to Counter Fake News? A Taxonomy of Anti-fake News Approaches’ (2018) 9(1) *Eur J Risk Reg* 1. See also Siva Vaidhyanathan, *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy* (Oxford, Oxford University Press 2018).

⁴ James Grimmelman, ‘The Virtues of Moderation’ (2015) 17 *Yale JL & Tech* 42.

⁵ Jack M Balkin, ‘Information Fiduciaries and the First Amendment’ (2016) 49(4) *UCD L Rev* 1183.

⁶ Mireille Hildebrandt, ‘Algorithmic regulation and the rule of law’ (2018) 376 *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 1.

⁷ Primavera de Filippi and Benjamin Loveluck, ‘The invisible politics of Bitcoin: governance crisis of a decentralised infrastructure’ (2016) 5(3) *Internet Policy Review*.

⁸ See for instance Martin Giles, ‘Ian Goodfellow GANfather’ *MIT Technology Review* (Cambridge, Mass, 21 February 2018) <<https://www.technologyreview.com/s/610253/the-ganfater-the-man-whos-given-machines-the-gift-of-imagination/>>;

Hugh Howey, ‘How to Build a Self-Conscious Machine’ *Wired* (San Francisco, 4 October 2017)

<<https://www.wired.com/story/how-to-build-a-self-conscious-ai-machine/>>; Gideon Lewis-Kraus, ‘The Great A.I.

Awakening’ *The New York Times* (New York, 14 December 2016) <<https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html>>;

Richard Socher, ‘AI’s Next Great Challenge: Understanding the Nuances of Language’ *Harvard Business Review* (Brighton, Mass, 25 July 2018) <<https://hbr.org/2018/07/ais-next-great-challenge-understanding-the-nuances-of-language>> all sources accessed 9 June 2019.

⁹ Vishal Maini, ‘Machine Learning for Humans 🤖’ (*Medium*, 19 August 2017) <<https://bit.ly/2wrDJTO>> accessed 9 June 2019.

¹⁰ In the very few papers where it is included, the history of artificial intelligence mostly refers to the specific subfield of legal informatics, as an expression of artificial intelligence applied to issues or questions mostly relevant for legal practice (e.g. e-discovery, etc.). See Sergio David Becerra, ‘The Rise of Artificial Intelligence in the Legal Field: Where We Are and Where We Are Going’ (2018) 11 *J Bus Entrepreneurship & L* 27; Richard E Susskind, ‘Artificial Intelligence, Expert Systems and Law’ (1990) 5 *Denning LJ* 105.

¹¹ Bruce G Buchanan, ‘A (Very) Brief History of Artificial Intelligence’ (2005) 26 *AI Magazine* 53.

¹² See for instance Frank Baum, *Tiktok of Oz* (Chicago, Reilly & Britton 1914).

year, based on the assumption that ‘[...] every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.’¹³ While earlier notable work on the idea of machines behaving intelligently had been theoretically designed and published,¹⁴ the 1956 Dartmouth conference is generally considered to be the moment when artificial intelligence was born as a concept and was shaped as a mission.¹⁵

The following decade was so full of optimism due to breakthroughs in natural language processing and machine vision:¹⁶ researchers started speculating it would only be a matter of time until computers would be able to beat humans in chess, or discover and prove new mathematical theorems.¹⁷ The optimism fuelled and was fuelled by an injection of funding into research on artificial intelligence. Between 1963 and approximately 1970, the Massachusetts Institute of Technology was heavily funded by the Advanced Research Projects Agency (later known as DARPA) with up to \$3 million per year.¹⁸ However, problems such as the lack of data and computational power, as well as design issues in the application of reasoning or logic systems, soon began to slow down the initial enthusiastic efforts of creating human-like intelligence in machines. This led to what is known as the first AI winter (see Figure 1), and it manifested itself through the cutting of financial support to artificial intelligence research projects and the general loss of interest in the topic.

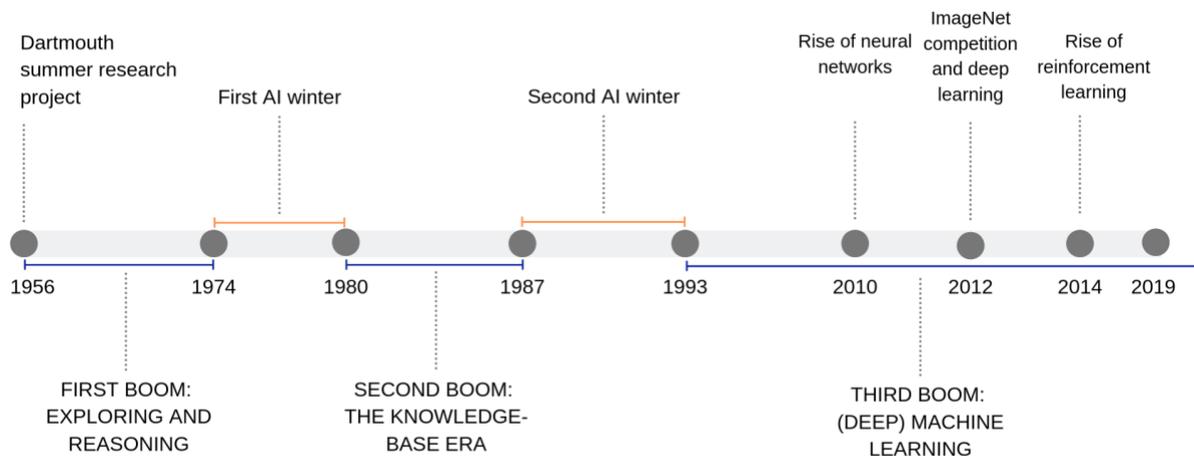


Figure 1 – A brief history of artificial intelligence¹⁹

¹³ John McCarthy and others, ‘A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence’ (31 August 1955) <<http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>> accessed 9 June 2019.

¹⁴ Alan Turing, ‘Computing Machinery and Intelligence’ (1950) 59 *Mind* 433; Claude E Shannon, ‘Programming a Digital Computer for Playing Chess’ (1950) 41 *Philosophy Magazine* 356.

¹⁵ Daniel Crevier, *AI: The Tumultuous Search for Artificial Intelligence* (New York, Basic Books, 1993) 49; Pamela McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospect of Artificial Intelligence* (25th anniversary edition) (Natick, A. K. Peters Ltd. 2004), 114.

¹⁶ Crevier, supra note 15, 84–102; 164–172; McCorduck, supra note 15, 300–305. See also Marvin L Minsky, *Semantic Information Processing* (Cambridge, MIT Press 1968); Marvin L Minsky, ‘Steps Toward Artificial Intelligence’ (1961) 49 *Proceedings of the Institute of Radio Engineers* 8.

¹⁷ Stuart J Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Upper Saddle River, Prentice Hall 2003); see also Arthur L Samuel, ‘Some Studies in Machine Learning Using the Game of Checkers’ (1959) 3 *IBM Journal of Research and Development* 210.

¹⁸ Crevier, supra note 15, 64–65.

¹⁹ Original scheme based on Crevier, supra note 15; McCorduck, supra note 15; Tokio Matsuzaki, ‘Ethical Issues of Artificial Intelligence in Medicine’ (2018) 55(1) *Cal W L Rev* 255; and Karen Hao, ‘We analyzed 16,625 papers to figure out where AI is headed next’ *MIT Technology Review* (Cambridge, Mass, 25 January 2019) <<https://bit.ly/2Tlj3oM>> accessed 9 June 2019.

In addition to the lack of funding, researchers from other disciplines (e.g. philosophy) started voicing concerns regarding the goals and visions of artificial intelligence projects. One was philosopher Hubert Dreyfus, who was already emphasizing the threats of reliance on machines for reasoning in his ‘Alchemy and AI’ paper dating from 1965: ‘if, using only digital programs, we try to process the ill-structured data in which real-life problems are posed, we face an infinite regress.’²⁰

In spite of growing criticism, the early 80s saw renewed funding opportunities, giving researchers the possibility to approach artificial intelligence in a new manner.²¹ There was a shift away from large, abstract concepts and targeting knowledge base approaches by zooming into specific domains – leading to the rise of expert systems.²² Once again penning too ambitious goals²³ and also finding itself under the pressure of external factors,²⁴ the research community could not consistently hold the attention of public and private investors since it seemingly failed to achieve commercial impact. The end of the 80s saw a new AI winter that required not only redesigning methods of research in artificial intelligence, but even the terminology used to describe the discipline as a whole (e.g. knowledge engineering, informatics, etc).

As of the early 1990s, the third boom of artificial intelligence research saw more media popularity than ever before, with new breakthroughs such as IBM’s Deep Blue defeating the world’s most famous chess player, Gary Kasparov, in a six-game match in 1997.²⁵ Between 1993 and approximately 2010, supervised learning in the form of empirical systems with algorithms such as support vector machines began to be used. The promising rise of deep learning as of 2012 enjoyed a similar degree of popularity.²⁶ In 2019, deep learning is the driving force behind Google searchers, Facebook news feeds and Netflix recommender systems, although it is predicted it may soon be replaced by other approaches.²⁷ In a study of all 16625 papers available in the ‘artificial intelligence’ section of the database arXiv for the Massachusetts Institute of Technology (MIT) Technology Review,²⁸ Karen Hao highlighted that interest in artificial intelligence is growing (see Figure 2 below). Three main trends

²⁰ Hubert L Dreyfus, ‘Alchemy and AI’ (1965) RAND Corporation Memo

<<https://www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf>> accessed 9 June 2019, 70. See also the expanded version of the paper, later published as a book, Hubert Dreyfus, *What Computers Can't Do* (New York, MIT Press 1972).

²¹ In 1981, the Japanese Ministry of International Trade and Industry launched a 10-year computer technology project focusing on natural language and knowledge-based information processing, totaling \$850 million. See Mikel Olazaran, ‘History of the Neural Network Controversy’ (1993) 37 *Advances in Computers* 335, 387.

²² McCorduck, *supra* note 15, 421. See also P. Jackson, *Introduction To Expert Systems* (Boston, Addison-Wesley Longman Publishing 1986), 2-10.

²³ Crevier, *supra* note 15, 212; McCorduck, *supra* note 15, 441.

²⁴ Progress in artificial intelligence research has always been dependent on the hardware used for development. The late 80s saw a revolution in personal computing that affected the development of computers specialised to support artificial intelligence. See McCorduck, *supra* note 15, 435.

²⁵ Steven Levy, ‘What Deep Blue Tells Us about AI in 2017’ *Wired* (San Francisco, 23 May 2017) <<https://www.wired.com/2017/05/what-deep-blue-tells-us-about-ai-in-2017/>> accessed 9 June 2019.

²⁶ A turning point in the use of deep learning is marked by the 2012 edition of the ImageNet Competition, where three researchers from the University of Toronto submitted a deep convolutional neural network architecture called AlexNet, which was the first to get a less than 25% error rate for the visual recognition challenge. Using the same method in the subsequent five years, competing teams managed to further bring down the error rate to less than 5%. See Dave Gershgorn, ‘The data that transformed AI research—and possibly the world’ *Quartz* (New York, 26 July 2017) <<https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/>> accessed 9 June 2019. See also the original paper, Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton, ‘ImageNet Classification with Deep Convolutional Neural Networks’ (2012) NIPS.

²⁷ Hao, *supra* note 19. See also Nils Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (New York, Cambridge University Press 2009).

²⁸ arXiv is an open access database owned and operated by Cornell University, currently storing 1545774 electronic papers on physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics, available at <<https://arxiv.org>>.

describe research on this topic since the early 2000s:²⁹ a paradigm shift from knowledge base systems toward machine learning in the early 2000s, the rise of neural networks in the early 2010s, and the popularity of reinforcement learning since 2014.³⁰ The trend of moving away from empirical systems which require huge amounts of labelled data, and towards more unsupervised learning raises additional questions that are very relevant for law, such as how to understand the liability of a system that is not controlled by its designers. This has led to the emergence of a new research field on Fairness, Accountability, and Transparency (FAT) in socio-technical systems.³¹

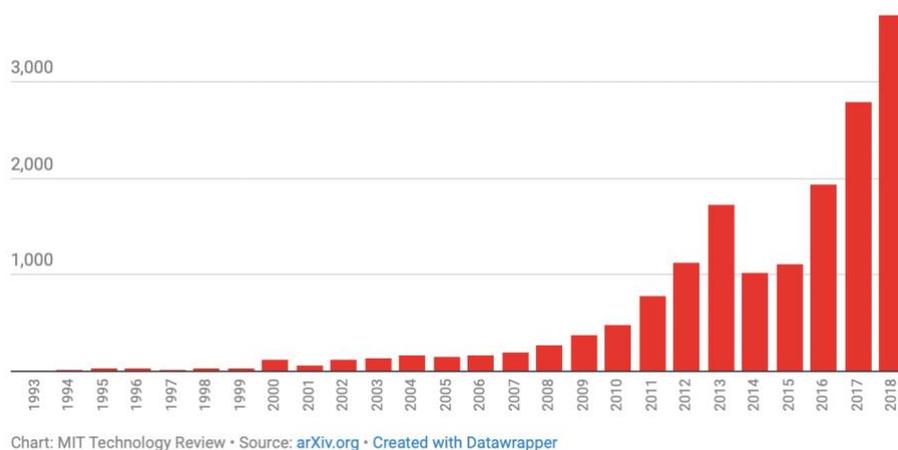


Figure 2 – arXiv papers in the section ‘artificial intelligence’³²

It is difficult to assess whether the current boom focused on machine learning will at some point result in disillusionment and consequently yet another AI winter. Well-known deep learning experts such as Andrew Ng claim artificial intelligence will benefit from stabilised financing in the years to come, given it has such a ‘strong underlying driver of real value that it won’t crash like it did in previous years.’³³

B. Legal Literature on Artificial Intelligence: The HeinOnline Corpus

Legal interest in artificial intelligence can be traced as early as late 1950s and early 1960s, when journals such as *MULL: Modern Uses of Logic in Law*³⁴ were publishing reports on conferences akin to the Law and Electronics Conference. Participants to the Conference were discussing how ‘data processing could effect accurate and rapid handling of vast amounts of data, improve legal research through more ready accessibility to repositories of the law, reveal gaps and contradictions in existing law, and could lead to redesigning the legal system and provide the opportunity for eventual reform.’³⁵

²⁹ Hao, supra note 19. The research takes into account papers indexed on arXiv on 18 November 2018.

³⁰ Ibid.

³¹ Solon Barocas and Moritz Hardt and Arvind Narayanan, *Fairness and Machine Learning: Limitations and Opportunities* (2018) <<https://fairmlbook.org/index.html>> accessed 9 June 2019.

³² Figure from Hao, supra note 19.

³³ Will Knight, ‘AI Winter Isn’t Coming’ *MIT Technology Review* (Cambridge, Mass, 7 December 2016) <<https://www.technologyreview.com/s/603062/ai-winter-isnt-coming/>> accessed 9 June 2019.

³⁴ The *MULL: Modern Uses of Logic in Law* journal, dating from 1959, became known as the *Jurimetrics Journal* in 1966, and was later renamed *Jurimetrics* in 1978.

³⁵ Martha Gibbs and Eldridge Adams, ‘A Report on the Second National Law and Electronics Conference’ (1962) 3 *MULL Mod Uses Log L* 215, 222.

Beyond basic metrics provided by databases, and a few glimpses into the existing body of literature on artificial intelligence,³⁶ it is not easy to answer questions on how we can identify trends in topics which have attracted legal scholars and what these trends are, since no such study currently exists. Moreover, as we have seen above,³⁷ commercial and public perceptions regarding artificial intelligence did not entirely match scientific progress, creating a disconnect between industry expectations and research output. A similar narrative may exist with respect to how artificial intelligence has been perceived in legal literature, and whether expectations and projections expressed in the latter actually match scientific developments.

Generating a knowledge map that can provide insights into how legal scholars relate to artificial intelligence brings multiple benefits:

- (i) It helps legal scholars get a faster grasp on legal literature available on artificial intelligence;
- (ii) It helps legal scholars understand on a macro scale what research questions have been asked in the past, therefore offering a way to reduce the risk of repetitive research;
- (iii) It helps computer scientists, and researchers from other disciplines, to get a simplified, bird's eye view over legal concerns regarding artificial intelligence;
- (iv) It allows for comparisons between literature trends based in different disciplines;
- (v) It has the potential of generating additional research questions regarding the path of legal scholarship on artificial intelligence.

To create such a knowledge map, we have designed a two-part project, consisting of an automated analysis of a body of legal literature that ought to reveal how the latter developed over time. Part one focuses on the descriptive and normative underpinnings of this analysis, which are described in this chapter. Part two focuses on the actual data analysis, and will result in a separate computer science publication. In what follows, we will discuss the Corpus, the collection methodology, and the characteristics of the Corpus as revealed using descriptive statistics and data visualisation.

The Corpus includes a total of 3931 journal articles obtained from HeinOnline as a result of a Data and Text Mining Agreement. Absent a centralised, comprehensive, open access repository for international legal scholarship,³⁸ we focused on one of the commercially-available databases: we chose HeinOnline as one of the leading international databases on legal materials, containing over 170 million pages of literature and indexing over 2700 law journals.³⁹ In the section 'Law Journal Library', section type 'Articles',⁴⁰ the Corpus covers literature available in the database between 1960 and 2018. Unlike arXiv, HeinOnline does not have a section on 'artificial intelligence'. The total number of retrieved articles reflects the results of a Boolean search using the keywords 'artificial intelligence', namely all articles which include both terms. Resulting articles therefore discuss a wide array of aspects relating to artificial intelligence, and do not, as such, focus on specific technical or legal issues. We do, however, account for the number of pages with keywords in a given article, and are thus able to distinguish between different parts of the Corpus (see Figure 3 below):

³⁶ Supra note 10. See also Anne von der Lieth Gardner, *An Artificial Intelligence Approach To Legal Reasoning* (Boston, MIT Press 1987), 67-76.

³⁷ Supra note 23.

³⁸ The problems arising out of publication models for legal research are well-known. See Gary Wolf, 'Who Owns the Law' *Wired* (San Francisco, 1 May 1994) <<https://www.wired.com/1994/05/the-law/>> accessed 9 June 2019. See also Graham Lawton, 'Paywall: The business of scholarship review – analysis of a scandal' *New Scientist* (London, 5 October 2018) <<https://www.newscientist.com/article/2181744-paywall-the-business-of-scholarship-review-analysis-of-a-scandal/>> accessed 9 June 2019.

³⁹ See <<https://home.heinonline.org>>.

⁴⁰ All additional sections have been omitted, namely: 'Comments', 'Notes', 'Reviews', 'Legislation', 'Cases', 'External', 'Miscellaneous', 'Index', and 'Editorial'.

	Number of pages with keywords	Number of journal articles
Journal articles focused on artificial intelligence	40-2	1463
Journal articles mentioning artificial intelligence	1	2468
TOTAL		3931

Figure 3 – A classification of the HeinOnline Corpus

As can be seen in Figure 3, the articles indexed by HeinOnline under the keywords ‘artificial intelligence’ may refer to the keywords between one and 40 times. While this metric may tell us for sure that an article referring to the keywords 40 times will certainly be a legal article on artificial intelligence,⁴¹ the opposite may not always hold true. For instance, a paper dealing with copyright and Google Books which only refers to the keywords once is not necessarily *not* a paper on artificial intelligence, seeing as the latter is a technological product discussed within such a paper.⁴² Given the inherent method of legal analysis, namely the doctrinal method, it is difficult to draw strong lines and categorise to what degree a paper discusses artificial intelligence. However, for the purposes of this study, we assume that even one reference to the keywords is sufficient to include an article in the Corpus. This is because sometimes articles refer to the shortened version of the term (e.g. AI), or in other cases they refer to different sub-categories of artificial intelligence, such as machine learning, and one reference to the keyword is sufficient to define its use in the body of the article. Moreover, excluding single references to the keyword would not show the full extent of the interest shown by legal scholars in the topic. Figure 3 shows this category to include most of the articles in the Corpus. Some of these articles are very much related to artificial intelligence and in rare occasions are even written by computer scientists,⁴³ or belong to the wider umbrella of law and technology scholarship, whether in its earlier iterations⁴⁴ or in a more contemporary format.⁴⁵ One possible classification on the basis of this metric is to divide the papers between papers that focus on artificial intelligence (between 2 and 40 references to the keywords) and papers that refer to artificial intelligence more casually (one reference).

The lack of indexed books is an additional drawback, since a lot of legal literature is published in edited books. More data sources can be added to the Corpus, whether from HeinOnline or externally, leaving room for the expansion of this initial framework and dataset in order to facilitate further comparisons (e.g. what are the differences between the body of literature on artificial intelligence published in journals, as opposed to that published in books, etc.). The distribution of journal articles in the Corpus can be observed in Figure 4 below.

⁴¹ Henry H Jr Perritt, ‘Artificial Intelligence Techniques for Evaluating Employee Terminations on a Personal Computer’ (1987) 13 Rutgers Computer & Tech LJ 341.

⁴² See for instance Maurizio Borghi and Stavroula Karapapa, ‘Non-Display Uses of Copyright Works: Google Books and Beyond’ (2011) 1 Queen Mary J Intell Prop 21.

⁴³ Daniela Rus, ‘The Robots Are Coming’ (2015) 94 Foreign Aff 2.

⁴⁴ Anthony D’Amato, ‘Can/Should Computers Replace Judges’ (1977) 11 Ga L Rev 1277; Alan Wolfe, ‘Algorithmic Justice’ (1990) 11 Cardozo L Rev 1409.

⁴⁵ Jonathan Zittrain, ‘Engineering an Election’ (2013-2014) 127 Harv L Rev F 335.

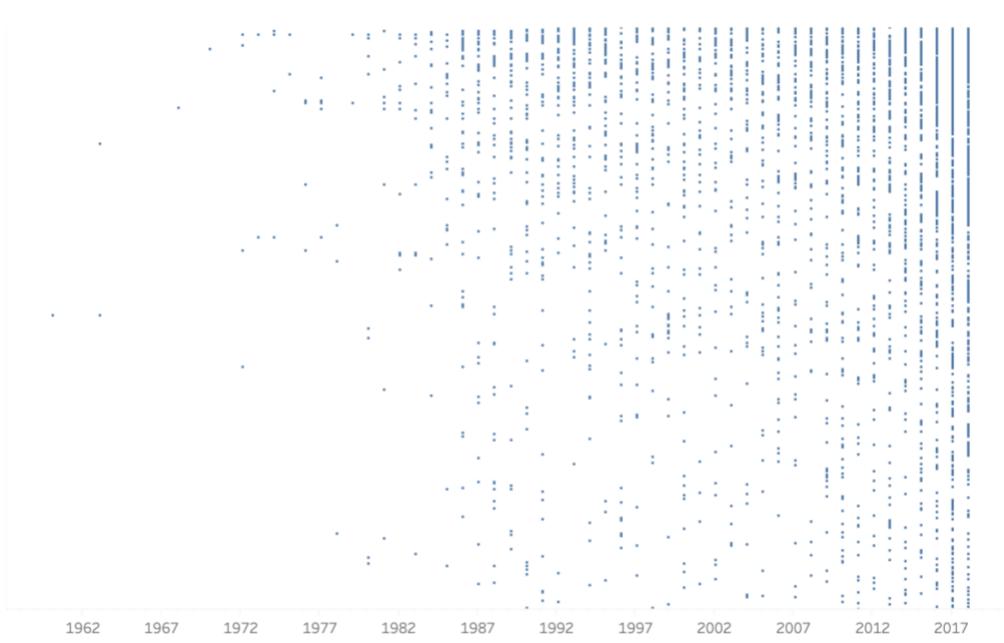


Figure 4 – The time distribution of the Corpus per journal per year⁴⁶

Figure 4 shows an increase in the popularity of artificial intelligence in the Corpus, especially between 2016 and 2018. A clearer overview can be seen in Figure 5, depicting the number of journal articles with keywords per year, and the number of pages with keywords per year:

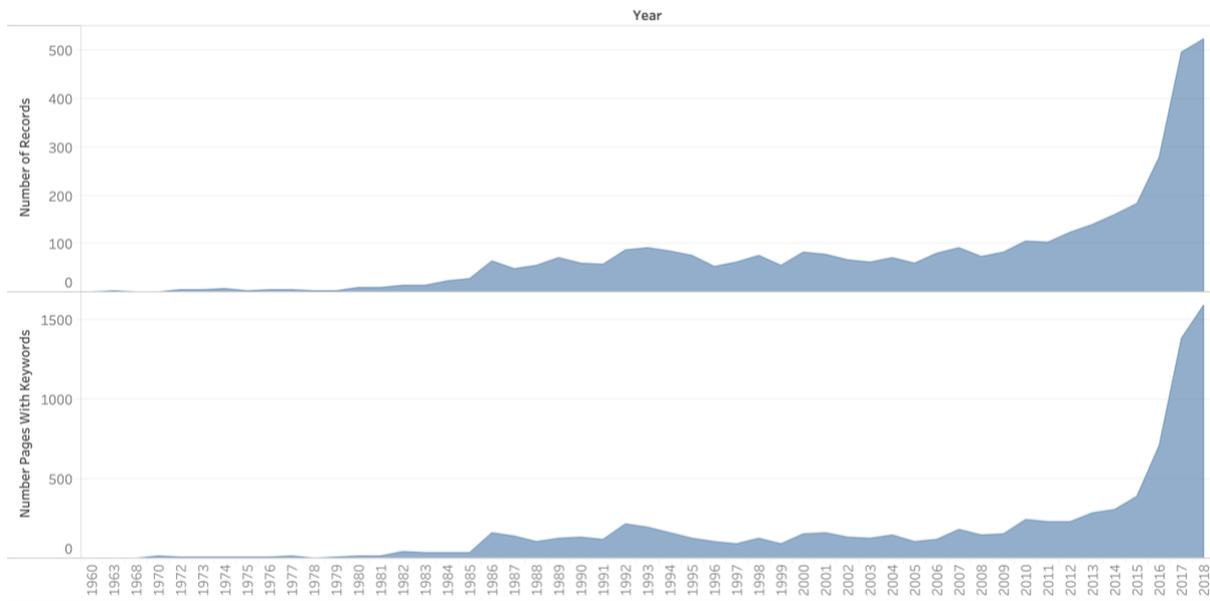


Figure 5 – Number of journal articles with keywords per year and number of pages with keywords per year⁴⁷

A detailed overview of the number of publications on artificial intelligence per journal – along with the complete list of the 1113 journals included in the analysis – are available on record and may be requested electronically until a project website is set up. Figure 6 shows a snippet of this analysis.

⁴⁶ Data visualisation done using Tableau.

⁴⁷ Data visualisation done using Tableau.

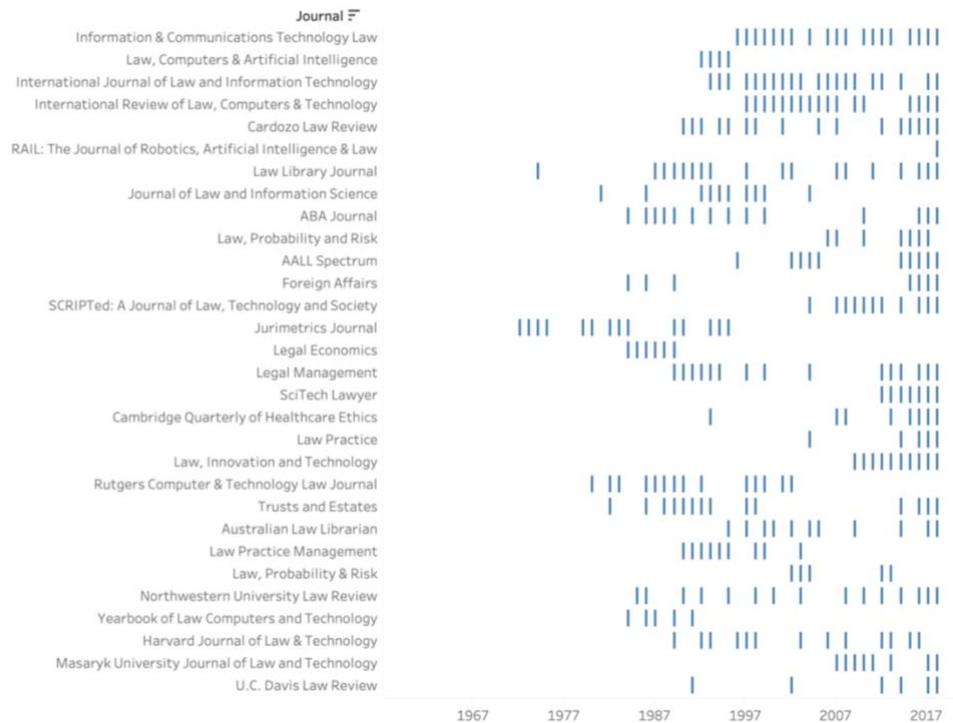


Figure 6 – Example of journal articles on artificial intelligence distributed per journal per year⁴⁸

The Corpus, especially as illustrated in Figure 5, shows similar patterns of growth in the volume of literature as the arXiv dataset from Figure 2. While computer science literature on artificial intelligence is obviously more voluminous than its legal equivalent,⁴⁹ between 2016 and 2018 we can identify a clear boom in scholarship pertaining to both disciplines.

In the next Section, we will further the overlap between law and computer science. Through the framework of the history of artificial intelligence described above, we will map examples of topics tackled by legal scholars in various parts of the Corpus.

II. Examples of Legal Research Topics on Artificial Intelligence

The illustration of topics in this Section is a qualitative exercise that proposes using the three AI waves identified in Section 1 (see Figure 1) to highlight some parts of the Corpus, in order to discuss selected topics found at its core. The insights below were derived from the data through a general literature review, albeit not systematic, given the total amount of articles. While not giving a full picture of the Corpus, the selected examples of articles and topics are sufficient to inform a normative discussion regarding the meaning and future of legal scholarship on artificial intelligence.

A. The First AI Boom: Exploring and Reasoning (1956-1974)

Early interest in artificial intelligence and law was a shy undertaking. The Corpus features 20 articles published between 1956 and 1974, most of them published in the *Jurimetrics Journal*; topics included

⁴⁸ Ibid.

⁴⁹ In 2018 alone arXiv indexed 3976 papers on artificial intelligence, compared to the 524 papers indexed by HeinOnline for the same year.

digital legal research,⁵⁰ the use of logic in the systematisation of law,⁵¹ or predictive legal analytics.⁵² During this period, one of the earliest comprehensive reflections on the impact of artificial intelligence on legal reasoning was authored by Bruce Buchanan and Thomas Headrick.⁵³ In this paper, interdisciplinarity was already regarded as a desirable but difficult to attain goal for two disciplines projecting misconceptions about one another.⁵⁴ In trying to achieve this goal, the two authors focus on the characteristics of the legal thought process, identifying four legal problem-solving processes: (i) finding conceptual linkages in pursuing goals, (ii) recognizing facts, (iii) resolving rule conflicts, and (iv) finding analogies.⁵⁵

Given its scarcity, this early work can be characterised even without further insights from data analysis. Its contributors show a high level of information literacy, due to interdisciplinary collaborations, like the one between Buchanan and Headrick, and also in the light of the other fields (e.g. logic or linguistics) necessary to conceptualise and systematise legal needs. This likely determined a concentration of thematic expertise, signalling the formation of a highly specialised research environment. Lastly, there seems to be a focus on processes relevant to the legal discipline (e.g. research and argumentation), and not so much on the content of theoretical legal concepts. From this perspective, this wave of law and artificial intelligence – while raising some normative questions on the role of technology in performing human functions associated with the legal profession –⁵⁶ focused on specific applications and not so much on the normative questions arising out of a potential paradigm shift.⁵⁷

B. The Second AI Boom: The Knowledge Base Era (1980-1987)

This era already sees an increase in the interest attached to law and artificial intelligence. In this period, the Corpus lists 208 articles published in a much broader array of publications, from law and technology journals such as the Rutgers Computer and Technology Law Journal, the Journal of Law and Information Science, the Yearbook of Law, Computers and Technology, to regular law journals such as the International Labour Review, the Southern California Law Review or the Modern Law Review.

Interestingly, during this period, parallels are drawn between novel issues posed by technology and other similar situations arising beyond the use of technology.⁵⁸ Among the topics discussed by authors writing between 1980 and 1987 we can find: management concerns with microcomputing in

⁵⁰ See for instance Philip Slayton, 'Electronic Legal Retrieval - A Report Prepared for the Department of Communications of the Government of Canada' (1974) 15 *Jurimetrics J* 108; Kenneth Katz, 'Computerized Research: An Editorial' (1973) 14 *Jurimetrics J* 25.

⁵¹ Peter B Maggs and Cary G deBessonnet, 'Automated Logical Analysis of Systems of Legal Rules' (1972) 12 *Jurimetrics J* 158.

⁵² Gerald Fleischmann and Philip Scaletta, 'Use of Computer Technology in Legal Analysis and Prediction' (1974) 11 *Am Bus LJ* 251.

⁵³ Bruce G Buchanan and Thomas E Headrick, 'Some Speculation about Artificial Intelligence and Legal Reasoning' (1970) 23 *Stan L Rev* 40.

⁵⁴ *Ibid.*

⁵⁵ *Ibid.*, 53.

⁵⁶ Buchanan and Headrick speak about the vision of 'removing the 'dog work' that underlies quality legal service, the computer system might make good service available to a larger segment of society', *supra* note 53, 61.

⁵⁷ For instance, at the time, Maggs and deBessonnet claimed the ultimate goal of research in law and artificial intelligence is 'ultimate goal is to enable machines to perform [the tasks of receiving, retaining information, and rendering cognitive responses to natural language questions] at levels of sophistication sufficient to warrant the use of formal expressions instead of natural language expressions in recording information, thereby making possible certain automated inferential and communicative process involving legal textual materials', *supra* note 51, 1.

⁵⁸ See for instance the comparison between a machine endowed with artificial intelligence and an organisation, Michael McDonald, 'The Personless Paradigm' (1987) 37 *U Toronto LJ* 212.

police activity,⁵⁹ reflections on the law applicable to robots as machines performing human tasks,⁶⁰ legal information processing,⁶¹ or text retrieval systems.⁶²

In addition to furthering the work on legal reasoning and information retrieval by discussing knowledge-based expert systems,⁶³ we see some emerging work on normative legal questions that do not focus on how artificial intelligence can be applied to law, but rather on whether artificial intelligence should be used at all in various societal spheres and, if so, what the legal repercussions ought to be.⁶⁴ At the same time, artificial intelligence is also mentioned in scholarship on jurisprudence.⁶⁵

C. The Third AI Boom: (Deep) Machine Learning (1993-present)

By far the largest part of the Corpus, this period (1993-2018) brings together a total 3353 articles on law and artificial intelligence. In this period, the number of journals on law and technology is on the rise (e.g. *Law, Innovation and Technology*; *Richmond Journal of Law and Technology*; *Journal of Business and Technology Law*; etc.), just as the scholarship not necessarily focusing on legal reasoning or information retrieval.

In the 90s, automated logic applied to legal information was still *en vogue*,⁶⁶ and had spread well beyond the United States,⁶⁷ yet it was soon replaced with applications of artificial intelligence – and digitalisation, more broadly speaking – pertaining to legal sub-fields such as intellectual property,⁶⁸ data protection,⁶⁹ fundamental rights,⁷⁰ or Internet governance.⁷¹ Philosophical considerations became more prevalent.⁷² The tone of a lot of the writings in this period became more sceptical, perhaps due to the initial scepticism clouding the commercial success and proliferation of personal computing and communication infrastructures, such as the Internet. Writings were focused on the limitations of existing technologies, and the need to maintain human control in certain aspects of the application of justice

⁵⁹ Armand P Hernandez, 'Police Microcomputing: Strategies and Concerns' (1986) 5 *Am J Police* 1.

⁶⁰ Michael Gemignani, 'Laying Down the Law to Robots' (1984) 21 *San Diego L Rev* 1045.

⁶¹ Cary G Debessonnet and George R Cross, 'An Artificial Intelligence Application in the Law: CCLIPS, a Computer Program That Processes Legal Information' (1986) 1 *High Tech LJ* 329. See also Richard Craswell, 'Interpreting Deceptive Advertising' (1985) 65 *BU L Rev* 657.

⁶² Jon Bing, 'Third Generation Text Retrieval Systems' (1981) 1 *JL & Inf Sci* 183.

⁶³ Richard E Susskind, 'Some Preliminary Considerations Concerning Expert Systems in Law' (1987) 14 *N Ky L Rev* 211; Richard E Susskind, 'Expert Systems in Law: A Jurisprudential Approach to Artificial Intelligence and Legal Reasoning' (1986) 49 *Mod L Rev* 168.

⁶⁴ Gemignani, *supra* note 60.

⁶⁵ See for instance the criticism of the goal of the artificial intelligence community to 'articulate a complete formalization of the mental processes involved in thought and action', George Wright, 'On a General Theory of Interpretation: The Betti-Gadamer Dispute in Legal Hermeneutics' (1987) 32 *Am J Juris* 191, 204.

⁶⁶ Antonio A Martino, 'Artificial Intelligence and Law' (1994) 2 *Int'l JL & Info Tech* 154; Kevin W Saunders, 'What Logic Can and Cannot Tell Us about Law' (1998) 73 *Notre Dame L Rev* 667; Edward S Adams and Daniel A Farber, 'Beyond the Formalism Debate: Expert Reasoning, Fuzzy Logic, and Complex Statutes' (1999) 52 *Vand L Rev* 1241.

⁶⁷ Henry Prakken, 'Logica en Recht: Wat Heeft de Kunstmatige Intelligentie ons Geleerd' (1993) 22 *R & R* 198; Giovanni Sartor, 'Defeasibility in Legal Reasoning' (1993) 24 *Rechtstheorie* 281.

⁶⁸ Samuel Oddi, 'An Uneasier Case for Copyright Than for Patent Protection of Computer Programs' (1993) 72 *Neb L Rev* 351.

⁶⁹ Eric Thiry, 'Personal Medical and Social Data: Their Processing and Legal Protection' (1993) 12 *Med & L* 643.

⁷⁰ Jacques Fremont, 'Computerized Administrative Decision Making and Fundamental Rights' (1994) 32 *Osgoode Hall L J* 817.

⁷¹ Klaus W Grewlich, 'Access to Global Networks - European Telecommunications Law and Policy' (1998) 41 *German YB Int'l L* 9.

⁷² Brett Lilly, 'Resistance is Futile: Nietzsche, Foucault, and the Aporia of Modern Legal Thought' (1995) 19 *Legal Stud F* 139. See also Louis E Raho and James A Belohlav and David R Drehmer, 'Expert Systems in Organizations: Legal and Ethical Considerations' (1994) 3 *L Computer & Artificial Intell* 47.

(e.g. judging).⁷³ While these trends seem to be maintained throughout the 2000s, with the explosion of literature as of 2016, writings start to become more granular and topics become more diversified, with some skyrocketing in popularity.⁷⁴

Precisely because of this popularity, methods of analysis often employed in legal research (e.g. qualitative content analysis) fall short of being able to generate an accurate overview of exactly what topics attract the interests of which legal scholars, and when. This is the main justification behind the need for part two of this study; in this section, we have merely given some illustrations of the type of literature included in the Corpus.

D. The Need for Further (Quantitative) Research: Topic Modelling

Part two of the research project focuses on applying quantitative data analysis methods to the Corpus, particularly topic modelling. As mentioned above, completing a qualitative systematic literature review of over 3900 papers is a highly problematic, if not impossible task to achieve for a legal scholar using traditional methods of legal inquiry. Yet in the absence of an overall image of existing literature, it is difficult to understand macro trends in a body of legal research. New and established researchers exploring topics at the intersection of law and artificial intelligence may have impressions of these trends, but evidence relating to the accuracy of such impressions is currently hard to come by. An example of this mismatch can be drawn from the field of empirical legal research: the general popularity buzz around empirical legal research has led to a general impression that more empirical research is currently being done in Europe, but this view was recently disproved through empirical evidence.⁷⁵

Understanding existing evidence with respect to what scholarship the legal community has been generating on law and artificial intelligence is the main goal behind the second part of the research project. In the field of empirical legal research, text analysis has so far been done with the help of human coding, a time-consuming and sometimes inaccurate method, especially when it needs to be applied to a high volume and high variety of unstructured text. Topic models are a category of statistical algorithms that address such shortcomings, as they are used to ‘summarize, explore and index large collections of text documents’.⁷⁶ Topic modelling algorithms such as Latent Dirichlet Allocation are also called ‘generative models’, since they are based on the assumption that any document is generated by sampling words given specific topic probabilities (and actually the model attempts to learn these probabilities. Given the assumption that observable data in the body of text may be clustered according to topics using probability theory, they are called ‘generative models’.⁷⁷ Topic modelling has been widely used in computer science,⁷⁸ but is increasingly used in the emerging field of computational social sciences, with the goal of augmenting social science analysis.⁷⁹

⁷³ See for instance Jeanne Lee, ‘The Era of the Computer Judge’ [1995] UCL Jurisprudence Rev 249.

⁷⁴ See for instance Jessica S Brodsky, ‘Autonomous Vehicle Regulation: How an Uncertain Legal Landscape May Hit the Brakes on Self-Driving Cars’ (2016) 31 Berkeley Tech LJ 851; Madeleine de Cock Buning, ‘Autonomous Intelligent Systems as Creative Agents under the EU Framework for Intellectual Property’ (2016) 7 Eur J Risk Reg 310; Andrew M Brown, ‘Blame It on the Machines: How Autonomous Vehicles Will Impact Allocation of Liability Insurance and the Resulting Impact on the Legal Community’ (2016-2017) 95 NC L Rev Addendum 29.

⁷⁵ Gijs van Dijck and Shahar Sverdlov and Gabriela Buck, ‘Empirical Legal Research in Europe: Prevalence, Obstacles, and Interventions’ (2018) 11(2) Erasmus L Rev 105.

⁷⁶ Ryan Wesslen, ‘Computer-Assisted Text Analysis for Social Science: Topic Models and Beyond’ (2018) <<https://arxiv.org/pdf/1803.11045.pdf>> accessed 9 June 2019, 1.

⁷⁷ Ibid.

⁷⁸ See also David M Blei, ‘Probabilistic topic models’ (2012) 55 Communications of the ACM 7784.

⁷⁹ Justin Grimmer and Brandon M Stewart, ‘Text as data: The promise and pitfalls of automatic content analysis methods for political texts’ (2013) 21 Political Analysis 267.

Topic models fall under the general umbrella of Natural Language Processing (NLP), which covers the whole area of making machines understand how people use language. In general, NLP takes into account all levels of language formation (morphology, syntax, semantics, discourse, etc.). However, for simplicity reasons, many statistical models are based on the so-called ‘bag-of-words’ (BoW) assumption: text documents are introduced in a document-term matrix (DTM), which will count the occurrence of specific words in specific documents, thus ignoring the word order or even the semantics of the words. While the performance of topic models in more complex NLP problems such as question answering is not great, for collections of documents that are large enough, ‘the BoW assumption provides the theoretical foundation for a richer set of statistical methods (mixture models) by the assumption of exchangeability’.⁸⁰ One of the most often used topic models is the Latent Dirichlet Allocation (LDA) topic model.⁸¹ LDA, as a generative model, assumes the existence of a number of topics and then tries to estimate the document-topic mixtures (i.e. the mixtures under which documents are generated conditioned on all the topics of the collection) and the topic-word probabilities (i.e. the mixture probabilities under which words are generated, conditioned on a specific topic). A more concrete explanation of how this topic model is applicable to legal literature is included in part two of the project, which will be published separately as a computer science publication.

This Section provided selected illustrations of the legal scholarship included in the Corpus, and highlighted some trends which remain to be proved by the analysis of the dataset using the latent Dirichlet allocation topic model. A lot of normative questions arise, however, from this brief descriptive assessment of the Corpus. In the following section, we will raise questions regarding the meaning and future of legal research on artificial intelligence.

III. Time and Research: Questions for the Legal Community

In 1970, Buchanan and Headrick, a computer scientist and a legal scholar from Stanford University, were writing: ‘[...] research proceeding in computer science could enhance our understanding of the processes by which lawyers work and think. So far lawyers have not attempted to explore its relevance. They should.’⁸² And so they did. Yet what has the legal community achieved so far in doing so, and where is this research headed? We will endeavour to reflect upon this question by going through a handful of observations based on the analysis above.

First, it seems there is a trend for legal research on artificial intelligence to generally become less specialised. As we could see from the early body of legal work on artificial intelligence, only a handful of researchers were interested in the topic, driven by the same appreciation of formal reasoning and the goal of applying artificial intelligence to very practical questions of legal research. Even without the results of topic modelling, the decentralisation of legal research interests in artificial intelligence is notable when becoming more familiar – albeit qualitatively – with the body of legal literature on law and technology from the past decades. The crowd of researchers covering artificial intelligence is the largest ever, yet what is equally visible is that technology literacy has dropped rather than improved, in spite of the increase in the volume of literature and the granularity of technical and legal issues. A sign of this shift is the coverage of law and technology questions by not only specialised journals such as *Jurimetrics*, dedicated to a niche audience of legal informatics aficionados, but also by general legal journals serving as outlets for legal communities with much broader interests. This may have hindered, to some extent, the process of authors specialising on topics that involves understanding highly complex

⁸⁰ Wesslen, *supra* note 76, 2.

⁸¹ David M Blei and Andrew Y Ng and Michael I Jordan, ‘Latent Dirichlet allocation’ (2003) 3 *Journal of Machine Learning Research* 993.

⁸² Buchanan and Headrick, *supra* note 53, 62.

technologies, especially in the light of the fast pace at which these technologies are developing. Of course, content experts – especially found in the emerging field of law and technology – who have been writing about artificial intelligence for more than the past couple of years, are the exception. However, the noticeable explosion of literature as of 2016 could very well have resulted in the dilution of the level of expertise necessary when dealing with the intricacies of artificial intelligence.

Second, the legal community must seriously reflect upon the cyclicity of questions in research on artificial intelligence and law. As we have seen above,⁸³ the idea of robots fulfilling human tasks considerably pre-dates any hype surrounding Sophia, the first robot to be offered the citizenship of a country.⁸⁴ While the field of robotics may have seen considerable technological progress since the mid 80s, the matrix of legal questions – as well as answers – which may be given in response to the challenge of integrating robots in society is rather finite. Should a robot, or any type of autonomous system, be held liable for losses caused by it? The answer to this question is either a negative or a positive one. Surely there can be many different variations built on each of these answers, however the factors that ought to be taken into account when dealing with this matter (e.g. personhood, insurance, etc.) do not raise an unlimited number of permutations. From this perspective, legal scholars must very carefully consider the added value of contributing to topics where thorough research unveils that most – if not all – of these permutations have already been addressed. This does not mean that writing one article on such a complex question as liability for autonomous agents will suffice, but we question whether too many articles on this topic actually advance palpable solutions or make considerable contributions to existing inquiries. From the perspective of time, we ought to consider what it means for the same legal questions to be asked for decades over and over again: does this mean the questions are so complex that a solution has simply not been found yet? It could be the case, for instance, when dealing with the early efforts of using computer-based logic to represent legal argumentation. The approach and the technology available to deploy it were simply not ripe enough in the 1960s for legal scholars to be able to apply automated content review software to contracts, but it is being applied with some commercial success today.⁸⁵ In this case, subsequent iterations of the question may contribute to solving the problem. However, there is also the option of spending decades asking questions that have the same (limited amount of) answers: should we allow for automated decision-making in a courtroom? This topic may be trending right now, but authors like D’Amato and Lee have been asking this question since 1977 and 1994 respectively,⁸⁶ and a lot of the points they make in their papers are currently circulated in literature on algorithmic discrimination. If the latter does not build on earlier work but rather reiterates it, we are left wondering why that would be the case. While there may be more explanations, at least two points are noteworthy: there can be a disconnect between earlier and current research (e.g. no citations between different time clusters), or perhaps the incentives behind asking these questions might be different. The latter point will be further explored under the fourth observation below.

Third, legal research ought to include more meta-studies and literature reviews. The legal issues arising out of artificial intelligence are becoming more and more technically and societally sophisticated (e.g. recommender systems; targeted advertising; etc.), thus driving the quest to understand and manage human behaviour from a legal point of view. From this perspective, legal researchers ought to be able to embrace law and technology – and artificial intelligence in particular – without necessarily having

⁸³ Gemignani, *supra* note 60.

⁸⁴ Emily Reynolds, ‘The agony of Sophia, the world’s first robot citizen condemned to a lifeless career in marketing’ *Wired* (San Francisco, 1 June 2018) <<https://www.wired.co.uk/article/sophia-robot-citizen-womens-rights-detriot-become-human-hanson-robotics>> accessed 9 June 2019.

⁸⁵ See for instance <<https://www.lawgeex.com/aboutus/>>.

⁸⁶ D’Amato, *supra* note 44; Lee, *supra* note 71.

had to do a PhD in the topic. However, this brings us back to the first observation made in this section: the extent to which contributions on this topic are specialised given that, as opposed to being experts on narrow yet interdisciplinary topics such as information retrieval or expert systems, authors increasingly originate from other fields of law (e.g. criminal law, contract law, administrative law) and look into various applications of artificial intelligence impacting the field in question. Specialisation, as a benefit arising out of expertise, is accrued when a researcher reaches an in-depth understanding of a given topic. For this reason, expertise takes time. However, the design of the legal research landscape itself hinders an efficient process of gaining sufficient information in order to develop meaningful expertise: legal literature is generally decentralised, paywalled and often not linked; the legal community has not managed to harmonise its research indexing (e.g. Web of Science); databases with legal information – wherever available – are rarely user friendly or grant access to meta data; many researchers do not use Boolean searches; legal journals generally do not publish literature reviews; and the list may continue. On the one hand, all these factors increase the time and complexity attached to legal research activities. On the other hand, this is a perfect environment to design and test new optimisation strategies. One example to this end is the generation of knowledge maps, as undertaken by the research project underlined in this chapter. Knowledge maps help researchers gain a complete picture of a specific field, topic or question; depending on the number of variables used to build it, researchers may apply those to understand a defined field at a faster pace, or uncover new patterns or ideas in a given dataset. Undoubtedly, this means researchers need to start attaching more value to building, maintaining, linking and reusing datasets. Similarly, this also means the legal field needs to embrace a paradigm shift with respect to how information is perceived and managed, as researchers can no longer keep track of knowledge by reading every single article available on a given research question – this will simply become physically impossible in the next decade, due to the exponential propagation of academic literature. In this context, legal information in general has the characteristics of big data and may benefit from more automated, systematic data analysis and visualisation.⁸⁷

Lastly, the research community could also benefit from reconsidering publication incentives. Doing legal research, just as any other kind of research, is ideally about furthering the boundaries of knowledge, and contributing to either the theory or practice represented by this knowledge. What hypes in legal research reveal – at least during the most considerable growth between 2016 and 2018 – is that sometimes, published legal research does not fulfil this goal. Instead, scholars may sometimes choose to focus on topics that are popular because they may increase their reputation or their funding chances. To the extent that each of these considerations has a direct bearing on the quality of legal research, current publication incentives may be detrimental to the type of consolidated and coordinated efforts that are necessary for a field to experience knowledge breakthroughs.

Conclusion

In this chapter, we have embedded the theme of the book – time and law – in the context of mapping several waves of legal literature on artificial intelligence. To do so, we centred the chapter around a dataset obtained from HeinOnline. This dataset reflects all legal journal articles indexed by the database between 1960 and 2018 referencing ‘artificial intelligence’. The dataset (called throughout the chapter ‘Corpus’) involves a two-fold analysis: the current chapter in a legal publication, where we contextualise it, describe it, and use examples to depict its content; and a subsequent computer science publication where the dataset is analysed using a statistical model called Latent Dirichlet Allocation, a central tool in NLP which explores the thematic structures of document collections.

⁸⁷ Catalina Goanta, ‘Big Law, Big Data’ (2017) October (Special Issue) *Law and Method* <<https://www.bjutijdschriften.nl/tijdschrift/lawandmethod/2017/10/lawandmethod-D-17-00007>> accessed 9 June 2019.

We started this chapter by outlining a brief history of artificial intelligence, and then proceeded to describing the dataset using descriptive statistics and data visualisation made with the help of the Tableau software.⁸⁸ We then overlapped the insights from the historical overview on artificial intelligence with literature available in the Corpus, and discussed examples of topics included therein. This analysis enabled a general reflection on the meaning and future of legal research on artificial intelligence.

Overall, in the chapter we proposed the generation of knowledge maps, such as the one to be made using topic modelling, because of a number of benefits. These include: helping legal scholars to get a faster grasp on legal literature available on artificial intelligence; helping legal scholars to understand on a macro scale what research questions have been asked in the past, therefore potentially reducing the risk of repetitive research; helping computer scientists, and researchers from any other discipline, to get a simplified, bird's eye view over legal concerns regarding artificial intelligence; allowing for comparisons between literature trends based in different disciplines; generating additional research questions regarding the path of legal scholarship on artificial intelligence. Additional research may complement the current project by adding new datasets, in order to better explore the potential of knowledge maps in legal research on artificial intelligence.

⁸⁸ Some parts of this statistical analysis are available on record and may be requested via e-mail until a project website is set.