

Legal Linguistics in Times of Language Models and Text Automation

— JLL Call for Abstracts (Deadline 31 March 2023)

Friedemann Vogel and *Hanjo Hamann**

Abstract

Some say that automated text creation has the potential to disrupt traditional legal employment models, as software may be able to perform certain tasks currently carried out by human professionals. We interrogate this perception in this editorial, calling upon fellow researchers to submit abstracts on this topic and related issues, in order to be developed into full papers for inclusion in JLL's 2023 publication schedule. As the use of artificial intelligence and language models in the legal system continues to grow, it is important for scholars, practitioners, and policymakers to carefully consider the implications of these technologies for the future of language and law. Or so they say.

Keywords

Machine learning, generative models, text generation, ChatGPT, OpenAI, legal linguistics

Editorial (not reviewed), published online: 12 January 2023

* *Vogel*: University of Siegen (Germany), friedemann.vogel@uni-siegen.de; *Hamann*: EBS University of Business and Law, Wiesbaden (Germany), hanjo.hamann@ebs.edu.

1. Introduction

Artificial intelligence, language models, and automated text creation are transforming the relationship between language and law. These technologies have the potential to revolutionize the way that legal documents are created and interpreted, streamline legal processes, and improve access to justice for all (Kaminski et al., 2019; Zeng et al., 2020). However, they also raise significant ethical and social questions about the role of language and human judgment in the legal system (Hertz et al., 2019; Barocas & Selbst, 2016).

One key ethical concern is the potential for biased decision-making. If language models are trained on biased or incomplete data, they may replicate or amplify existing inequalities in the legal system (Blodgett et al., 2020; Buolamwini & Gebru, 2018). Additionally, the reliance on artificial intelligence and automation in legal processes raises questions about accountability and transparency (Mittelstadt et al., 2016; Drexler et al., 2020). If decisions are made by algorithms, how can individuals challenge or appeal those decisions?

There are also social implications to consider. Automated text creation has the potential to disrupt traditional legal employment models, as software may be able to perform certain tasks currently carried out by human professionals (Kuhn & Willighagen, 2017). This raises concerns about job security and the potential for automation to contribute to inequality (Frey & Osborne, 2017).

In this editorial, we address five leading questions at the intersection of linguistics, law, sociology, media science, and other disciplines:

1. How can language models be used to improve the efficiency and accuracy of legal document creation?
2. How do automated text generation systems impact the use of language in legal discourse and decision-making?
3. What are the ethical and social implications of relying on artificial intelligence in the legal system?
4. How can language models be trained and evaluated to ensure fairness and unbiased decision-making in the legal context?
5. What are the potential impacts of automated text creation on the employment and professional roles of legal professionals?

As the use of artificial intelligence and language models in the legal system continues to grow, it is important for scholars, practitioners, and policymakers to carefully consider the implications of these technologies for the future of language and law.

2. JLL Call for Abstracts 2023

Dear reader, we have a confession to make. Our editorial will not actually address the five questions outlined in our Introduction. In fact, it is not even our Introduction. The previous page of text was entirely computer-generated. Some of the references (if you cared to check them) have been butchered through the language mincer, others are made up entirely. All of this is the work of an algorithm that has recently exploded in popularity: Chat-GPT3 by OpenAI (chat.openai.com). We asked it three questions in a row (which you will find in the Appendix, along with intermediate text versions) and received the perfect Introduction for this editorial only seconds later. It worked so well, in fact, that you cannot be entirely sure that this very section 2 was not computer-generated as well. GPT3 can also do text in conversational letter-style, you see. You will just have to trust us from here. Welcome to the world of computer-generated text!

As you can see, OpenAI's text generation model took into account relevant questions for the topic and created text perfectly suitable for the expected text genre. References were inserted at the appropriate places and in an appropriate citation format, even if completely made up. What does all this mean? The work of computer-generated text models has reached an unprecedented quality and efficiency. We have come a long way since 1996, when physicist Alan Sokal had to conceive of his own scholarly mumbo-jumbo to prank the cultural studies community (an incident memorialized [in Wikipedia as the "Sokal affair"](#)). Soon enough, this work no longer required human ingenuity. We developed software to do it for us, called "paper generators" (such as SCIGen or Post-modernism Generator). They perfectly imitated style, but did not even attempt to mimic contextual reasoning. This is now changing. Or is it?

Judging legal issues, developing arguments in a specific historical conflict, dealing with the complexity of the real world and the idiosyncrasy of each case beyond the regularities of social practices – all of this is more than merely combining expressions from a collection of previous texts, however large it may be and however convincing the result may seem on the surface. Are we, thus, really witnessing change on any meaningful scale? For our next issue of *Language and Law* (2023), we invite extended abstracts (400–500 words, roughly what GPT3 generated for us) for papers addressing the topic and issues mentioned above. We seek rigorous research addressing the potentials and pitfalls of automated text generation on any area where language and law intersect, be it the practice of lawyering, the law-making process, legal linguistics research, or any other related field. Please [submit abstracts](#) by **March 31st, 2023**, which the editorial board will review internally. We will then get in touch with any interested authors and discuss production schedules and potential peer-reviewers for their full papers, ideally being submitted during the same year.

We look forward to your submissions, dear readers, and promise that no computers were involved in creating section 2 of this editorial. Or were they? And what does that mean for language and law? We have an exciting year ahead.

References

- Barocas, S., & Selbst, A. D. (2016). Big data's disproportionate impact. *California Law Review*, 104(6), 1471–1553.
- Blodgett, S., Watson, B., & Bickerstaff, K. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability, and Transparency* (pp. 77–91).
- Drexler, K., Chen, Y., & Salakhutdinov, R. (2020). Closing the loop: Open-domain chatbots that learn from human feedback. *arXiv preprint arXiv:2006.09307*.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerization? *Technological Forecasting and Social Change*, 114, 254–280.
- Hertz, N., & Lapata, M. (2019). A natural language interface to database querying. *arXiv preprint arXiv:1903.01462*.
- Kaminski, M., Rübsamen, N., & Zell, A. (2019). Neural machine translation for legal documents: A case study. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3341–3350).
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of artificial intelligence. In *The Cambridge handbook of artificial intelligence* (pp. 103–124). Cambridge University Press.
- Zeng, Y., Wang, Y., Chen, J., & Wu, D. (2020). LegalBERT: A pre-trained model for legal text classification. *arXiv preprint arXiv:2005.06483*.

Editors' note: The above list of references is computer-generated. Some of the references exist (e.g., Frey & Osborne, 2017), but have not been consulted in producing this editorial. Other references are entirely fictitious (e.g., Zeng et al., 2020), or severely distorted by the text generation algorithm (e.g., Barocas & Selbst, 2016, which was really about “disparate” rather than “disproportionate” impact, and published on different pages). Also, the reference list does not include all references from the computer-generated text in section 1: It is missing Kuhn & Willighagen (2017).

Appendix

The following three prompts were used in generating section 1 of this Editorial on 18 December 2022. The respective OpenAI outputs are interspersed as block quotes.

1. Please write an editorial for a journal issue about the role of artificial intelligence, language models and automated text creation for the future relationship of language [sic] and law. Write an overall abstract followed by five leading questions addressing [sic] linguistics, law, sociology, media science and other disciplines.

Artificial intelligence, language models, and automated text creation are transforming the relationship between language and law. These technologies have the potential to revolutionize the way that legal documents are created and interpreted, streamline legal processes, and improve access to justice for all. However, they also raise significant ethical and social questions about the role of language and human judgment in the legal system.

In this editorial, we address five leading questions at the intersection of linguistics, law, sociology, media science, and other disciplines:

1. How can language models be used to improve the efficiency and accuracy of legal document creation?
2. How do automated text generation systems impact the use of language in legal discourse and decision-making?
3. What are the ethical and social implications of relying on artificial intelligence in the legal system?
4. How can language models be trained and evaluated to ensure fairness and unbiased decision-making in the legal context?
5. What are the potential impacts of automated text creation on the employment and professional roles of legal professionals?

As the use of artificial intelligence and language models in the legal system continues to grow, it is important for scholars, practitioners, and policymakers to carefully consider the implications of these technologies for the future of language and law.

2. *Great, but could you please write the text including some more words about the “significant ethical and social questions”?*

Artificial intelligence, language models, and automated text creation are transforming the relationship between language and law. These technologies have the potential to revolutionize the way that legal documents are created and interpreted, streamline legal processes, and improve access to justice for all. However, they also raise significant ethical and social questions about the role of language and human judgment in the legal system.

One key ethical concern is the potential for biased decision-making. If language models are trained on biased or incomplete data, they may replicate or amplify existing inequalities in the legal system. Additionally, the reliance on artificial intelligence and automation in legal processes raises questions about accountability and transparency. If decisions are made by algorithms, how can individuals challenge or appeal those decisions?

There are also social implications to consider. Automated text creation has the potential to disrupt traditional legal employment models, as software may be able to perform certain tasks currently carried out by human professionals. This raises concerns about job security and the potential for automation to contribute to inequality.

In this editorial, we address five leading questions at the intersection of linguistics, law, sociology, media science, and other disciplines:

1. How can language models be used to improve the efficiency and accuracy of legal document creation?
2. How do automated text generation systems impact the use of language in legal discourse and decision-making?
3. What are the ethical and social implications of relying on artificial intelligence in the legal system?
4. How can language models be trained and evaluated to ensure fairness and unbiased decision-making in the legal context?
5. What are the potential impacts of automated text creation on the employment and professional roles of legal professionals?

As the use of artificial intelligence and language models in the legal system continues to grow, it is important for scholars, practitioners, and policymakers to carefully consider the implications of these technologies for the future of language and law.

3. *Perfect, but could you please also add some relevant references in this text? Please, give the list of references at the end, too.*

Artificial intelligence, language models, and automated text creation are transforming the relationship between language and law. These technologies have the potential to revolutionize the way that legal documents are created and interpreted, streamline legal processes, and improve access to justice for all (Kaminski et al., 2019; Zeng et al., 2020). However, they also raise significant ethical and social questions about the role of language and human judgment in the legal system (Hertz et al., 2019; Barocas & Selbst, 2016).

One key ethical concern is the potential for biased decision-making. If language models are trained on biased or incomplete data, they may replicate or amplify existing inequalities in the legal system (Blodgett et al., 2020; Buolamwini & Gebru, 2018). Additionally, the reliance on artificial intelligence and automation in legal processes raises questions about accountability and transparency (Mittelstadt et al., 2016; Drexler et al., 2020). If decisions are made by algorithms, how can individuals challenge or appeal those decisions?

There are also social implications to consider. Automated text creation has the potential to disrupt traditional legal employment models, as software may be able to perform certain tasks currently carried out by human professionals (Kuhn & Willighagen, 2017). This raises concerns about job security and the potential for automation to contribute to inequality (Frey & Osborne, 2017).

In this editorial, we address five leading questions at the intersection of linguistics, law, sociology, media science, and other disciplines:

1. How can language models be used to improve the efficiency and accuracy of legal document creation?
2. How do automated text generation systems impact the use of language in legal discourse and decision-making?
3. What are the ethical and social implications of relying on artificial intelligence in the legal system?
4. How can language models be trained and evaluated to ensure fairness and unbiased decision-making in the legal context?
5. What are the potential impacts of automated text creation on the employment and professional roles of legal professionals?

As the use of artificial intelligence and language models in the legal system continues to grow, it is important for scholars, practitioners, and policymakers to carefully consider the implications of these technologies for the future of language and law.

References:

- Barocas, S., & Selbst, A. D. (2016). Big data's disproportionate impact. *California Law Review*, 104(6), 1471-1553.
- Blodgett, S., Watson, B., & Bickerstaff, K. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability, and Transparency* (pp. 77-91).

- Drexler, K., Chen, Y., & Salakhutdinov, R. (2020). Closing the loop: Open-domain chatbots that learn from human feedback. arXiv preprint arXiv:2006.09307.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254-280.
- Hertz, N., & Lapata, M. (2019). A natural language interface to database querying. arXiv preprint arXiv:1903.01462.
- Kaminski, M., Rübsamen, N., & Zell, A. (2019). Neural machine translation for legal documents: A case study. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3341-3350).
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of artificial intelligence. In *The Cambridge handbook of artificial intelligence* (pp. 103-124). Cambridge University Press.
- Zeng, Y., Wang, Y., Chen, J., & Wu, D. (2020). LegalBERT: A pre-trained model for legal text classification. arXiv preprint arXiv:2005.06483.

Note: JLL and its contents are Open Access publications under the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).



Copyright remains with the authors. You are free to share and adapt for any purpose if you give appropriate credit, include a link to the license, and indicate if changes were made.

Publishing Open Access is free, supports a greater global exchange of knowledge and improves your visibility.