Applying NLP to Support Legal Decision-making in Administrative Appeal Boards in the EU

Henrik Palmer Olsen[1], Malte Højmark-Bertelsen[2] and Sebastian Felix Schwemer[3]

[1]iCourts, University of Copenhagen, Karen Blixens Plads 16, 2300 Copenhagen, Denmark
[2]KMD Denmark, 8000 Aarhus, Denmark
[3]Centre for Information and Innovation Law, University of Copenhagen, Karen Blixens Plads 16, 2300 Copenhagen, Denmark

Abstract
While Natural Language Processing (NLP) is being applied in an increasing number of contexts, including law, it remains a difficult task to leverage NLP for the purpose of real-life support of legal decision-making. This is because 1) legal-decision making must be made in a way that is sensitive not only to legislation but also to evolving case practice (prior decision-making that functions as precedent), 2) legal-decision making is sensitive to open-ended legislative language and shifting factual contexts, 3) traditional methods of NLP are capable of processing long texts, but they are suboptimal compared to novel methods, i.e., transformer-based models, e.g., BERT [1], etc. 4) however the transformer-based models are limited by maximum input lengths, which makes it difficult to apply in real-life scenarios, where legal documents exceed the maximum input length. In this paper, we show how we tackle the problem of providing NLP-based intelligence support to legal decision-makers in a real-world setting using transformer-based NLP.

Keywords
Legal information retrieval, NLP, public administration, automation bias, decision support, legal decision-making

1. AI for legal decision-making in public administrative practice: rule of law challenges

Legal decision-making takes place in numerous contexts every day: in courts, in public administration, and in private businesses. These decisions are often complex and rely to a large extent on costly human expertise. Such decision-making is often slow and expensive, but it is also often morally and socially important and performed by the use of discretionary and context-sensitive judgments.

If AI can be used to support real-life legal decision-making, there will be a significant potential to reduce cost and processing time. Moreover, there is a potential that AI support may even increase decision quality and advance justice by furthering a more homogeneous use of discretion in decision-making practice [2][3][4].[1] Yet, the field of AI and law is also beset with difficulties as noted, e.g., by Deakin and Markou [5].

This paper contributes to state-of-the-art in two ways. First it sets out to explain how NLP can be used to enhance legal decision-making in administrative decision-making that is characterized by both having a large number of cases decided under the same legal rule and by being very sensitive to discretionary judgment by case-workers deciding these cases. More specifically we explain how automated information from precedent cases have the potential to advance both efficiency and overall legal equality. Secondly we show how the maximum input length in transformer based language models provides a challenge for application of these models in the context of legal information retrieval, and propose a solution to overcoming this challenge.

Recent research documents several examples of AI-driven legal decision-making in public administration which exhibit failure to comply with rule-of-law principles and fundamental rights [6]. A prominent concrete example is the so-called SyRI case from the Netherlands, where the use of an algorithmic social welfare fraud detection system was found by the Hague District Court to violate the European Convention of Human Rights [7][8]. A similar use of an algorithmic system in the US State of Michigan was found to be fraught with error, unjustly depriving thousands of citizens of unemployment benefits [9]. Yet another recent AI and law scandal in the Netherlands, where citizens were wrongly accused of child care benefits fraud led to the government resigning in early 2021 following a parliamentary inquiry.[2]

These examples show that, while desirable in theory, it is difficult in practice to develop automated legal decision-

Proceedings of the Sixth Workshop on Automated Semantic Analysis of Information in Legal Text (ASAIL 2023), June 23, 2023, Braga, Portugal
henrik.palmer.olsen@jur.ku.dk (H. P. Olsen); hjb@kmd.dk (M. Højmark-Bertelsen); sebastian.felix.schwemer@jur.ku.dk (S. F. Schwemer)

[1]There are also ongoing political initiatives to try to leverage the advantages of AI in legal information and legal decision-making practices, see e.g: https://joinup.ec.europa.eu/collection/better-legislation-smoother-implementation (for the EU); https://en.digst.dk/policy-and-strategy/digital-ready-legislation/ (for Denmark).

[2]https://en.wikipedia.org/wiki/Dutch_childcare_benefits_scandal

making that is ethically sound and lawful. There are numerous reasons why this is so, but here we focus on one specific challenge: In most legal systems there is a requirement under public administrative law to perform individual discretion based on specific facts in each individual case. What this means is that public administrators are not allowed to reduce the discretionary scope set out in the law by introducing easy-to-use rules as these would deprive citizens of their right to have their case decided on the basis of a full appreciation of how the relevant facts in their case are judged against the rules and standards that apply to the case at hand. At the same time, public agencies are required to decide like cases alike, which means that they must not arbitrarily treat citizens differently in like situations. Navigating this decision space is notoriously difficult to break down into fixed criteria embedded in a code [5]. Legal decision-making can in other words not be automated in a simple decision tree. Thus, there is a need to rethink the way AI can be used to support legal decision-making processes in public administration and beyond.

2. Overcoming rule-of-law challenges: Using AI to support case-based reasoning

If full automation is not an option (because it is neither feasible nor desirable in certain case-handling scenarios), then what part of the legal decision-making process in public administration could be AI-assisted in order to unlock potential efficiency and quality gains without undermining legal compliance?

In the LEGALESE project, we develop an information retrieval module for case-handling software that uses an NLP model to match new case descriptions to descriptions of prior cases that have been decided manually by caseworkers. We implement this model to a specific decision-making practice in a highest instance administrative agency and we take the agency's prior decisions in the selected practice area to be a gold standard, meaning that new cases should probably (but not certainly) be decided in the same way as similar previous cases.

In centralized public administration, there often exist a lot of repetitive cases. No cases are of course identical, but they may often be very similar in regard to the facts of the case that is relevant to the law in question. In LEGALESE, we operationalize our case match system in the context of decisions on Danish welfare law, more specifically a rule, selected in collaboration with the Appeals Board, that provides a right for families with children who suffer from reduced physical or mental ability to get coverage of necessary additional expenses.[3] We selected decisions relating to this rule because of the large volume of cases and because caseworkers at the Appeals Board called out these cases as being particularly difficult to deal with. Therefore this case area has a high potential for both quality enhancement (obtaining a better articulated and homogeneous practice) and efficiency gain (less time spend per case).

Denmark is divided into 98 municipalities, and each municipality has a social welfare administration unit that makes decisions (on delegation from the municipal board) on applications for welfare support under the specific rule in the Danish welfare law mentioned above (§41). When a citizen has its application for welfare support under this article rejected, they can file a complaint to the Appeals Board. The Appeals Board receives complaints from all municipalities in Denmark and decides around 800 complaints on §41 every year.[4]

Deciding these cases cannot easily be automated because there is no clear metric for deciding when a disability is "significant", when a disorder is "long-term", when an expense is "necessary", or when an expense is "additional". Each of these criteria is spelled out in the decision-making practice of the Appeals Board. This practice is described in general terms in the Board's practice guidelines, but these guidelines cannot be transcribed to unambiguous rules. There is, as mentioned above a requirement to perform a concrete assessment in each individual case, which must not be reduced to a formulaic rule. For this reason, we focus on supporting inductive reasoning from previous decision practice.[5]

This approach to AI and law is not new. It has been previously explored under the heading of "case-based reasoning systems"[11][12].[6] Case-based reasoning systems aim at solving new problems by retrieving stored 'cases' that describe prior problem-solving episodes similar to a new problem (case).[7]

[3] §41: "The municipal board must provide coverage of necessary additional expenses for providing at home for a child under the age of 18 with a significant and permanently reduced physical or mental ability to function or an intervening chronic or long-term disorder. It is a condition that the additional expenses are a consequence of the reduced functional capacity and cannot be covered according to other provisions of this Act or other legislation." The original Danish version of this rule can be found here: https://www.retsinformation.dk/eli/lta/2022/170 (visited 18 December 2022). The Appeals Board decides cases that are appealed to the Board after a decision is made in the municipality.

[4] The same caseworkers also decide cases on §42, which provides access to the salary loss experienced by parents who opt to care for their children at home. The Appeals Board decides more than 1000 of these cases per year. These cases contain sensitive personal information and we can therefore not make this dataset available.

[5] For a similar view, see Branting et al. [10] who emphasize that: "Denial of benefits by an automated process, no matter how accurate, raises significant due-process issues ..."

[6] For an overview of various artificial intelligence approaches applied to law, see [13].

[7] Note that this is different from the approach by Branting et al. [10] who use attention network-based prediction to find relevant text.

In human decision-making practice, case-based reasoning is a well-known method used in bureaucratic institutions. New cases are often resolved by seeking out similar past decisions from decision archives. Such retrieval of prior cases is either based on the memory of individual human caseworkers who have worked up an experience with deciding cases of the same kind or on getting information from well-informed colleagues or both. Sometimes information can also be retrieved from case archives, by searching through these. Various ways of systematizing such archives exist and there are various ways of searching through these. Existing computer-operated case retrieval systems often have limited search functionalities and provide less than optimal search results when queried. Our aim is therefore to improve both case retrieval efficiency and case retrieval accuracy by implementing an NLP model.

In the LEGALESE project, we introduce an NLP model that reads selected documents from the corpus of all prior §41 cases and compares these documents against the same kind of documents in the new case. This model could be called a document match algorithm, but because the ultimate aim is to compare cases we refer to it as Case Match. To operationalize a workable Case Match for our real-life situation we needed to reduce computational complexity and this meant selecting the same specific documents from all cases as representative of case content for the purposes of calculating document-to-document similarity.

Selecting which documents from a case archive are the most relevant representations of the full case content is a problem that can only be solved by relying on domain expertise. Hence for the construction of our document match algorithms, we conducted interviews with caseworkers at the Appeals Board with experience in deciding §41 cases. More specifically we first conducted a collective unstructured interview with three caseworkers and their team manager with a view to reaching a consensus on which documents in the case files contain the most essential elements relevant to represent the cases on file. We used a workshop format to conduct these interviews (see further below in section 4.1.). Subsequently, we conducted individual semi-structured interviews with three caseworkers with varying work experience (from a few months to several years) in regard to deciding on §41 cases and two managers with institutional responsibility for the decisions made. Through these interviews, we learned that caseworkers are tasked with and given the competence to decide cases on their own after a learning period, where their work is supervised by a more experienced case worker. We also learned that caseworkers are expected to decide (on average) one case per day. We also noted significant differences between the interviewed caseworkers in regard to what knowledge sources they rely on when handling their cases.

The knowledge we gained from these interviews allowed us to identify the most relevant documents in the case files, thereby reducing algorithmic and computational complexity. Still, as we shall discuss below, even with this reduction, we face the challenge that there is a significant gap between state-of-the-art transformer-based NLP and real-world legal document length: Transformer-based NLP performance is limited to 4096 tokens, but many of the documents we need to match are up to 3-4 times longer and sometimes even longer than that. In section 4, we explain how we overcome this problem.

After computing a similarity score between case documents (see further details below) Case Match shows the entire case files associated with the documents that have the highest similarity score. This allows human caseworkers to receive faster and more qualified information about the most similar previously decided cases, thereby enabling a smoother case-based reasoning process and better decision-making efficiency and quality.

It should be noted that in designing this model we made the deliberate choice not to showcase outcomes directly to caseworkers as this could advance unwanted automation bias, i.e. the "possible tendency of automatically relying or over-relying on the output produced" by automated legal decision-making tools.[8]

The primary focus of the LEGALESE project is to bring relevant legal reasoning from prior cases forward to the caseworkers so that they may draw inspiration from this. Thereby LEGAELSE makes it easier for caseworkers to decide on their own whether to follow reasoning laid out in prior decisions (if the facts of the new case are judged to be sufficiently similar to one or more of the matched cases) or to depart from this and create new reasoning more specifically tailored to the new case at hand (if it is found not to match).[9] This approach is central to the LEGALESE project as it supports the requirement in public administration that like cases should be treated alike a requirement that is sometimes referred to as a principle of equality.[10] The principle of equality

Our approach is to find relevant cases that contain reasoning that the case worker can rely on in the new case. We therefore operationalise a case similarity system rather than identifying specific text passages from former cases that may be deemed relevant in the new case. Computationally though, there is a overlap in the techniques used.

[8] Defined in Article 14(4) lit.b of the draft Artificial Intelligence Act [14]. If passed, the provision would require "high-risk AI systems" to be designed and developed so they are subject to human oversight and that individuals remain aware of automation bias [15][16][17].

[9] Whereas not indicated by our interviewees, we note that there may be instances where caseworkers would rely on previous decisions that might be relevant even though they are not similar in most parts of the document.

[10] For an introduction to the principle of equality in the context of EU law, see, e.g., [18].

builds on the fundamental idea that everyone is equal in front of the law and that the law applies in an equal manner to all. Hence, when two cases are alike in all relevant aspects they should be decided the same way. What counts as "relevant aspects", however, is a matter of discretion and cannot be automated [19]. The advantage of Case Match is that in instances where a caseworker decides that cases are sufficiently similar and need to be decided in the same way, they can copy the language in the prior decision into the new decision, thereby giving the "likeness" judgment a textual representation that will streamline decision-making in future cases. Similarly, when cases are considered to be not sufficiently similar, the decision will be flagged as not sufficiently similar by the creation of new decision text that departs from the most similar prior decisions. We estimate that this, over time, may enhance both decision efficiency and quality.

3. Operationalizing Natural Language Processing models in the context of legal case data

Caseworkers at the Social Appeals Board begin their work on a new case by picking it from an online folder containing all new incoming cases. Once the case is picked, the caseworker will be able to see all the metadata for the case as well as all the documents and appendices belonging to the case. Furthermore, they are presented with a column presenting a number of the most similar previous cases for the given case, i.e. Case Match results.

There are some important design choices to be made for the Case Match functionality. How many prior similar cases should be shown? Should the system be set up so that it shows the best matching cases in different outcome categories? Should recent similar cases be given priority over older similar cases?[11] We will test and discuss various solutions in collaboration with the domain experts testing the system as the LEGALESE project unfolds.

As mentioned above, the similarity function in Case Match operates by transforming selected documents from all case files in a database of previously decided cases into vectors. In the LEGAELSE project, we test three different methods for document vectorization: 1) TF-IDF vectorization, 2) a transformer-based language model with legal domain adaptation, and 3) a transformer-based language model, also with legal domain adaptation, but furthermore, trained with spectral decoupling to mitigate the inherent biases of language models [20]. These three methods all allow for an efficient vector-based search and calculation of cosine distance similarity scores between documents.

[11] This also relates to the question of how to deal with changes in the administrative practice (i.e., when there is a change in interpretation). As noted above, case-based reasoning systems aim at solving new problems by retrieving stored 'cases' that describe prior problem-solving episodes similar to a new case. In the (rare) instance of a change in interpretation (or law), thus, the system must reflect these developments.

4. Overcoming the text length problem

As mentioned above, Case Match uses either TF-IDF or transformer-based language models for document vectorization. Using transformer-based language models, however, poses a problem regarding the maximum input length for the language models [21], which has also been mentioned in previous work about finding similar cases [22]. The way these language models vectorize text is by first, tokenizing the text and then indexing these with their vocabulary to create a general vector representation of the text. These models are however often limited to a maximum input length of 512 or fewer tokens [23], which is far less than the average total case text length of the documents domain experts at the Danish Appeals Board pointed out as being of essential importance to represent case content. To overcome this limitation, we extended the length of the Danish BERT[12] from 512 tokens to 4096 tokens, which is also one of the mentioned future directions in a recent survey on long text modelling with transformers [24]. This solves some of the issues, but a maximum input length of 4096 tokens is still not sufficient for generating vector representations for all the text in many of the relevant case documents. We, therefore, developed a method for identifying the most salient parts of the different documents attached to each of the cases stored in the database of previously decided cases.[13]

4.1. Creating an accurate vector representation with unstructured data

Cases decided in the social appeals contains many different documents: applications from parents; statements from doctors; reports from teachers, pedagogues, etc; decisions from the municipality, etc. Comparing a new case to an old case is therefore a complex matter involving comparison across many documents in each case. case complexity and diversity is a major obstacle in operationalizing an automated case retrieval system for similar cases. We therefore set up a workshop with the participating caseworkers at the appeals board to try to reduce

[12] https://huggingface.co/Maltehb/danish-bert-botxo

[13] It should be noted, of course, that TF-IDF does not have a length limit, so when testing this, we are fitting it on all the text in the documents, and not using the method for overcoming the maximum text length problem of the transformers.

case complexity without loosing depth of information about the cases. During this workshop we found that there are in general four documents in every case that contain the most salient information about the content of the case. We use the four documents in every case to calculate case similarity. The four documents are: 1) the initial decision of the municipality in the case; 2) the citizen complaint about the municipality's decision; 3) the reevaluation of the case by the municipality; and 4) the Danish Appeals Board's decision.

We know that the Appeal Board's decision constitutes the ultimately correct decision for a case[14], and is, therefore, the document which contains the information most relevant to decision outcome. Moreover, the Appeals Board decisions all resemble each other in terms of style and length as they are written up using a standard format. We also found that these documents, would usually not exceed 4096 tokens, whereas the other three documents could be of any length (usually above 4096 tokens) and format. With this knowledge, we created a method for using the Appeal Board's decision as a reference point for identifying relevant information in the other three documents. The method consisted of first dividing all documents into text windows of 4096 tokens, where the Appeal Board's decision document would consist of 1 window, whereas the other three documents could consist of multiple windows, depending on their word length. We then vectorize the windows (except for when we test tf-idf which do not have the same restraints as the transformer models). Having vectorized all the constituent window parts we could now use the Appeal Board's decision document and use it to calculate a similarity between it and each of the other different document windows allowing us to identify which 4096 token window in each of the other documents had the most representative information about the case. This allowed us to find the most relevant part of the two documents from the municipality and the citizen complaint (as measured against the final decision in the case, which is the measure we used for overall relevance in the case). We saved both the document vectors and the calculated similarity values. These could then be used for calculating a weighted case vector, where each similarity was applied as a weight for the average sum between the documents, thus, obtaining the most accurate vector representation for each case.

4.2. Calculating the case similarity for open cases

While the above method allowed us to calculate similarities between all existing closed cases in the Appeals Board database, we still needed a way to handle new and

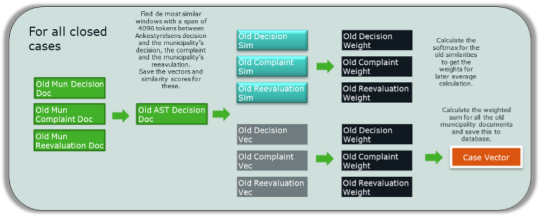

Figure 1: Creating an accurate vector representation with unstructured data

open cases, where no decision document exists yet. We did this by again taking the three documents from the citizen complaint and the two municipality decision documents (initial decision and reevaluation) respectively and dividing these into windows of 4096 tokens. Hereafter, for every closed case, we took the vector of each relevant document (see section 4.1. above) and compared these with same kind of documents in the open case. This allowed us to find the part of the three documents, where the text was most similar, compared to the same documents in the closed cases. With these new open case document vectors and similarity scores, we used the closed case document weights to calculate the weighted sum of the similarities, thus obtaining an overall case similarity score, allowing us to calculate the cosine similarity between a given open case and closed cases.

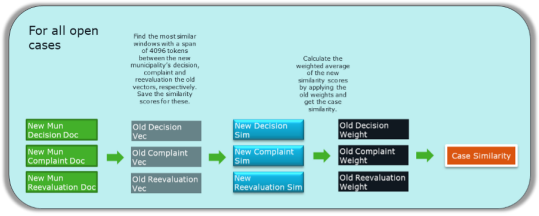

Figure 2: Calculating the case similarity for open cases

5. Conclusion, challenges, and suggestions for further research

Using transformer-based language models to build automated decision support for legal decision-making is demanding for two reasons: Firstly, document length, legal complexity, and demands for a comprehensive examination of circumstances in each case make it difficult. Secondly, increasing demands from European regulation relating to personal data protection[15] [25][26] and devel-

[14]Decisions by the Appeals Board are very rarely subject to judicial review, and when it is the review is constrained to procedural matters.

[15]Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free

opment and use of AI systems [14][16][15][27] in addition to the requirements under general administrative laws make it a demanding exercise with considerable legal uncertainty to build compliant automated-handling practices.

The approach in our LEGALESE project is therefore to avoid these issues by closely supporting existing non-automated case-handling practices. Instead of relying on profiling and fully automated decision-making which raises data protection concerns, we use an approach to decision-making support that is recognizable and comprehensible to caseworkers (intelligence assistance rather than automated decision-making): searching for similar previous cases and using these as inspiration to decide new cases. By doing so we do not suggest a whole new method for administrative decision-making, but instead seek to provide enhanced legal information retrieval skills to support a case-work practice that is already well-established in the Social Appeals Board. LEGALESE also aims to avoid automation bias. Rather than suggesting a decision outcome or producing an automatic draft of the decision in the new case, the system only brings relevant previous cases forward to the case worker. The caseworker then has to make an active choice about how to use the cases shown to them in Case Match. In LEGALESE we test the Case Match functionality with three different models, where we transformed the text into vectors representing the text in the case documents. However, when using document length-limited transformer-based language models we had to develop a novel comparison algorithm, where we compared the case documents to previous decisions made by the Danish Appeals Board to identify the most relevant piece of text. Conclusively, this allowed us to calculate representative similarity values for all of the cases, allowing the caseworkers to see the most similar cases in their document database.

It is one thing to succeed in automating information retrieval through a model for measuring similarity across complex legal files; it is another to succeed in achieving perceived value of such an automated retrieval system. In LEGALESE we will perform evaluation through a questionnaire format that will be issued to those caseworkers who are testing the system. The questionnaire focuses on caseworkers' perceived experience of whether or not the system provides them with similar cases. We deliberately use an empirical approach to the evaluation of the systems performance because our aim is to assist the legal reasoning process as it is perceived by real life caseworkers.

Within this approach for the implementation of decision support, there are still improvements that can be made. Here we shall highlight a few:

- Firstly, there is a need and potential for improving the methods for processing long documents. There has been conducted a lot of research regarding improving transformer-based language models' ability to process longer sequences and reducing the computational cost. The Nyströmformer [28], for example, is a novel modeling approach that significantly reduces the cost, while having the ability to process long documents. However, no such Danish model was available at the time of the LEGALESE project. This, thus, entails a need for more development within Danish natural language processing, which could be training better Danish language models with novel model architectures.
- Secondly, a feature of the system that could significantly improve the Case Match functionality would be to incorporate a feedback system, where users could give feedback. The feedback could consist of the caseworker evaluating whether a match was good or bad. This would result in concrete training data for Case Match which would allow the training of models from human feedback. Other types of data and information that could be utilized in such a feedback system could be metrics about user behavior in the system. E.g., by using something similar to "internet cookies" we could investigate how much time caseworkers spend on different cases and try to infer, from data, if a case was a good or a bad match.
- Thirdly, it could be considered to highlight specific textual fragments in prior cases predicted to match the information needed for the decision of a current case. By this we mean that if it were possible to predict which part of the closed decision document would be most useful to copy into the open case decision document, then we could automatically highlight this part, making it easier for a caseworker to identify and copy this. It should also be remembered though, that this would also increase the risk of introducing automation bias because it could have a nudging effect and simultaneously make it easier for the caseworker to use that specific text fragment in the new decision. There is a trade-off between increasing automation and preventing automation bias in a legal decision making process about issues that are sensitive for citizens.
- Lastly, going beyond Case Match, information extraction techniques could be applied to enrich the metadata of the cases, which could provide case workers with more information in their decision-making process.

movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

Acknowledgments

This research is part of the LEGALESE project at the University of Copenhagen, co-financed by the Innovation Fund Denmark (grant agreement: 0175-00011A).

References

[1] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[2] D. Hoadley, Beyond classical retrieval: Case law and natural langauge processing, Austl. L. Libr. 28 (2020) 116.

[3] G. Peruginelli, S. Faro, Knowledge of the Law in the Big Data Age, volume 317, ios Press, 2019.

[4] B. Alarie, A. Niblett, A. H. Yoon, How artificial intelligence will affect the practice of law, University of Toronto Law Journal 68 (2018) 106–124.

[5] S. Deakin, C. Markou, Is law computable?: critical perspectives on law and artificial intelligence, Bloomsbury Publishing, 2020.

[6] M. Zalnieriute, L. B. Moses, G. Williams, The rule of law and automation of government decision-making, The Modern Law Review 82 (2019) 425–455.

[7] S. Bekker, Fundamental rights in digital welfare states: The case of syri in the netherlands, Netherlands Yearbook of International Law 2019: Yearbooks in International Law: History, Function and Future (2021) 289–307.

[8] R. D. Haag, Syri legislation in breach of european convention on human rights, 2020. URL: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx.

[9] S. Ranchordas, Empathy in the digital administrative state, Duke LJ 71 (2021) 1341.

[10] L. K. Branting, C. Pfeifer, B. Brown, L. Ferro, J. Aberdeen, B. Weiss, M. Pfaff, B. Liao, Scalable and explainable legal prediction, Artificial Intelligence and Law 29 (2021) 213–238.

[11] C. Hafner, Legal reasoning models, in: N. J. Smelser, P. B. Baltes (Eds.), International Encyclopedia of the Social Behavioral Sciences, Pergamon, Oxford, 2001, pp. 8675–8677. URL: https://www.sciencedirect.com/science/article/pii/B0080430767005866. doi:https://doi.org/10.1016/B0-08-043076-7/00586-6.

[12] K. D. Ashley, Artificial intelligence and legal analytics: new tools for law practice in the digital age, Cambridge University Press, 2017.

[13] J. Dias, P. A. Santos, N. Cordeiro, A. Antunes, B. Martins, J. Baptista, C. Gonçalves, State of the art in artificial intelligence applied to the legal domain, 2022. URL: https://arxiv.org/abs/2204.07047. doi:10.48550/ARXIV.2204.07047.

[14] E. Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, COM/2021/206 final, 2021.

[15] S. F. Schwemer, L. Tomada, T. Pasini, Legal ai systems in the eu's proposed artificial intelligence act, in: Proceedings of the Second International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2021), held in conjunction with ICAIL, 2021.

[16] M. Veale, F. Z. Borgesius, Demystifying the draft eu artificial intelligence act—analysing the good, the bad, and the unclear elements of the proposed approach, Computer Law Review International 22 (2021) 97–112.

[17] E. U. A. for Fundamental Rights, Bias in Algorithms – Artificial Intelligence and Discrimination, Publications Office of the European Union, 2022.

[18] G. Barrett, Re-examining the concept and principle of equality in ec law, Yearbook of European law 22 (2003) 117.

[19] S. Wachter, B. Mittelstadt, C. Russell, Why fairness cannot be automated: Bridging the gap between eu non-discrimination law and ai, Computer Law & Security Review 41 (2021) 105567.

[20] I. Chalkidis, A. Søgaard, Improved multi-label classification under temporal concept drift: Rethinking group-robust algorithms in a label-wise setting, 2022. URL: https://arxiv.org/abs/2203.07856. doi:10.48550/ARXIV.2203.07856.

[21] J. Ainslie, S. Ontanon, C. Alberti, V. Cvicek, Z. Fisher, P. Pham, A. Ravula, S. Sanghai, Q. Wang, L. Yang, Etc: Encoding long and structured inputs in transformers, arXiv preprint arXiv:2004.08483 (2020).

[22] C. Xiao, H. Zhong, Z. Guo, C. Tu, Z. Liu, M. Sun, T. Zhang, X. Han, Z. Hu, H. Wang, J. Xu, Cail2019-scm: A dataset of similar case matching in legal domain, 2019. arXiv:1911.08962.

[23] D. Mamakas, P. Tsotsi, I. Androutsopoulos, I. Chalkidis, Processing long legal documents with pre-trained transformers: Modding LegalBERT and longformer, in: Proceedings of the Natural Legal Language Processing Workshop 2022, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 2022, pp. 130–142. URL: https://aclanthology.org/2022.nllp-1.11.

[24] Z. Dong, T. Tang, L. Li, W. X. Zhao, A survey on long text modeling with transformers, 2023.

arXiv:2302.14502.

[25] S. Wachter, B. Mittelstadt, L. Floridi, Why a right to explanation of automated decision-making does not exist in the general data protection regulation, International Data Privacy Law 7 (2017) 76–99.

[26] L. Tosoni, The right to object to automated individual decisions: resolving the ambiguity of article 22 (1) of the general data protection regulation, International Data Privacy Law 11 (2021) 145–162.

[27] C. of Bars, L. S. of Europe (CCBE), CCBE position paper on the proposal for a regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act), 2021.

[28] Y. Xiong, Z. Zeng, R. Chakraborty, M. Tan, G. Fung, Y. Li, V. Singh, Nyströmformer: A nyström-based algorithm for approximating self-attention, 2021. URL: https://arxiv.org/abs/2102.03902. doi:10.48550/ARXIV.2102.03902.