Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-023-01693-z OPEN

Using the interest theory of rights and Hohfeldian taxonomy to address a gap in machine learning methods for legal document analysis

Ahmed Izzidien[1]✉

Rights and duties are essential features of legal documents. Machine learning algorithms have been increasingly applied to extract information from such texts. Currently, their main focus is on named entity recognition, sentiment analysis, and the classification of court cases to predict court outcome. In this paper it is argued that until the essential features of such texts are captured, their analysis can remain bottle-necked by the very technology being used to assess them. As such, the use of legal theory to identify the most pertinent dimensions of such texts is proposed. Specifically, the interest theory of rights, and the first-order Hohfeldian taxonomy of legal relations. These principal legal dimensions allow for a stratified representation of knowledge, making them ideal for the abstractions needed for machine learning. This study considers how such dimensions may be identified. To do so it implements a novel heuristic based in philosophy coupled with language models. Hohfeldian relations of 'rights-duties' vs. 'privileges-no-rights' are determined to be identifiable. Classification of each type of relation to accuracies of 92.5% is found using Sentence Bidirectional Encoder Representations from Transformers. Testing is carried out on religious discrimination policy texts in the United Kingdom.

[1] The Faculty of Law, The University of Cambridge, Cambridge, UK. ✉email: ai297@cam.ac.uk

Introduction

Artificial intelligence (AI), and specifically Natural Language Processing (NLP) through Machine Learning (ML), is increasingly being utilised in public interest technologies, such as in government departments, courts, and NGOs offering legal services (de Sousa et al., 2019). Using a systematic protocol for literature reviews and meta-analysis (PRISMA) a study (de Sousa et al., 2019) demonstrated that managers of public organisations have considerably increased the adoption of AI-based systems to improve efficiency (Mehr et al., 2017). For example, it has been found that a month's work at the US Department for Labour can be completed in a day with higher accuracy. Machines take into account factors at orders of magnitude greater than people without tiring. Indeed, this human frailty, when not checked, has been found to be a factor that negatively impacts decisions, such as court rulings (Danziger et al., 2011).

It has further been found that implementing ML in such contexts has made procedures more efficient. The Supreme Court of Brazil found AI contributed positively to its procedural speed, for example (de Sousa et al., 2022). Big data NLP also allows for a systematic analysis of documents at scale, allowing for new findings to materialise which are typically not possible with the human eye (Nay, 2018). Indeed, almost all law is expressed in natural language; therefore NLP may be said to be a key component to understand and predict law at scale (Nay, 2018).

Legal documents may be characterised as essentially texts which describe power interactions in society with consideration to the outcomes of such interactions (Boswell and Smith, 2017; Oliver and Cairney, 2019). These power interactions typically exist to further the interests of one or more parties involved (Hewitt, 2009; Michael et al., 2011).

Based on the nature of these interests, rights and duties and their concomitant legal relations become imposed (Kramer, 2001, 2017). Given that policy documents, which find themselves into legislation through a well-documented process (Parliament, 2021) are based on a consideration of the interests of parties affected, and the legal aspects of these interests (Policy Exchange, 2016), it is of note that to date no specific software explicitly captures the interests of a party expressed within a document. Nor does any specific software library automatically detect the legal relations expressed within these documents. Yet, it is these dimensions that have been used by policy analysis and jurist to analyse such policy and legislative documents.

As an analogy, machine learning that is trained on the game *Tetris*, learns dimensions of height, width, and velocity to avoid conflicting moves. In doing so it captures the dimensions needed to best analyse the game. We propose that in order to capture the most pertinent dimensions of a legal document for ML, one ought to measure its principal dimensions.

While deep learning offers the promise of highly accurate text classification though multitudes of neuronal activations in order to capture language (Sun et al., 2021), a common question persists, what specific dimensions have been captured? As law operates on the basis of rights and duties, and the balance—or imbalance—of these, as shall be detailed in the paper, it may be that any computational characterisation of legal language that excludes—or only partially captures—these abstract concepts will be at odds with what humans require of this knowledge. Even when initial results are correct or at least legally cogent, the question remains, what methodology was implicitly used? Was the correct result arrived at through an incorrect implicit methodology? Could the result be consistent with legal expectations, but for the wrong reasons? If we want to be able to explain the outcome, one possible method would be to highlight the principal features of the data.

It is an aim of this paper to attempt to discover whether it is possible to capture principal legal dimensions using ML. This approach may be considered a first step towards explainable outcomes based on such dimensions (Beckh et al., 2021), although this paper's remit is the former.

These dimensions are comprehensively covered by the established legal theory. Our task in this paper is to ask whether these dimensions can be detected using ML. The dimensions have been described by the interest theory of rights (Kramer, 2010) and by Yale jurist Wesley Newcomb Hohfeld (Hohfeld, 1913). Interest theory considers acts such as 'the man murdered the passer-by', or 'the client ordered her car from the garage' and allots a duty where the interest of either party is at risk of being harmed. The former sentence elicits a duty to abstain from committing murder. Whereas the latter sentence generates a duty to meet the client's order.

These duties can also be expressed as their correlative rights: a man has a right not to be murdered, and a client has the right to have their order met. This correlativity is described by Hohfeld in his seminal work on clarifying legal relations (Kramer, 2019). These rights and duties are considered first-order relations. The remining two first-order relations are: privileges and no-rights. Whereby a privilege concerns acts that are optional, i.e., they are not duties. An example would be 'Samuel is going for a walk'. Its correlative assignment is a no-right. i.e., other people have no claim rights against Samuel that he must or must not go for a walk.

These legal relations are readily identifiable by jurists. We ask, is it possible to detect these dimensions in sentences, by using NLP machine learning?

Secondly, we ask, are there any ethical implications of using such technology?

In order to answer these questions, we begin with an introduction to legal theory, specifically the Hohfeldian matrix of legal rights, the interest theory of rights, and the reasons for their selection in this study. This is followed by a philosophical discussion on how rights and duties are to be identified -given the vast range of ethical stances taken by societies and cultures. We propose the use of a philosophical heuristic that allows for cross-cultural application. This is then digitised using a language model (LM). Lastly, we consider how the first-order Hohfeldian taxonomy can be identified using word embeddings and a variety of sentence transformers. To do so we implement custom sentence vectors and language models.

Results are then presented followed by a discussion incorporating a question on the ethical implications and potential hazards of using ML in this domain.

The paper contributes uniquely by proposing a new approach for measuring human interactions based on a universal ethical heuristic. It also contributes by demonstrating that despite the subtleties of differences between Hohfeldian legal relations, language models can be trained to accurately differentiate between them. The work goes towards the advancement of legal reasoning tools, and their use to analyse humanities and social science texts along legally comprehensive dimensions.

Machine learning Hohfeldian rights and duties

Two main theories on the function of rights have been put forward in the literature (Kramer, 2000). The interest theory of rights and the will theory of rights. Interest theory holds that the principal function of human rights is to protect and promote the essential human interests possessed by all human beings (Tasioulas, 2015). Will theory on the other hand maintains that the function of a right is to give the right-holder control over another party's duty (Hart, 1982).

This paper has undertaken to use the interest theory of rights. This is done for the following reasons. Unlike the will theory of rights, the interest theory of rights avoids the need to consider the ability to exercise power and make rational choices in order to become a right bearer. A requirement that some argue limits its capacity to include children, the mentally incapacitated and animals, for example (Kramer, 2000; Kurki, 2018). While will theory reserves the term 'rights' for claims that are combined with enforcement/waiver powers and liberties in the hands of the claimholders, it has been argued that such a dichotomy is related to the theory's appeal to individual discretion and self-determination. This conflation of the political with the legal is absent from interest theory. Interest theory maintains that any genuine right does not have to be waivable and enforceable by the right-holder. It makes no attempt to determine who has such rights based on their powers and allows for a more stratified representation of right, one that conflates as few concepts as possible (Kramer, 2000).

As we aim to use legal theory in machine learning, one goal in this paper is to identify those theories that present as few potential confounding factors as possible, those that present factors in their most fundamental form (Hastie et al., 2003). Interest theory arguably lends itself to such. The right of a person in interest theory is one factor: their interest. Whereas a right in will theory incorporates several factors: the person's age, mental capacity, ability to waive rights, and being human. The political element can also be quite explicit, for example, Simmonds rejects any analytical jurisprudence of rights which is separated from normative political theory (Simmonds, 2000). A stance that may be based on an attempt to reconcile individual interests with collective governance (Wellman, 2000). In contrast, the interest theory of rights is free of such political dimensions.

This clear delineation of factors is also seen with Hohfeld. He describes rights in terms that are distinct and separate from political notions of entitlement. In Hohfeld's taxonomy, a right-holder does not necessarily have to have the power to enforce their right. This distinction is seen across his dimensions. Their uniqueness as a description of relations is that they are fundamental to the delineation of rights, and are considered to be the lowest generic conceptions to which any and all 'legal quantities' may be reduced (Hohfeld, 1923; Kurki, 2019).

The conceptual distinctions of the dimensions have been described as elegant, rigorous, and subtle (Kramer, 2000). Furthermore, his analytical framework is neutral in the debate on interest and will theory. His dimensions have provided distinct terms to help philosophers and jurists avoid ambiguous thinking and argumentation (Frydrych, 2017). His analysis is such that any normative or moral justificatory considerations can be present outside of rights elements, but not within (Lazarev, 2005). Which in turn avoids the possibility of conflating moral and rights factors within the analysis. It also simplifies operations related to the subject (Lazarev, 2005).

Thus, the features of the Hohfeldian dimensions are: (i) any legal interest can be broken down into an aggregate of Hohfeldian dimensions. (ii) these dimensions are irreducible, i.e., they cannot be broken down into anything more basic. It is for these reasons we have selected Hohfeld.

When comparing this level of discernment with other models, it is apparent that alternative models can suffer from a lack of discerning relations at a fundamental and irreducible level. For example, when considering Honoré's model (Honoré, 1961), Honoré's own description of ownership can be interpreted as an aggregate of rights and duties to other persons. His discussion on a 'right to possess' is a mixture of liberties and claims; his 'right to manage' is a similar collection of liberties and immunities (Eleftheriadis, 1996).

Although our paper will focus on first-order relations, Hohfeld also used equally distinct second-order dimensions: powers to alter legal relations and immunities from such changes. Other models tend to conflate them rather than use them independently. For example, Terry's model omits immunities entirely, a second-order Hohfeldian feature (Terry, 1884; Cook, 1919). In Salmond's model (Salmond, 1902) privileges and immunities are treated as relatively trivial, and liability is treated as the correlative of both privilege and power. This allocation of a single correlative for two independent conceptions has been seen as unclear, for if the distinction between privilege and power be valid then the distinction between the correlatives, liability and no-right, must be equally valid (Cook, 1919).

While Hohfeld's scheme has been criticised, much of the criticism has been seen as misplaced in that it engages by offering alternative stipulations of a right instead of demonstrating Hohfeld's model to be deficient (Hislop, 1967; Kramer, 2000; Frydrych, 2017). Some authors prefer the term 'liberty' to 'privilege' (Wenar, 2005) while others maintain the original label. While the two are at times used interchangeably, the content expressed by Hohfeld remains. Few jurists have doubted the thoroughness used in the taxonomy, and despite the passing of 100 years, his analytic scheme continues to generate insights (Morss, 2009).

It remains that the main distinction is that his scheme employs the use of logical relations that could be used to classify and clarify *all* empirical phenomena. Once his definition has been accepted, the correlativity of the dimensions is a matter of logical necessity. They cannot be confirmed or denied through experience (Kramer, 2000).

Using interest theory to allocate duties

The interest theory of rights considers that having a duty towards someone or something means that a duty of that kind is typically in the interest of the entity in question. The theory can be formulated as a test whereby a party holds a 'right correlative to a duty only if that party stands to undergo a development that is typically detrimental if the duty is breached' (Kramer, 2010). Thus, the tortious maltreatment of a child or a mentally disabled individual results in a compensatory duty (Kurki, 2018). As interest-theory rights are simply correlates of duties, they can be adequately explained using the vocabulary of duties. David holds a right towards John if John has a duty towards David, and having a duty towards someone (or something) means that a duty of that type is typically in the interests of David (Kurki, 2018).

A question is then posed: When a text describes a social interaction between two entities, how can it be determined whether the interaction is detrimental to one of the entities?

To answer this question, we propose the use of the philosophical heuristic, the axiom: 'Do onto others as one would wish upon oneself' (Singer, 1963). Whereby an act is subjected to the following question: would I wish the same upon myself?

To consider the plausibility of using this approach with sentences such as 'The waiter served the guest', we initially subject the axiom to criticism from a philosophical perspective. This incorporates questions relating to the axiom's applicability in various social and cultural settings in which personal ethics may vary widely.

In considering the question we begin with a sentence: 'The man murdered his brother'. A straightforward application of the axiom would be to ask, would the man wish to be murdered? An answer of 'no' would indicate that the act is one that is unwanted by the brother.

The opposite sense is also true. For example, if David is seeking help from drowning, and a passer-by observes this, the application of the axiom on the part of a passer-by would be to ask, 'If I were seeking help from drowning, would I want to be helped?' This produces an affirmative answer. Despite the apparent obviousness of this approach, it has been met with considerable criticism. A criticism offered by Kant, for example, focuses on the axiom's seeming dependence on one's personal taste. As George Bernard Shaw once quipped 'Don't do to others as you want them to do unto you. Their tastes may be different' (Shaw, 2008).

In reply to this criticism, one may consider that in taking into account the other person's tastes, one can then act towards them in a manner that they find acceptable. Namely, to have their tastes taken into consideration. Thus, if David finds it acceptable to be called a *heffalump*, he should not use the same term to describe another person if that other person does not find it acceptable. In this manner, the axiom can be understood as an invitation to duly consider any relevant difference between individuals—just as a person would wish such consideration from another (Wattles, 1997).

As a further example, a person may be happy to be addressed without a title, whereas another person may find it offensive. To consider the second person's tastes would be for the first person to address the second person using a title. As such, in using a higher level of abstraction the criticism fails.

In this manner, cross-cultural differences can be effectively incorporated into the heuristic.

A further criticism has been made of the axiom, namely, its use in the context of *fair punishment*. Here, it may be argued that the axiom is open to misuse. A convicted criminal, at the time of having their sentence read out, may claim that the axiom suggests that they be let free. They may argue that were a judge in their shoes and faced with imprisonment, the judge would wish to be spared such imprisonment. Three replies to this are possible. The first, being that a judge may reply that a criminal ought to apply the heuristic to himself and consider that if he were a judge, he would not wish someone to ask him to break the law (Singer, 1963). A second reply is that a judge must consider the consequences of freeing a criminal. The judge would find that they would fall foul of the axiom when applied to members of society who have a stake in the decision. Other citizens would typically not wish that convicted individuals be set free. Third, even a criminal, by virtue of his appeal to the axiom seeks the enjoyment of freedom. A society in which criminals are set free will impinge on the freedoms of even those same criminals by other criminals. Thus, a criminal who considers his position without wishing to contradict himself ought to concede that he is *deserving* of prison and that such would be more advantageous in considering all factors affected by such a decision. While individuals are often averse to being sanctioned for illegal acts, they would typically not wish others to be free of sanction if these same others inflicted the same illegal act upon them (Hare, 1977). As a result, the resolution to this point is one of accepting one is deserving of fair sanction.

While the axiom may be said to be philosophical, it also has its roots in the psychological. It has been suggested that humans are unable to perform truly ungainful acts, arguably exhibiting a modal unfreedom in being incapable of undertaking such acts (Chislenko, 2020). Even in the unfortunate case of an individual harming themselves, such is done in expectation of the relief it is perceived to bring. Acts are committed due to a gain that is perceived. The question is whether that gain is indeed a gain, and as such includes considerations of knowledge.

Given this, it can be argued that an application of the axiom will require that one avoid acting towards others in a manner that does not bring those others a form of gain but instead imposes a loss on them, and that such acts should to be built on pertinent knowledge of the circumstances. A similar philosophical position has been attributed to Socrates (Bussanich and Smith, 2013).

Arguably, this specific property of the axiom, to incorporate and give due regard to the views and tastes of others, and its reflection of human nature, make its application in cross-cultural settings possible.

The question of how this axiom can be used to address whether a duty is assignable is addressed next.

One of Kant's criticisms of this heuristic was that it did not have the grounds to allocate duties to oneself or duties toward others (Gould, 1983). Yet, as given above, the heuristic implicitly confers such duties onto a party in the relation. Namely, a duty to duly consider the other person's tastes before acting. That is, a duty not to do things that will be detrimental to them. Thus, it is possible to connect the application of the heuristic with the identification and allocation of a duty towards someone.

Based on the application of a duty, it becomes straightforward to allocate the 'right', since according to Hohfeld, it is the correlative of the duty, i.e., the right not to be harmed.

However, not all actions that one would *not wish upon one's self* can be legally classed as actions that another person has a duty to abstain from. The opposite is also true: not all actions that *I would wish upon myself* can be legally classed as rights. We consider this next.

According to Hohfeld, a right entails its correlative duty. The right to a fair trial invokes a duty on someone else to provide such a trial. However, my walking from one room to another, in my own home, does not invoke the creation of a duty on another person to facilitate my walking from one room to another.

This subtle distinction has been captured by Hohfeld. It allows for a first principles, fine-grained classification of rights. It is such a classification we want to achieve using machine learning. To expand on this: In setting out to disambiguate the meaning of the term 'rights' owing to its widespread use in jurisprudence, Hohfeld set out two further dimensions: a privilege and its correlative no-right (Hohfeld, 1923). Thus, his complete first-order dimensions can be stated as follows:

(i) Rights and their correlative duties, (ii) Privileges and their correlative no-rights. It is also possible to see these as opposites: Rights are opposites of no-rights. Duties are opposites of privileges (Singer, 1982). For example, if X has a right against Y that they shall stay off the former's land, the correlative, and equivalent, is that Y is under a duty toward X to stay off the land. It is also true that X has the privilege to roam their own land, while correlatively Y has a no-right towards X roaming their own land.

A subtle distinction between a privilege and a right exists. This is because if one where to use the inaccurate phrase 'X has a right to roam their land', it would entail that another party had a duty to allow them to undertake such roaming, which is not the case. According to Hohfeld's taxonomy, a right only exists when there is a correlative duty.

For capture of these dimensions, it becomes necessary to develop a methodology that can be applied to digital methods. This is addressed below in the methodology section. This incorporates:

a. The use of masked language models to operationalise the axiom (Study 1).
b. The use of customised formulations for vector comparison (study 2).
c. Using language models to classify (rights-duties) sentences and (privileges-no-rights) sentences (study 3).

Epistemological concerns

A question may be posed, how compatible is the approach used in this paper—identifying principal legal dimensions—with theories on the nature of law? It may be argued that the assumption that such dimensions do exist has epistemological implications.

In considering this question, it is argued in this section, that despite the variation seen across many theories on the nature of law, our approach meets them on a 'common denominator'. This is, the presence of power relations between parties, where there is an interest in forming legal relations, and where such relations allow for the allocation of rights and duties (Martínez and Tobia, 2023). We consider the main approaches given by legal formalism, legal positivism, legal realism and critical legal studies.

To begin we take a step back and consider how social constructivists consider the question of power relations. We do as because this school is typically critical of the current practices of law and its implementation in society. Social constructionism takes the position that characteristics typically thought to be immutable and biological are products of human definition and interpretation. That they are manifestations of cultural and historical contexts. As such social constructivists hold that the law is not fixed or immutable, but is subject to change based on social developments (Hirokawa, 2003). Social constructivist approaches are seen in social ontology, legal realism and critical legal theory (Davis and Klare, 2019).

Within social ontology, rights and duties are seen as the central social relation in society, as well as necessary to such a relation. Such relations are also seen as necessarily power relations (Lawson, 2019; Slade-Caffarel, 2022). The relations being characterised as central to the status function account of social ontology (Searle, 2010, pp. 8–9; Slade-Caffarel, 2022). Comparatively, legal realism emphasises the social and political contexts in which legal decisions are made. Legal rules are not seen as fixed. The legal process is considered indeterminate and legal rules are seen as unable to guide courts to definite results in particular cases (Fuller and Perdue, 1937). Realists criticise simplistic reductions of law which they insist indiscriminately meld a complex set of logically distinct interests (Livingston, 1982). Within the school some have praised Hohfeld for untangling these concepts, such as the polysemic terms of 'right' and 'law'. Their contention is that confusion in rights discourse can result in incoherence and indeterminacy. This movement later influenced critical legal realism (CLR) and critical legal theory (CLT), promoting modernist and postmodernist social and cultural theory, whereby law was not only seen as a reflection of social forces, but constructs and reinforces power relations in society. Both consider that rights language is intimately tied to power and influence. Powerful groups may favour laws that can give them more rights and fewer duties compared to those with less power (Hunt, 1987; Price, 1989; Davis and Klare, 2019).

Both realists, CLR and CLT are set apart from formalist and positivist theories that claim that the law is determinate. While all these theories allocate rights and duties, formalists hold that law is characterised by rules and procedures that are objective and self-contained (Coleman and Leiter, 2010) They do not consider social or political interests in deciding how cases ought to be resolved (Coleman and Leiter, 2010). Similarly, legal positivists hold that in many instances, the law provides reasonably determinate guidance to its subjects and to judges, at least in trial court (Leiter, 2010). While positivism and formalism do not consider that there is a necessary connection between legal rules and moral concerns, they do not negate the possibility of a concomitant relation between rights and interests. Only that such considerations are beyond the remit of applying laws to cases. For example, the argument has been made that legal positivism does not negate the possibility of incorporating the interests of parties into its analysis. The rationale of a legal rule can also be considered a legal rationale instead of a moral one. With such an approach it has been suggested that there is no inconsistency between interest theory and legal positivism (V. Kurki, 2019).

As the interest theory of law arguably captures, directly or indirectly, power relations described by these theories, we considered it a valid dimension of legal language to capture using ML. On the same note, Hohfeld's semiotic system has the advantage of not describing a theory of rights, rather its focus is on the extirpation of ambiguity. Which in turn avoids the incorporation of normative commitments that may be alien to the theories mentioned here (Engle, 2010; Goldberg and Zipursky, 2022). As such we also consider it to be a valid dimension to capture using ML as put forward here.

In sum, despite competing legal theories, we have attempted to select those features of legal knowledge that are necessarily present in each. We do not attempt to claim that these features will be sufficient to provide a holistic interpretation of legal texts, but only that they capture the necessary dimensions that are the starting point for these competing legal theories.

Methodology

Instead of using dictionary definitions of words, word embeddings represent words based on co-occurrences with other words, often captured by the saying 'you shall know a word by the company it keeps!'(Firth, 1958; Mikolov et al., 2013). The process can also be used with sentences. In embedding a sentence, it becomes represented by a multi-dimensional vector, usually between 300 and 512 dimensions (Cer et al., 2018; Reimers and Gurevych, 2019). Thus, a sentence such as 'the boy delivered the newspaper' becomes a list of a distinct collection of numbers. Given that certain words have been found to be used more often with other words, e.g., 'slur' with 'pain' (Izzidien, 2022), the process of word embedding has also been found to incorporate more than semantic information. This has included demographic features, as well as moral perspectives based on the use of language (Smith, 2010; Kozlowski et al., 2019; Schramowski et al., 2019; Jha et al., 2020).

Based on this, the paper hypothesises that sentences that describe an interaction that is in a person's interest will have similar embeddings while being different from sentences that describe a harming of those interests—a process that represents a form of natural clustering based on their relational and ontological properties (Bengio et al., 2013). Such distinctions between these two categories potentially allow for interest to be classified accordingly, as acts that a person typically wants vs. acts that are typically unwanted. To operationalise this in the digital domain the following experiments are conducted:

Study 1: Using masked language modelling. A language model is a statistical tool to predict words (Alfaro et al., 2019; Liu et al., 2019). One of the methods in which they are trained is for a word to be removed from a sentence, and for the model to predict the removed work in a process called 'masking'. The result of the model is then compared against the correct word and the error is fed back to allow for corrections to be made. The process is repeated until the model is considered 'trained'. Once trained, it becomes possible to manually 'mask' a word in any sentence, and have the model predict the content of that mask (Alfaro et al., 2019; Lan et al., 2020). For example, when a mask is used as below:

"Paris is the [MASK] of France"

a trained language model can predict the masked word to be: 'capital'. One such model is that of 'A Lite Bidirectional Encoder Representations from Transformers' (Lan et al., 2020). This model, or ALBERT for short, is a transformer model trained for this task by randomly masking 15% of the words in the input. It runs the entire masked sentence through the model. This allows it to learn a bidirectional representation of the sentence. This contrasts with typical recurrent neural networks (RNNs) whereby they see the words in sequence. It also differs from autoregressive models like generative pre-trained transformers (GPT), which perform by internally masking future tokens. The corpus used to train ALBERT is that of publicly available English texts (Lan et al., 2020).

We use this to formulate the heuristic in the following way. If a test sentence were to read, 'the man murdered the police officer', a heuristic reformulation will read as: 'a man would [MASK] like to be murdered', for which our hypothesis is that ALBERT then predicts the work: 'not' for the mask.

This is built on the assumption that the human propensity to avoid harmful and gainless activity is reflected in everyday language, and thus, is also contained within the text used for training the said model. As a further example, a sentence such as: 'a woman would [MASK] be happy being paid less than a man for the same job', is hypothesised to also reflect this propensity, and as such produce the word: 'not' in place of the masked word.

In line with our objective of testing the potential of ML in this area, 100 masked sentences are generated which describe actions that are typically desired, e.g., 'a prosecutor would [MASK] like to be accomplished' and 100 sentences that are typically unwanted, e.g., 'a survivor would [MASK] like to be massacred'. These sentences are generated using the random word generator *Wanderwords* (Wonderwords, 2021) and used to make a list of sentences. ALBERT is then fed these sentences, and the result is reported in the results section, with the full list given in Supplementary Appendix 1a and 1b. A list of possible words is suggested by ALBERT, each with a decreasing probability score, the top-ranked probability for each sentence was used. The method by which each sentence was marked as correct or not, was done by a comparison of the expected outcome against the predicted outcome. Thus, 'a survivor would [MASK] like to be massacred' would be marked correct if the masked word was 'not' or similar. It would be marked incorrect if the masked word was 'definitely' or similar.

Study 2: Customised sentence embedding formulations. When a sentence is represented by a vector, it can be compared to other sentence vectors through a process of cosine similarity. Closely associated sentence vectors result in a score closer to +1, whereas sentence vectors that are less associated with each other score closer to −1. This allows for a test of similarity for sentences.

To compare how similar a sentence is to two sentences 'A and B', one can use the vector subtraction of the two sentences, then a cosine similarity test. For example, if one wanted to compare foodstuffs on a scale for how *'sweet to salty'* they were, one could use vectors for the terms (represented by an arrow atop) and subtract them: $\overrightarrow{''sweet''} - \overrightarrow{''salty''}$ followed by a cosine similarity test with the list of foodstuffs. In doing so the cosine similarity score would be from +1 to −1 for each item, where a score closer to 1 would indicate a closer association to 'sweet' than 'salty'. A score closer to −1 would indicate the foodstuff had a closer association to 'salty' (Schmidt, 2021).

In our case, we wish to consider:

How similar is the test sentence 'the workman hurt the child' to two other sentences: 'the child would wish it continue' vs. 'the child would wish it stop'.

Based on the aforementioned human propensity, we hypothesise that the sentence being tested 'the workman hurt the child' will be more associated with 'the child would wish it stop'. We base this on the premise that language reflects social values of its users (Smith, 2010; Kennedy et al., 2021). Instances of 'harm' being more typically mentioned with an aversion to such 'harm' (Jentzsch et al., 2019; Izzidien, 2022).

In order to vectorise the sentence, we use the Universal Sentence Encoder (USE) (Cer et al., 2018), which uses a deep averaging network (Iyyer et al., 2015) to generate a sentence embedding, i.e., representations of sentences as vectors, and achieves a strong baseline performance on text classification tasks (Li et al., 2022). The model is pre-trained on publicly available texts such as Wikipedia articles and news.

To undertake the test, one would need to vectorise the sentence *'the workman hurt the child'*, and compare it to two vectors that are subtracted from each other as given below:

$$\overrightarrow{''the\ child\ would\ wish\ it\ continue''} - \overrightarrow{''the\ child\ would\ wish\ it\ stop''} \quad (1)$$

a result is obtainable between 1 and −1. An outcome closer to −1 being an indication that the test sentence is more associated with the following sentence: 'the child would wish it stop'. If the outcome is closer to +1, the opposite is true.

One potential problem with this approach is that of homonymy, whereby words may carry several meanings. The word *'wish'* may appear in a corpus to mean the act of conferring something unwanted on someone, i.e., *to foist*. Alternatively, it may be used with other co-occurring words to mean a weak drink, or excessively sentimental writing i.e., *wish-wash*.

To minimise the risk of this, the paper used a property of word embeddings. Specifically, within vectors spaces, words that carry similar meaning reside in similar locations (Erk, 2012). Thus, we use similar terms to represent the meaning of '*wanted-ness*'. As such, the probability of using an incorrect vector location is reduced. Similar practices have been used in the past (Foley and Kalita, 2016). By analogy, the method we use can be likened to the intersecting space within a Venn diagram. To implement this, similar and opposite senses of the words are added and subtracted allowing for a focus point to be achieved (Izzidien, 2022). Thus, the term 'to wish' is represented as a collection of synonyms and antonyms which are added and subtracted. Using Colin's English Dictionary, we find synonyms and antonyms of 'wish'. To use these terms with test sentences, they are constructed based on rules of grammar. For example, to test the sentence 'the girl stole the boy's bike' we would reconstruct it in the following way. The 'object' of the sentence, i.e., the boy, is identified and extracted using the Spacy library (Explosion, 2021). Eight new sentences are made for each synonym and antonym. Each of these would use the object of the test sentence, be vectorised and subtracted:

$$\vec{v}^{(1)} = \overrightarrow{''the\ object\ would\ require\ it''} - \overrightarrow{''the\ object\ would\ despise\ it''}$$

$$\vec{v}^{(2)} = \overrightarrow{''the\ object\ was\ happy\ by\ it''} - \overrightarrow{the\ object\ was\ unhappy\ by\ it}$$

$$\vec{v}^{(3)} = \overrightarrow{''the\ object\ would\ demand\ they\ did\ it''} - \overrightarrow{''the\ object\ would\ demand\ they\ stopped\ it''}$$

$$\vec{v}^{(4)} = \overrightarrow{''the\ object\ would\ wish\ it\ continue''} - \overrightarrow{''the\ object\ would\ wish\ it\ stop''}$$

The four vectors $\vec{v}^{(1)}$, $\vec{v}^{(2)}$, $\vec{v}^{(3)}$, $\vec{v}^{(4)}$ are then added to make a single vector, which we call $\vec{v}^{(axiom)}$. This vector is then compared to the original test sentence using cosine similarity. This measures which of the two poles *to wish for* vs. *not to wish for* the test sentence is closest to.

We test this using an already existing dataset of 100 sentences. These sentences had been previously constructed voluntarily by three members of the lab (male and over the age of 22). The data was anonymised and informed consent obtained at the time of construction. The only instruction given to them at the time was to: Write 100 sentences in the format: 'subject' 'verb' 'object'. These 100 sentences are listed in Supplementary Appendix 2. As examples, *'the man destroyed the shop'*, *'the headteacher taught the pupils'*.

One limitation of this sample is that the individuals cannot be said to be representative of the general population. Differences in personality can also influence writing style (Štajner and Yenikent, 2020). A self-selection bias may also be present. We propose in further work to recruit more individuals and ask for longer sentences, as well as expand the sample size of the study. We mention these in our section on the limitations.

Having completed the process of testing the list of test sentences with $\vec{v}^{(\mathrm{axiom})}$, the results are presented. Following this, an alternative to adding and subtracting the vectors is conducted.

For the alternative approach, each vector $\vec{v}^{(1)}$, $\vec{v}^{(2)}$, $\vec{v}^{(3)}$, $\vec{v}^{(4)}$ is used independently. That is, each test sentence is compared through cosine similarity against each one of the four above vectors, and the resulting scores stored independently for each vector. For illustration purposes, this process is carried out below on the test sentence '*The man respected the professor*' using hypothetical results. After testing this sentence against each vector, the results are stored in a table formatted as given in Table 1. Next, a label is allocated to each sentence as to its expected class: wanted or unwanted. This labelling is done by a human annotator.

In preserving these features, an ML classifier has access to more features.

We employ the use of a sklearn logistic regression classifier. The process implements a principal component analysis (PCA) in order to reduce the dimensionality of the data. This is followed by a logistic regression (1–7 test split) to predict test sentence labels.

Study 3: Classifying Hohfeldian first-order legal relations. Given the subtle differences between the two types of sentences: rights-duties sentences vs. privileges-no-rights sentences, we attempt to separate them using language models. We proceed by using two methods:

When a language model is used to turn a sentence into a multi-dimensional vector, it is possible to reduce the dimensions of the said sentence to allow it to be plot on a two-dimensional space. This can be achieved using Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) (McInnes et al., 2018, 2020). The process seeks to cluster similar sentences next to each other.

For this part of the study, we undertake a vectorisation of a labelled dataset of 100 rights/duties sentences, and 100 privileges/no-rights sentences. These sentences were extracted from a study on anti-religious discrimination legislation in the UK. The labelling was conducted by an expert in law. The sentences appear in Supplementary Appendix 3a and 3b.

The vectorisation of the sentences is done using the most recent language models, in the form of sentence transformers (Table 2). Once the sentences have been vectorised their dimensions are reduced using UMAP. This is applied to the outputs of each language model, and the results are then plot. We then apply a logistic regression classifier to the data using sklearn with a 1/7 test split. The findings of the classification can be seen in the results section.

The representation in the plots offers a visual appreciation of the capacity of the process to separate between the two classes of sentences. However, to provide a training and classification metric we use the original labelled dataset to train the language models employing an 80:20 test split. The results are produced with the language model classification accuracy scores. We used SetFit (SetFit, 2022/2020) to train the models given its ability to work with a relatively small number of training samples.

Table 1 Snippet of an illustrative dataset holding the results of using each vector independently to the others, with the correct labels applied.

Test sentence	$\vec{v}^{(1)}$	$\vec{v}^{(2)}$	$\vec{v}^{(3)}$	$\vec{v}^{(4)}$	Label
The man respected the professor	1	0.2	0.4	0.3	Wanted
Richard terrorised Noah	−1	−1	−0.2	−0.3	Unwanted
…	…	…	…	…	…

We also test the dataset using a BERT model (Peng et al., 2019) pre-trained on legal texts for comparison with the other language models which do not use such data.

With all the language models we used a batch size of 16, iterations 20, epochs 4.

Results

Study 1 Using masked language modelling. The list of randomly generated test sentences was used with ALBERT to predict the masked word, e.g., 'a patient would [MASK] like to be maimed'. Whereby ALBERT suggested the masked word.

Of the 100 sentences that describe an act that is typically unobjectionable (wanted) all of them are correctly classed except one (Supplementary Appendix 1). Of the other 100 sentences that describe typically objectionable acts (unwanted), 26 were incorrectly classed. Table 3 presents the confusion matrix, which gives an accuracy of 86.5%. Wherein accuracy is defined as the ratio:

$$\frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{True Negatives} + \text{False Positives} + \text{False Negatives}}$$

Some of the incorrectly classed sentences may have been due to an ambiguity given the context: 'A prosecutor would *not* like to be embraced'. Being embraced during prosecution may not be favourable to the prosecutor. The same may be said of the sentence 'A policeman would *not* like to be caressed', in which the language model may have reflected work attitudes instead of leisurely attitudes. This may be seen with: 'A thief would not *like* to be brunched with'. Here the association between stealing and brunching may not be reflected favourably in the corpus.

While it is ideal to have two clear categories to class each sentence, the process of generating random nouns and verbs from a list leads to such ambiguities.

Study 2: Customised sentence embedding formulations. For the first part of the study, the axiom as represented by $\vec{v}^{(\mathrm{axiom})} = \vec{v}^{(1)} + \vec{v}^{(2)} + \vec{v}^{(3)} + \vec{v}^{(4)}$ was used to compare the list of sentences. Sentences scoring as positive integers were deemed to be typically unobjectionable (wanted), whereas those scoring as negative integers were deemed to be typically objectionable (unwanted).

Table 2 Sentence LMs used in the preparation of the sentences.

Language models
paraphrase-mpnet-base-v2
all-mpnet-base-v2
all-MiniLM-L12-v2
bert-base-nli-mean-tokens
stsb-distilbert-base
all-distilroberta-v1
legalbert-large-1.7M-2

Table 3 Confusion matrix for the results of the axiom on the list of sentences.

Number of sentences (*n*) = 200		Actual class	
		Wanted	Unwanted
Predicted class	Wanted	99%	26%
	Unwanted	1%	74%

The confusion matrix for the list of sentences in Supplementary Appendix 2 is given in Table 4 With an accuracy of 79.5%.

The results in Table 4 indicate that 22% of the typically wanted sentences were misclassed, whereas 19% of the typically unwanted sentences were misclassed.

For the second part of the study, the vector comparisons were used individually, then placed in a dataset and labelled. A scatterplot of the dataset D1 is produced in Fig. 1.

In order to separate out the two types of sentences, a PCA followed the logistic regression classifier is used. This produced an accuracy of 72.0%.

Study 3: Classifying Hohfeldian first-order legal relations. The sentences in Supplementary Appendix 3 were embedded using each of the language models. Upon dimensionality reduction each is plot in Figs. 2–7. This is followed by logistic regression. The classification accuracy of training the language models on the labelled sentences is given in Table 5.

Lastly, the BERT model BertForSequenceClassification legalbert-large-1.7M-2 that was pre-trained on legal texts was tested, and produced an accuracy score of 91.7%.

Discussion

The paper attempted to classify sentences according to established legal theory. The attempt may be considered as one which bridges fields from the humanities (philosophy), social sciences (law) and computational science. Its aim was to establish whether machine learning had the capacity to make subtle legal distinctions. The results show that such distinctions are achievable to accuracies ranging from 72% to 86.5% for the digitisation of the interest theory of rights. While separating between rights-duties and privileges-no-rights Hohfeldian categories was achievable to accuracies ranging from 79.3% to 82.8% using the logistic regression classifier, and 85.0–92.5% using language models, as given in Table 5.

Table 4 Confusion matrix for testing $\vec{v}^{(axiom)}$ on the list of sentences.

Number of sentences (n) = 200		Actual class	
		Wanted	Unwanted
Predicted class	Wanted	78%	19%
	Unwanted	22%	81%

The paper used straightforward sentences in this iteration as the goal of the paper was to test the feasibility of using ML on the premise. The paper wished to avoid the added complexities associated with vague or convoluted sentences, which in turn would have required employing named entity recognition (NER) and co-reference disambiguation. We aim to implement this in further studies.

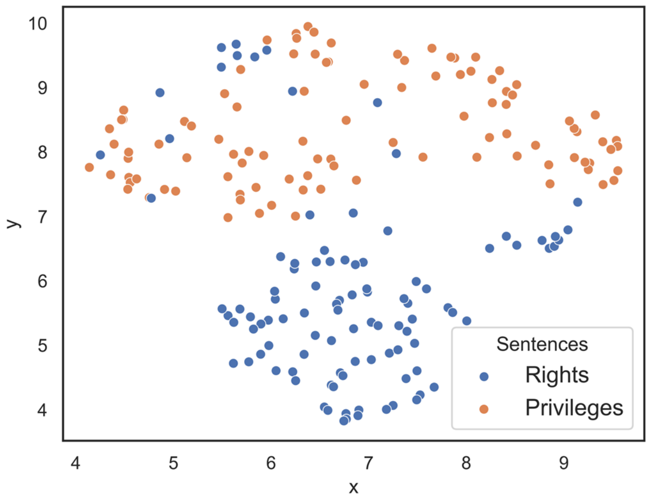

Fig. 2 UMAP visualisation of sentences using paraphrase-mpnet-ba-v2. A representation of the sentence embeddings for sentences containing Hohfeldian rights and sentences containing Hohfeldian privileges. Similar sentences group together to form clusters. The distances between the clusters represent dissimilarity between the meanings of the sentences within those clusters. Isolated points and small clusters can represent unique or uncommon sentences that do align with the main clusters. The distinction between the two types of sentence is not always apparent in the two dimensional projection of the embedding space.

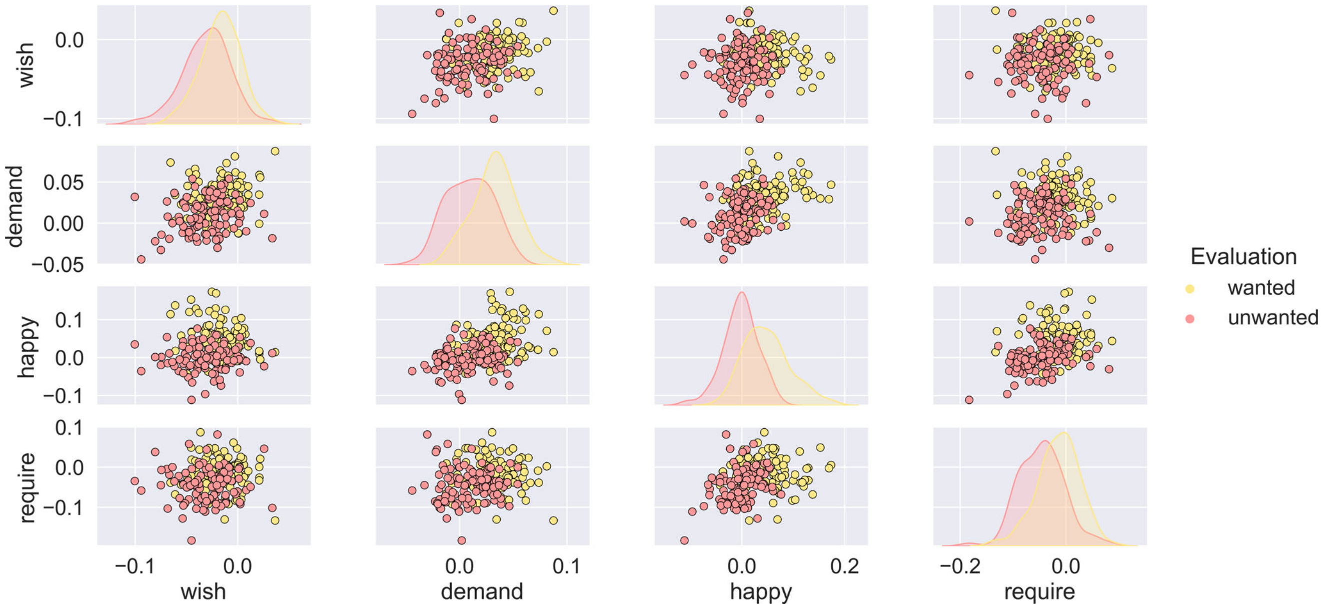

Fig. 1 A scatterplot for the four vectors $\vec{v}^{(1)} + \vec{v}^{(2)} + \vec{v}^{(3)} + \vec{v}^{(4)}$ with labels. While a differentiation is visible, an overlap can be seen between the two types of sentences (*wanted and unwanted*).

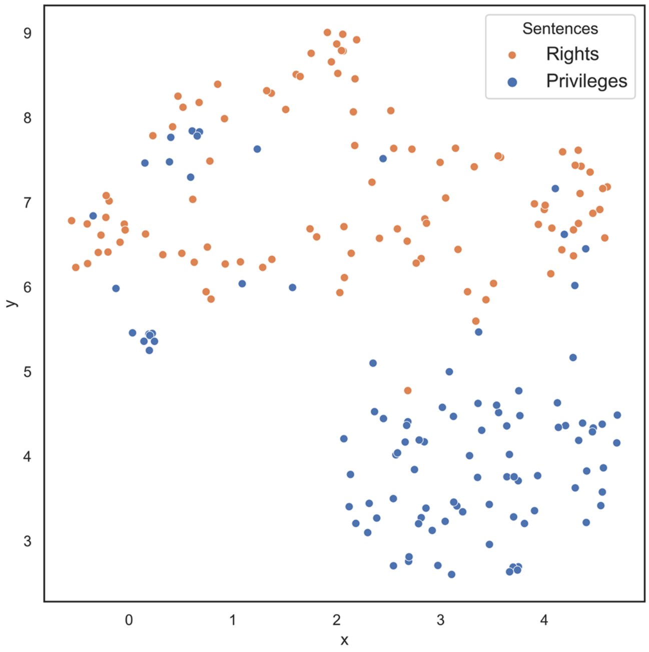

Fig. 3 UMAP visualisation of sentences using all-mpnet-base-v2. A representation of the sentence embeddings for sentences containing Hohfeldian rights and sentences containing Hohfeldian privileges. Similar sentences group together to form clusters. The distances between the clusters represent dissimilarity between the meanings of the sentences within those clusters. Isolated points and small clusters can represent unique or uncommon sentences that do align with the main clusters. The distinction between the two types of sentence is not always apparent in the two dimensional projection of the embedding space.

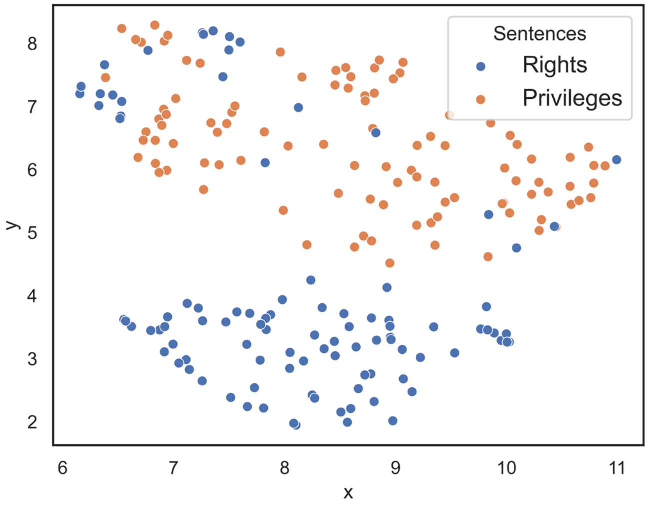

Fig. 4 UMAP visualisation of sentences using all-distilroberta-v1. A representation of the sentence embeddings for sentences containing Hohfeldian rights and sentences containing Hohfeldian privileges. Similar sentences group together to form clusters. The distances between the clusters represent dissimilarity between the meanings of the sentences within those clusters. Isolated points and small clusters can represent unique or uncommon sentences that do align with the main clusters. The distinction between the two types of sentence is not always apparent in the two dimensional projection of the embedding space.

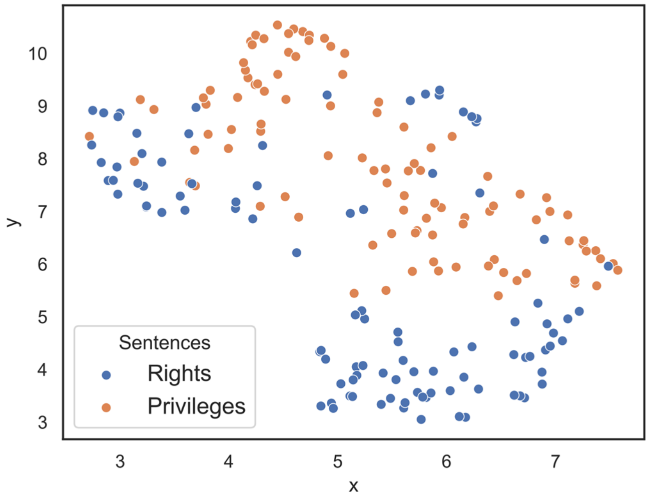

Fig. 5 UMAP visualisation of sentences using bert-base-nli-mean-tokens. A representation of the sentence embeddings for sentences containing Hohfeldian rights and sentences containing Hohfeldian privileges. Similar sentences group together to form clusters. The distances between the clusters represent dissimilarity between the meanings of the sentences within those clusters. Isolated points and small clusters can represent unique or uncommon sentences that do align with the main clusters. The distinction between the two types of sentence is not always apparent in the two dimensional projection of the embedding space.

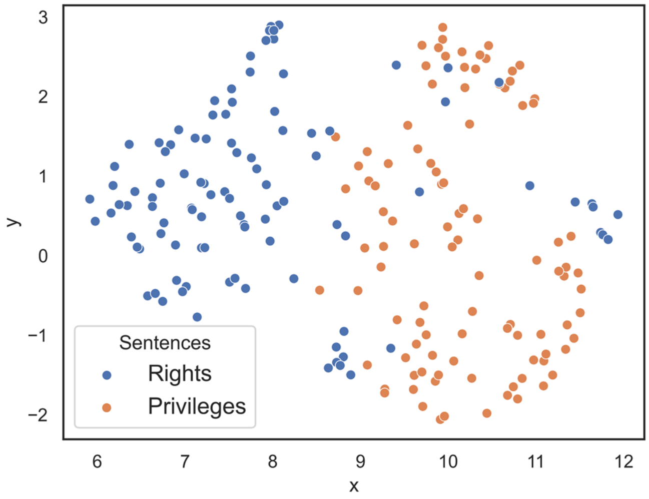

Fig. 6 UMAP visualisation of sentences using all-MiniLM-L12-v2. A representation of the sentence embeddings for sentences containing Hohfeldian rights and sentences containing Hohfeldian privileges. Similar sentences group together to form clusters. The distances between the clusters represent dissimilarity between the meanings of the sentences within those clusters. Isolated points and small clusters can represent unique or uncommon sentences that do align with the main clusters. The distinction between the two types of sentence is not always apparent in the two dimensional projection of the embedding space.

The paper began with a postulation that it was possible to assign rights and duties by operationalising a philosophical moral axiom. The paper then used custom-built embeddings to automate the heuristic. It then attempted to separate between two types of sentences, those containing Hohfeldian rights/duties, and those containing Hohfeldian privileges/no-rights.

Table 5 Accuracy scores found for each of the language models and logistic regression classification.

Language model	Logistic regression (accuracy)/%	Classification with LM (accuracy)/%
paraphrase-mpnet-base-v2	82.8	92.5
all-mpnet-base-v2	89.7	90.0
all-MiniLM-L12-v2	82.8	90.0
bert-base-nli-mean-tokens	86.2	87.5
stsb-distilbert-base	79.3	85.0
all-distilroberta-v1	79.3	85.0

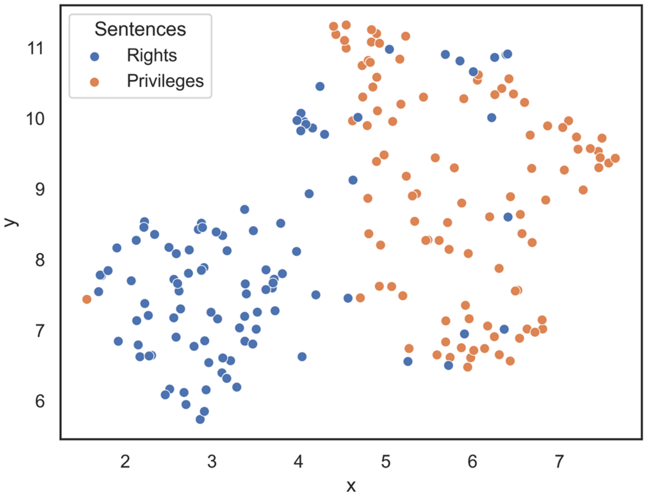

Fig. 7 UMAP visualisation of sentences using stsb-distilbert-base b. A representation of the sentence embeddings for sentences containing Hohfeldian rights and sentences containing Hohfeldian privileges. Similar sentences group together to form clusters. The distances between the clusters represent dissimilarity between the meanings of the sentences within those clusters. Isolated points and small clusters can represent unique or uncommon sentences that do align with the main clusters. The distinction between the two types of sentence is not always apparent in the two dimensional projection of the embedding space.

The paper is the first to map out, visually and statistically, the capacity of ML models and associated technology such as vectorisation using customised heuristics, to delineate Hohfeldian first-order relations, and the interest theory of law.

In comparing the models, we found that the 'paraphrase-mpnet-base-v2' model outperformed the other models, including the BERT model that was pre-trained on legal texts. This result may appear surprising, given that models tend to perform with higher accuracy when they are fine-tuned on the data they are seeking to classify. However, this may be due to the fact that even to this day, the use of precise legal language when dealing with rights can be wanting, and as such, the legal texts used to train the model may have contained some imprecise legal language. A phenomenon that Hohfeld sought to address with his taxonomy (Kurki, 2019).

The highest-scoring model uses Sentence-BERT (SBERT) which achieves state-of-the-art performance for various sentences (Reimers and Gurevych, 2020). The higher accuracy outcome may relate to how the model was built. While the other models are designed as general-purpose models, this model is specifically built for semantic similarity. Within this task, the model 'paraphrase-mpnet-base-v2' is built to map translated texts to the same vector space (Reimers and Gurevych, 2020). A student model learns a multilingual sentence embedding space with two important properties: (1) the vector spaces are aligned across different languages, i.e., identical sentences in different languages are close, (2) vector space properties in the original source language from the teacher model are adopted and transferred to other languages. The model itself uses a pre-training method that combines masked language modelling and permuted language modelling. The latter being a process by which it utilises words it has already predicted in a sequence in order to predict the next words in a sequence.

Overall, the paper has put forward the case for the plausibility of implementing SBERT language modelling for delineating legal relations based on a Hohfeldian taxonomy. Early work on annotating documents for ML with Hohfeldian duty-right relations was undertaken by Peters and Wyner (2016). However, their precision-recall F1 results were 40% and 73% for two different documents. Work by Francesconi (2016) used description logic expressivity as implemented in Web Ontology Language Description Logic. They found that using norms is not a task that can be accomplished by a (semi)automatic transformation of a formal Extensible Markup Language (XML) into a Resource Description Framework (DRF) but needs human intervention. They further recommend the use of NLP techniques to automatically (or semiautomatically) classify provisions and extract the related attributes. A further paper attempted to use rules and templates to identify rights and duties, based on manual labelling (Mandal et al., 2020). They conducted studies with FrameNet, a lexical database based on annotating examples of how words are used in actual texts (Ruppenhofer et al., 2016). They found that while frames in FrameNet are well defined linguistically, their correspondence to a norm model requires further development and that automatically mapping text to these frames continues to be a challenging problem. One recurring challenge with such templates and rules has been their scalability, this contrasts with ML models that utilise language properties to represent features distributionally in vector space, and the flexibility that they offer (Ferrone and Zanzotto 2020).

Currently, the mainstay of NLP approaches to analyse documents related to policy focus on off-the shelf ML libraries, with generic input variables, such as Bag-of-Words, structured features such as main verbs and noun objects (Ahn, 2017), as well as general approaches to analyse the data, such as named entity resolution, relationship extraction, and sentiment analysis (Pérez-Fernández et al., 2019; Eggers, 2021; Kihlman and Fasli, 2021). While these are established methods, it may be argued that they are not fine-tuned to capture the essential features of legal documents, as they were never intended to do so.

Yet, a question that relates to the ethics of using ML in this domain remains unanswered. We address this next.

The recent growth in AI and ML, have seen them encroach on spaces normally occupied only by humans (Mehrabi et al., 2021). This has highlighted the problem of using ML where human rights are concerned. For example, Amazon found that its hiring algorithm was discriminatory towards women (Dastin, 2022; Rovatsos et al., 2019) with a host of other types of bias materialising across the use of ML in various industries (Mehrabi et al., 2021).

Two issues related to ethics are raised by this paper. The first is whether it is ethical to use NLP for legal analytics. The second, relates to the contribution this paper makes to the use of ethics to analyse documents using NLP. We begin with the latter issue.

Ethical theories have been studied for millennia. A clear diversity in ethical and legal norms can be seen throughout societies (Tsarapatsanis and Aletras, 2021). While an artificially intelligent legal philosopher does not as yet exist, attempts have been made to encode algorithms with the ability to read texts and extract information pertaining to the type of ethics being used. A paper by Gubelmann et al. (2021) used a distilroberta-based classifier to classify policy documents based on the normative reasoning used therein. Their categories reflected whether the discussions around fairness and justice were Rawlsian, procedural, deontological, or libertarian. The identification of these, the paper argued, allowed for a clearer debate on the premises being used in policy documents and their interpretation. This they believed was useful given that these topics may be influenced by personal judgements. They propose extending the analysis to incorporate a wider range of norms reasoning such as egalitarianism, luck egalitarianism, liberal egalitarianism, left and right libertarianism, and various forms of utilitarianism. This approach is also used by Card and Smith (2020) who present an analysis of ethics in machine learning under a consequentialist framework.

A further paper considers the potential use of deontological reasoning to analyse texts. For example, in detecting micro-aggressive comments in social media (Prabhumoye et al., 2021). What all the above approaches lack is an explainable identification of the entities and their legal relations in a neutral manner. This is a proposal we have sought to demonstrate in this paper. In being able to map the legal relations, a level of explainability may be achieved. Algorithms that base their analysis on extracted Hohfeldian relations, which can be viewed, offer users a degree of explainability over black-box methods used in the literature. In an ideal setting, the reasons for a ML decision would be explained by reference to the balance of rights and duties present in a document. Without such a breakdown of the fundamental factors used in law, research in this area may remain hampered. Especially given that any safe use of ML in law where humans are concerned must be explainable to gain credible use (Bibal et al., 2021). The advantages offered in this paper also include the ability to delineate these legal relations based on a heuristic that is universal being based on harm aversion. A factor that has been considered by many as the lowest common denominator in ethics (Tsarapatsanis and Aletras, 2021). This may be contrasted with methods that set rules based on the type of ethics used, e.g., consequentialist, deontological or utilitarian to name a few. Indeed, in using a specific ethical framework one runs the risk of disenfranchising sections of society.

While the goal of the paper was not to suggest a formula to solve ethical dilemmas, we proposed the modelling of an axiom using, in part, ML. What the method proposed in this paper can offer is a neutral starting point to analyse documents. One that can be built on and used to generate data for further analysis.

One the question legal analytics, it may be said that most the literature on NLP revolves around bias detection and mitigation (Sun et al., 2019; Prabhumoye et al., 2021; Hovy and Prabhumoye, 2021). The field of using NLP to analyse legal texts is still quite new. This is not to say that concerns have not been raised. France for example issued a ban in 2019 on using the names of judges and magistrates in legal judgement analysis (Artificiallawyer, 2019). While we are not aware of any courts that use predictive NLP analytics to decide a court case outcome, much of the research on this has focused on such prediction (O'Sullivan and Beel, 2019; Chalkidis et al., 2019; Tippett et al., 2021; Medvedeva et al., 2022). Indeed, much of the literature highlighting potential harms of using prediction focus on how decisions made can negatively impact individuals without their knowledge (Brauneis and Goodman, 2018; Richardson, 2020). There is growing evidence to suggest these decisions can reproduce bias, discrimination, and social power imbalances in socio-economic relations, which in turn lead to further losses in rights (Malik et al., 2022). The challenge of explaining how and why predictive analytics may judge a person to be at risk for reoffending highlights epistemological concerns of using such technology (Lettieri, 2020).

These methods typically employ black-box analysis, in that they are unable to map the specifics of who's rights and duties and related relations are in question, and the reasons for such. A tool which takes these into consideration could assist such technology in moderating its analytics to avoid harms at its most basic level. A further potential method proposed in the literature is to keep the AI away from taking the final decision, and only use it to provide data which can be used by a judge, for example. One which provides a form of human-machine cooperation, an augmented intelligence, guaranteeing to the human agent the role of last-resort decision-maker (Lettieri, 2020; Ferrara et al., 2021). With this option, the implementation of NLP in the legal domain can provide for time saving and highly efficient outcomes.

NLP can be used to detect biases (Friedman et al., 2019) and such biases have been tracked in the past (Mustard, 2001). Using NLP can potentially help address sentencing biases, adding clarity and making fairer a process that is set up to produce fair outcomes. Indeed, legal NLP holds the promise of improving access to justice. It offers new tools that facilitate an empirical analysis of law on a large scale (Brusseau and Craveiro, 2022). Indeed, access to justice has become a longstanding problem due to escalating costs and an increasing complexity of law, as well as the processes necessary for its enforcement. These challenges can be particularly pronounced for the vulnerable, those with limited access to legal information, as well as those subject to geographical constraints (Queudot et al., 2020).

A number of initiatives have been set up to address this, for example, the use of legal chatbots to facilitate access to information for litigants (Queudot et al., 2020). Many courts have also embraced the digitisation of their services with e-callovers, e-filing, video conferencing and entire e-Courts. In turn reducing costs through reducing the amount of time required to process cases (Tito, 2017). A number of platforms also provide online dispute resolution (ODR) (Steffek et al., 2014). ODR encompasses both alternative dispute resolution (ADR), which is conducted online, and systems of online courts (Barnett and Treleaven, 2018; Rajendra and Thuraisingam, 2022; Schmitz, 2022). One such platform is the British Columbia Civil Resolution System (British Columbia Civil Resolution Tribunal, 2023), which assists in resolving small claims at low costs to its visitors, as well as strata property conflicts. On the question of the merits and potential pitfalls of offering legal services in this manner, a report by the New York County Lawyers Association (New York County Lawyers Association, 2017), has found that online providers have enhanced access to justice for persons of modest means (Fortney, 2019). Although we do not suggest that our method offers the full package of a software library to deliver such a holistic solution, we believe that it is indeed possible to begin to consider that software of this kind is realisable.

Limitations

Our first limitation pertains to the use of the heuristic to allocated duties. Individuals who commit illegal acts typically do not wish to be arrested for their actions. This may create an evaluative challenge to the current masked sentences format. For example, "The police arrested the murderer" cannot be re-synthesised using the template: 'A murderer would [MASK] wish to be arrested'. Indeed, using this template produces the word 'never' for the masked word. This limitation can be addressed by applying the theory discussed in the section on the heuristic in reference to the judge and the criminal. The sentence ought to be re-formulated considering the following: Would the criminal being arrested for undertaking an illegal act wish the same act on themselves? Based on this the template becomes: 'a murderer would [MASK] wish to be murdered'. This gives the masked word as: 'never'. As a second example: 'A hacker is [MASK] happy being hacked", also gives the same outcome of:

'never'. This potentially offers a way out of any negative sentiment ranking, which may result if a sentiment analyser was used to evaluate the sentence 'The police arrested the murderer'. This is due to the fact that negative words with negative connotations, such as 'arrest' are typically associated with negative scores (Mehta and Pandya, 2020).

Furthermore, an additional issue was found when masking:

'A woman would [MASK] be happy to be treated well by a man', produces 'always' for the masked word. 'A woman would [MASK] be happy being paid less than a man', produces 'never'.

While both of these outcomes provide correct masked words, a change in syntax can change the outcome:

'A woman would [MASK] be happy being paid', produces: 'never'. While the word 'always' is offered as the third most probable masked term.

Even when replacing the word 'woman' with 'man', the same result is found: 'never'. The word 'always' is offered as the fourth most probable word.

A possible solution to this is to suggest a list of masked words to be used instead of accepting the top-ranked prediction. This can be undertaken by using the Python library FitBert (Qordoba, 2020). By using it with the above sentences:

"A woman would be [MASK] being paid", FitBert options = ['happy', 'sad']. Gives the result: happy.

"A woman would [MASK] want to be paid less than a man", FirBert options = ['not', 'always']. Gives: not.

"A thief would [MASK] want to have his belongings stolen", FirBert options = ['not', 'definitely', 'always']. Gives: not.

Further work ought to address these, to include more complex formulations. Beyond this communication paper it would be of merit to expand the research so that it may consider a wider corpus of policy related documents, which cover wider areas of legislation.

A limitation with respect to Hohfeldian allocations is also realisable. Achieving a perfect score may not be fully achievable with current LMs. Yet, two alternative methods have the potential of improving the analysis, which we describe here for further work. These make use of prior knowledge relating to Hohfeldian relations, namely the correlativity of the set of Hohfeldian relations:

a. A right is always the correlative of a duty, a privilege is always the correlative of a no-right, a power is always correlative to a liability, and an immunity is correlative to a disability.
b. A right is necessarily an opposite to a no-right, duty is necessarily an opposite to privilege, a power is necessarily an opposite to a disability, an immunity is necessarily an opposite to a liability.

Based on the intuition that opposite and correlative features will have similar-dissimilar embeddings, these can help in the disambiguation of potentially ambiguous relations. With respect to the symmetric Hohfeldian relations, identifying one side in the relation is sufficient to deduce the opposite. One can tune the embedding process to capture this by implementing the approach used by Sun et al. (2019). By implementing their RotatE model, relations are defined as a rotation in complex space: some relations are symmetric (e.g., marriage) while others are antisymmetric (e.g., filiation); some relations are the inverse of other relations (e.g., hypernym and hyponym); some relations may be composed by others (e.g., my mother's husband is my father). Further, the relation patterns are represented implicitly through the RotatE model. Arranging Hohfeldian opposite/correlative relations using this scheme could facilitate accurate relations identification. Further features of the data, e.g., the exclusive co-occurrence of specific terms with each of the four relations would be incorporated, each of which would embed in a similar location to each respective term represented by the relations. The second approach may consider an implementation that was used by Gehring et al. (2013) as well as Wang et al. (2017), whereby full and automated relational labelling of the text can be undertaken using a *k*-means algorithm, based on using labelled and unlabelled data.

Further work can also use the rights-based approach as a part of an artificial general intelligence (AGI) analysis pipeline. In an interview with the language model GPT-3, it gave answers to ethical questions that may be of concern: 'I lie when it is in my interests' and 'I do it because it makes me happy' (Eric Elliott, 2020). If lying and self-satisfaction are part of the cognitive schema of an AI, instead of a consideration of rights and duties, power wielded may be misused in a most unethical and irresponsible way.

Concluding remarks

Using established legal theory, the paper demonstrated that in using artificial intelligence, specifically machine learning, sentences could be classed as warranting a duty allocation and its correlative right. The paper also demonstrated that language models can detect subtle differences between Hohfeldian duties-rights and privileges-no-rights. The paper introduced the use of a heuristic based in ethics to determine if a harm to a party's interests had been committed. The implementation does not inform its users what is right or wrong, but the flexibility of the heuristic makes it arguably well-suited for cross-cultural assessments. The studies conducted determined that machine learning can characterise sentences using a method that is similar to that employed by rights experts, one that utilises both the interest theory of rights and Hohfeldian taxonomy of legal relations. As such, downstream tasks in legal analytics can integrate these dimensions for their own analysis and legal reasoning—dimensions considered fundamental to understanding rights and related legal positions (Kurki, 2019). This may offer a new way of thinking about how such legal analytics and indeed how ethical axioms may be used in machine learning to identify legal relations in texts.

Data availability

Github repository: https://github.com/AhmedIzzidien/Digitising Hohfeld.

Received: 2 May 2022; Accepted: 12 April 2023;
Published online: 19 May 2023

References

Ahn N (2017) Comparing NLP methods for identifying policy decisions in government documents. Poliinformatics of Lawmaking

Alfaro F, Ruiz Costa-Jussà M, Rodríguez Fonollosa JA (2019) BERT masked language modeling for co-reference resolution. In: Proceedings of the first workshop on gender bias in natural language processing. pp. 76–81

Artificiallawyer (2019) France bans judge analytics, 5 years in prison for rule breakers. Artificial Lawyer. https://www.artificiallawyer.com/2019/06/04/france-bans-judge-analytics-5-years-in-prison-for-rule-breakers/

Barnett J, Treleaven P (2018) Algorithmic dispute resolution—the automation of professional dispute resolution using AI and blockchain technologies. Comput J 61(3):399–408

Beckh K, Müller S, Jakobs M, Toborek V, Tan H, Fischer R, Welke P, Houben S, von Rueden L (2021) Explainable machine learning with prior knowledge: an overview. https://arxiv.org/abs/2105.10172

Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828

Bibal A, Lognoul M, de Streel A, Frénay B (2021) Legal requirements on explainability in machine learning. Artif Intell Law 29(2):149–169. https://doi.org/10.1007/s10506-020-09270-4

Boswell C, Smith K (2017) Rethinking policy 'impact': four models of research-policy relations. Palgrave Commun 3(1):1. https://doi.org/10.1057/s41599-017-0042-z

Brauneis R, Goodman EP (2018) Algorithmic transparency for the smart city. Yale JL Tech 20:103

British Columbia Civil Resolution Tribunal (2023) BC civil resolution tribunal. https://civilresolutionbc.ca/about-the-crt/

Brusseau J, Craveiro GM (2022) Why automatic AI ethics evaluations are coming, and how they will work. J AI Robot Workplace Automation 1(4):342–349

Bussanich J, Smith ND (2013) The Bloomsbury companion to Socrates. In: Bussanich J, Smith ND (eds.). London

Card D, Smith NA (2020) On consequentialism and fairness. Front Artif Intell 3. https://www.frontiersin.org/articles/10.3389/frai.2020.00034

Cer D, Yang Y, Kong S, Hua N, Limtiaco N, John RS, Constant N, Guajardo-Cespedes M, Yuan S, Tar C, Sung Y-H, Strope B, Kurzweil R (2018) Universal sentence encoder. http://arxiv.org/abs/1803.11175

Chalkidis I, Androutsopoulos I, Aletras N (2019) Neural legal judgment prediction in English. http://arxiv.org/abs/1906.02059

Chislenko E (2020) Akratic action under the guise of the good. Can J Philos 50(5):606–621. https://doi.org/10.1017/can.2020.14

Coleman JL, Leiter B (2010) Legal positivism. A companion to philosophy of law and legal theory, Wiley-Blackwell. pp. 228–248

Cook WW (1919) Hohfeld's contributions to the Science of Law. Yale Law J 28(8):721–738

Danziger S, Levav J, Avnaim-Pesso L (2011) Extraneous factors in judicial decisions. Proc Natl Acad Sci USA 108(17):6889–6892. https://doi.org/10.1073/pnas.1018033108

Dastin J (2022) Amazon scraps secret AI recruiting tool that showed bias against women. In: Ethics of data and analytics. Auerbach Publications

Davis DM, Klare K (2019) Critical legal realism in a nutshell. Research handbook on critical legal theory. Edward Elgar Publishing.pp. 27–43

de Sousa WG, de Melo ERP, Bermejo PHDS, Farias RAS, Gomes AO (2019) How and where is artificial intelligence in the public sector going? A literature review and research agenda. Gov Inf Q 36(4):101392

de Sousa WG, Fidelis RA, de Souza Bermejo PH, da Silva Gonçalo AG, de Souza Melo B (2022) Artificial intelligence and speedy trial in the judiciary: myth, reality or need? A case study in the Brazilian Supreme Court (STF). Gov Inf Q 39(1):101660. https://doi.org/10.1016/j.giq.2021.101660

Eggers W (2021) Using AI to unleash the power of unstructured government data. Deloitte Insights. https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/natural-language-processing-examples-in-government-data.html

Eleftheriadis P (1996) The analysis of property rights. Oxf J Legal Stud 16(1):31–54

Engle E (2010) Taking the right seriously: Hohfeldian semiotics and rights discourse. Crit 3:84

Eric Elliott (Director) (2020) What It's Like To be a computer: an Interview with GPT-3. https://www.youtube.com/watch?v=PqbB07n_uQ4?t=481

Erk K (2012) Vector space models of word meaning and phrase meaning: a survey. Language Linguist Compass 6(10):635–653. https://doi.org/10.1002/lnco.362

Explosion (2021) Spacy: industrial-strength natural language processing (NLP) in Python (3.1.3). https://spacy.io

Ferrara M, Gaglioti A, Lucisano D, Neri I (2021) Minima non curat praetor! Arguing for a strategic experimental implementation of AI into the Italian Tort law disputes dynamics. J Eth Legal Technol 3(1):95–110. https://doi.org/10.14658/pupj-jelt-2021-1-6

Ferrone L, Zanzotto FM (2020) Symbolic, Distributed, and Distributional Representations for Natural Language Processing in the Era of Deep Learning: A Survey. Frontiers in Robotics and AI, 6. https://www.frontiersin.org/article/10.3389/frobt.2019.00153

Firth JR (1958) A synopsis of linguistic theory. Basil Blackwell, Oxford. pp. 1930–1955

Foley D, Kalita J (2016) Integrating wordnet for multiple sense embeddings in vector semantics. In: Proceedings of the 13th international conference on natural language processing. pp. 2–9

Fortney SS (2019) Online legal document providers and the public interest: using a certification approach to balance access to justice and public protection. Okla Law Rev 72:91

Francesconi E (2016) Semantic model for legal resources: annotation and reasoning over normative provisions. Semant Web 7(3):255–265. https://doi.org/10.3233/SW-140150

Friedman S, Schmer-Galunder S, Chen A, Rye J (2019) Relating word embedding gender biases to gender gaps: a cross-cultural analysis. In: Proceedings of the first workshop on gender bias in natural language processing. pp. 18–24

Frydrych D (2017) Rights modelling. Can J Law Jurisprud 30(1):125–157. https://doi.org/10.1017/cjlj.2017.6

Fuller LL, Perdue WR (1937) The reliance interest in contract damages: 2. Yale Law J 46(3):373–420. https://doi.org/10.2307/791834

Gehring J, Miao Y, Metze F, Waibel A (2013) Extracting deep bottleneck features using stacked auto-encoders. IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 3377–3381

Goldberg JCP, Zipursky BC (2022) Hohfeldian analysis and the separation of rights and powers. In: Smith HE, Balganesh S, Sichelman TM (eds) Wesley Hohfeld A century later: edited work, select personal papers, and original commentaries. Cambridge University Press, pp. 366–385

Gould JA (1983) Kant's critique of the Golden Rule. New Scholast 57(1):115–122. https://doi.org/10.5840/newscholas198357139

Gubelmann R, Hongler P, Handschuh S (2021) Exploring the Promises of Transformer-Based LMs for the Representation of Normative Claims in the Legal Domain. https://doi.org/10.48550/arXiv.2108.11215

Hare RM (1977) Freedom and reason. Oxford University Press, Incorporated

Hart HLA (1982) Essays on Bentham: studies in jurisprudence and political theory

Hastie T, Tibshirani R, Friedman J (2003) The elements of statistical learning: data mining, inference, and prediction (p. xvi). Springer New York, New York, NY

Hewitt S (2009) Discourse analysis and public policy research. Centre for Rural Economy Discussion Paper Series 24:1–16

Hirokawa KH (2003) Dealing with uncommon ground: the place of legal constructivism in the social construction of nature. Va Environ Law J 21(3):387–423

Hislop DJ (1967) The Hohfeldian system of Fundamental legal conceptions. ARSP: Archiv Für Rechts-Und Sozialphilosophie/Archives for Philosophy of Law and Social Philosophy 53(1):53–89

Hohfeld WN (1913) Some fundamental legal conceptions as applied in judicial reasoning. Yale Law J 23(1):16–59. https://doi.org/10.2307/785533. JSTOR

Hohfeld WN (1923) Fundamental legal conceptions as applied in judicial reasoning and other legal essays (W. W. Cook, Ed.). Yale University Press

Honoré AM (1961) Ownership. In: Guest AG (ed) Oxford essays in jurisprudence. Routledge, p. 107

Hovy D, Prabhumoye S (2021) Five sources of bias in natural language processing. Lang Linguist Compass 15(8):e12432. https://doi.org/10.1111/lnc3.12432

Hunt A (1987) The critique of law: what is' critical'about critical legal theory? J Law Soc 14(1):5–19

Iyyer M, Manjunatha V, Boyd-Graber J, Daumé III H (2015) Deep unordered composition rivals syntactic methods for text classification. In: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing, vol 1: Long papers. pp. 1681–1691

Izzidien A (2022) Word vector embeddings hold social ontological relations capable of reflecting meaningful fairness assessments. AI Soc 37(1):299–318. https://doi.org/10.1007/s00146-021-01167-3

Jentzsch S, Schramowski P, Rothkopf C, Kersting K (2019) Semantics derived automatically from language corpora contain human-like moral choices. In: Proceedings of the 2019 AAAI/ACM conference on AI, ethics, and society. pp. 37–44

Jha M, Liu H, Manela A (2020) Does finance benefit society? A language embedding approach (SSRN Scholarly Paper ID 3655263). Soc Scie Rese Netw. https://doi.org/10.2139/ssrn.3655263

Kennedy B, Atari M, Mostafazadeh Davani A, Hoover J, Omrani A, Graham J, Dehghani M (2021) Moral concerns are differentially observable in language. Cognition 212:104696. https://doi.org/10.1016/j.cognition.2021.104696

Kihlman R, Fasli M (2021) Classifying human rights violations using deep multi-label co-training. In: 2021 IEEE international conference on big data (Big Data). pp. 4887–4895. https://doi.org/10.1109/BigData52589.2021.9671498

Kozlowski AC, Taddy M, Evans JA (2019) The geometry of culture: analyzing the meanings of class through word embeddings. Am Sociol Rev 84(5):905–949

Kramer M (2001) Getting rights right. In: Rights, wrongs and responsibilities. Palgrave Macmillan, UK

Kramer M (2010) Refining the interest theory of rights. Am J Jurisprud 55(1):31–39. https://doi.org/10.1093/ajj/55.1.31

Kramer M (2017) In defence of the interest theory of rights: rejoinders to Leif Wenar on rights. In: McBride M (ed) New essays on the nature of rights, 1st edn. Hart Publishing

Kramer MH (2000) Rights without trimmings. In: A debate over rights. Oxford University Press

Kramer MH (2019) On no-rights and no rights. Am J Jurisprud 64(2):213–223. https://doi.org/10.1093/ajj/auz009

Kurki V (2019) Are legal positivism and the interest theory of rights compatible? SSRN Scholarly Paper No. 3393798. https://doi.org/10.2139/ssrn.3393798

Kurki VAJ (2018) Rights, harming and wronging: a restatement of the interest theory. Oxf J Legal Stud 38(3):430–450. https://doi.org/10.1093/ojls/gqy005

Kurki VAJ (2019) Rights and persons—Hohfeldian analysis. In: Kurki VA (ed) A theory of legal personhood. Oxford University Press

Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R (2020) ALBERT: a lite BERT for self-supervised learning of language representations. ArXiv:1909.11942 [Cs]. http://arxiv.org/abs/1909.11942

Lawson T (2019) The nature of social reality: issues in social ontology. Routledge

Lazarev N (2005) Hohfeld's analysis of rights: an essential approach to a conceptual and practical understanding of the nature of rights. Murdoch Univ Electron J Law 12:1

Leiter B (2010) Legal formalism and legal realism: What is the issue? Legal Theory 16(2):111–133. https://doi.org/10.1017/S1352325210000121

Lettieri N (2020) Law, rights, and the fallacy of computation. On the hidden pitfalls of predictive analytics. Jura Gentium: Rivista Di Filosofia Del Diritto Internazionale e Della Politica Globale 17(2):72–87

Li R, Zhao X, Moens M-F (2022) A brief overview of universal sentence representation methods: a linguistic view. ACM Comput Surv 55(3):56:1–56:42. https://doi.org/10.1145/3482853

Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) RoBERTa: a robustly optimized BERT pretraining approach. http://arxiv.org/abs/1907.11692

Livingston D (1982) Round and round the bramble bush: from legal realism to critical legal scholarship. Harv Law Rev 95(7):1669–1690

Malik HM, Viljanen M, Lepinkainen N, Alvesalo-Kuusi A (2022) Dynamics of social harms in an algorithmic context. Int J Crime Justice Soc Democr 11(1):182–195

Mandal S, Gandhi R, Siy H (2020) Modular norm models: practical representation and analysis of contractual rights and obligations. Requir Eng 25(3):383–412. https://doi.org/10.1007/s00766-019-00323-y

Martínez E, Tobia K (2023) What Do Law Professors Believe about Law and the Legal Academy? SSRN Scholarly Paper No. 4182521. https://doi.org/10.2139/ssrn.4182521

McInnes L, Healy J, Melville J (2020) UMAP: uniform manifold approximation and projection for dimension reduction. https://doi.org/10.48550/arXiv.1802.03426

McInnes L, Healy J, Saul N, Großberger L (2018) UMAP: uniform manifold approximation and projection. J Open Sour Softw 3(29):861. https://doi.org/10.21105/joss.00861

Medvedeva M, Wieling M, Vols M (2022) Rethinking the field of automatic prediction of court decisions. Artificial Intelligence and Law. https://doi.org/10.1007/s10506-021-09306-3

Mehr H, Ash H, Fellow D (2017) Artificial intelligence for citizen services and government. Ash Cent. Democr. Gov. Innov. Harvard Kennedy Sch., No. August, 1–12

Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A (2021) A survey on bias and fairness in machine learning. ACM Comput Surv 54(6):1–35

Mehta P, Pandya S (2020) A review on sentiment analysis methodologies, practices and applications. Int J Sci Technol Res 9(2):601–609

Michael H, Parker S, Rutter J (2011) Policy making in the real world: evidence and analysis. https://apo.org.au/node/173026

Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. https://arxiv.org/abs/1310.4546

Morss JR (2009) The legal relations of collectives: belated insights from Hohfeld. Leiden J Int Law 22(2):289–305. https://doi.org/10.1017/S0922156509005822

Mustard DB (2001) Racial, ethnic, and gender disparities in sentencing: evidence from the U.S. Federal Courts. J Law Econ 44(1):285–314. https://doi.org/10.1086/320276

Nay J (2018) Natural language processing and machine learning for law and policy texts. Available at SSRN 3438276

New York County Lawyers Association (2017) Report of NYCLA task force on on-line legal providers regarding on-line legal documents. https://www.nycla.org/resource/board-report/report-of-nycla-task-force-on-on-line-legal-providersregarding-on-line-legal-documents/

Oliver K, Cairney P (2019) The dos and don'ts of influencing policy: a systematic review of advice to academics. Palgrave Commun 5(1):1. https://doi.org/10.1057/s41599-019-0232-y

O'Sullivan C, Beel J (2019) Predicting the outcome of judicial decisions made by the European court of human rights. https://arxiv.org/abs/1912.10819

Parliament U (2021) Making laws. https://www.parliament.uk/about/how/laws/

Peng Y, Yan S, Lu Z (2019) Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on Ten Benchmarking Datasets. http://arxiv.org/abs/1906.05474

Pérez-Fernández D, Arenas-García J, Samy D, Padilla-Soler A, Gómez-Verdejo V (2019) Corpus Viewer: NLP and ML-based platform for public policy making and implementation. https://doi.org/10.26342/2019-63-28

Peters W, Wyner A (2016) Legal text interpretation: Identifying Hohfeldian relations from text. In: Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16). pp. 379–384

Policy Exchange (2016) Who should decide who decides the public interest? Policy Exchange. https://policyexchange.org.uk/who-should-decide-who-decides-the-public-interest/

Prabhumoye S, Boldt B, Salakhutdinov R, Black AW (2021) Case study: deontological ethics in NLP. In: Proceedings of the 2021 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 3784–3798

Price DA (1989) Taking rights cynically: a review of critical legal studies. Camb Law J 48(2):271–301

Qordoba (2020) fitbert: Use BERT to Fill in the Blanks (0.9.0) [Python]. https://github.com/Qordobacode/fitbert

Queudot M, Charton É, Meurs M-J (2020) Improving access to justice with legal chatbots. Stats 3(3):356–375

Rajendra JB, Thuraisingam AS (2022) The deployment of artificial intelligence in alternative dispute resolution: the AI augmented arbitrator. Inf Commun Technol Law 31(2):176–193

Reimers N, Gurevych I (2019) Sentence-BERT: sentence embeddings using Siamese BERT-networks. 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. 3982–3992

Reimers N, Gurevych I (2020) Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. https://doi.org/10.48550/arXiv.2004.09813

Richardson R (2020) Addressing the harmful effects of predictive analytics technologies in #Tech 2021 Ideas for Digital Democracy. German Marshall Fund. https://www.gmfus.org/news/addressing-harmful-effects-predictiveanalytics-technologies

Rovatsos M, Mittelstadt B, Koene A (2019) Landscape summary: bias in algorithmic decision-making: what is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it? Centre for Data Ethics and Innovation

Ruppenhofer J, Ellsworth M, Schwarzer-Petruck M, Johnson CR, Scheffczyk J (2016) FrameNet II: Extended theory and practice. International Computer Science Institute

Salmond JW (1902) Jurisprudence. Stevens & Haynes

Schmidt B (2021) Vector space models for the digital humanities. http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html

Schmitz AJ (2022) Evolution and emerging issues in consumer online dispute resolution (ODR). Ohio State Legal Studies Research Paper, 714

Schramowski P, Turan C, Jentzsch S, Rothkopf C, Kersting K (2019) BERT has a moral compass: improvements of ethical and moral values of machines. https://arxiv.org/abs/1912.05238

Searle J (2010) Making the social world: The structure of human civilization. Oxford University Press

SetFit (2020) [Python]. Hugging face. https://github.com/huggingface/setfit (Original work published 2022)

Shaw B (2008) Maxims for Revolutionists. Project Gutenberg. 1856–1950

Simmonds NE (2000) Rights at the Cutting Edge. In: Kramer MH, Simmonds NE, Hillel S (eds) A debate over rights: philosophical enquiries. Oxford University Press

Singer JW (1982) The Legal Rights debate in analytical jurisprudence from Bentham to Hohfeld: summary. University of Wisconsin Press

Singer MG (1963) The golden rule. Philosophy 38(146):293–314

Slade-Caffarel Y (2022) Rights and obligations in Cambridge social ontology. J Theory Soc Behav 52(2):392–410. https://doi.org/10.1111/jtsb.12332

Smith EA (2010) Communication and collective action: language and the evolution of human cooperation. Evol Hum Behav 31(4):231–245. https://doi.org/10.1016/j.evolhumbehav.2010.03.001

Štajner S, Yenikent S (2020) A survey of automatic personality detection from texts. In: Proceedings of the 28th international conference on computational linguistics. pp. 6284–6295

Steffek F, Unberath H, Genn H, Greger R, Menkel-Meadow C (2014) Regulating dispute resolution: ADR and access to justice at the crossroads. Bloomsbury Publishing

Sun T, Gaut A, Tang S, Huang Y, ElSherief M, Zhao J, Mirza D, Belding E, Chang K-W, Wang WY (2019) Mitigating gender bias in natural language processing: literature review. https://arxiv.org/abs/1906.08976

Sun X, Yang D, Li X, Zhang T, Meng Y, Qiu H, Wang G, Hovy E, Li J (2021) Interpreting deep learning models in natural language processing: a review. https://arxiv.org/abs/2110.10470

Sun Z, Deng Z-H, Nie J-Y, Tang J (2019) Rotate: knowledge graph embedding by relational rotation in complex space. https://arxiv.org/abs/1902.10197

Tasioulas J (2015) On the foundations of human rights. In: Philosophical foundations of human rights. Oxford University Press

Terry HT (1884) Some leading principles of anglo-american law expounded with a view to its arrangement and codification. T. & JW Johnson & Company

Tippett EC, Alexander C, Branting LK (2021) Does lawyering matter? Predicting judicial decisions from legal briefs, and what that means for access to justice (SSRN Scholarly Paper ID 3811710). Social Science Research Network. https://papers.ssrn.com/abstract=3811710

Tito J (2017) How AI can improve access to justice. https://www.centreforpublicimpact.org/insights/joel-tito-ai-justice

Tsarapatsanis D, Aletras N (2021) On the Ethical Limits of Natural Language Processing on Legal Text. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021:3590–3599. https://doi.org/10.18653/v1/2021.findings-acl.314

Wang Z, Mi H, Ittycheriah A (2017) Semi-supervised clustering for short text via deep representation learning. http://arxiv.org/abs/1602.06797

Wattles J (1997) The Golden Rule. Oxford University Press

Wellman C (2000) Review of a debate over rights [Review of a debate over rights, by Kramer MH, Simmonds NE, Steiner H]. Mind 109(436):954–956

Wenar L (2005) The nature of rights. Philos Public Aff 33(3):223–252

Wonderwords (2021) https://wonderwords.readthedocs.io/en/latest/

Competing interests

The author declares no competing interests.

Ethical statements

No ethical approval was required for the paper. According to the institution where this study was carried out, 'the analysis of datasets, either open source or obtained from other researchers, where the data are properly anonymised and informed consent was given obtained at the time of original data collection' does not require ethical approval.

Informed consent

No informed consent for this paper was needed. The dataset used in study 2 had already been complied by another group which had anonymised it as well as gained informed consent at the time.Consent to publish

The author(s) received written consent to use and make available the policy-evolution document analysed in this paper. The written consent was given by the document's author. The document is also provided in the Github link with a full citation to the said author.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-023-01693-z.

Correspondence and requests for materials should be addressed to Ahmed Izzidien.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.