Cautiously Making Friends with AI: Machine Learning for human rights research and practice

By Helga Molbæk-Steensig[1] and Alexandre Quemy[2]

Introduction

The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law(Holmes, 1897: 461)

Humans are fallible and prejudiced and tend to be grumpy when hungry. Hence human judges may be fallible and prejudiced and tend hand down harsher sentences if lunch is approaching and they haven't had a good breakfast. So have legal realists argued since the father of legal realism, Oliver Wendall Holmes first positioned his prediction theory of law(Holmes, 1897). Both Holmes and his European counterpart, Alf Ross (Ross, 2019[1948]) were of the notion that it should be possible to predict a case-outcome by studying the written law, but the realist scholars they inspired took a more cynical view of the justice system and the human judge; claiming that judges routinely base their judgments not on the law but on their intuitions and biases, resulting in a messy and unpredictable judicial system (Schauer, 2009). Furthermore, a judicial system populated by human judges tends to be slow and cumbersome. By way of example, at the European Court of Human Rights (ECtHR), the right to a fair trial (Article 6) is the most frequently violated right, and roughly a third of all applications contain a complaint about the length of proceedings (Edel, 2007). There is therefore good reason for human rights advocates to be concerned about the inadequacies of human judges.

The hope that Artificial Intelligence(AI) and Machine Learning (ML) applications might assist in overcoming the problem of inefficiency and bias has been around since computers first became commonplace. The idea being, that if judges are predictable mouths of the law, computers might be able to do their jobs faster, while if they are not, computers might be better at ignoring problematic human biases that have nothing to do with the law. Dystopian visions of futuristic robotic judges and pre-emptive crimefighting have, however, also been around for just as long. [3] Today, ML algorithms are employed as decision support systems(DSS) by police, judicial bodies, and administrations in numerous jurisdictions predominantly in the western world, and in the future they might also assist the ECtHR, which has its own problem with backlogged cases (Madsen and Spano, 2020: 179) and according to some commenters with predictability as well (Lord

[1] Helga Molbæk-Steensig is a PhD-researcher at the European University Institute's Law Department. Her dissertation uses a mixed methodology approach to investigate how the principle of subsidiarity works at the European Court of Human Rights both internally within the Court's caselaw and externally in its relationship with the High Contracting Parties.

[2] Alexandre Quemy holds a M.Sc. in Mathematics, from the National Institute of Applied Sciences (France) and a Ph.D. in Machine Learning with applications to the judicial domain, from Poznan University of Technology (Poland). His research focuses on automated machine learning (AutoML) with a special attention on explainability and bias. In addition to academic research, he works on real-life global-scale models, previously at IBM, now for the financial crime mitigation department of a large financial institution.

[3] For example: Minority Report by Philip K. Dick 1956

Lester of Herne Hill, 1998: 75, Zwart, 2013: 86). The contention emerges in the debate on what these DSS are in fact predicting and what data they base these predictions on. Depending on the data input and the parameters given to the AI, algorithmic decisions may be just as biased as human ones, may miss subtle but vital clues, or may be modelled on a world that no-longer exists. Moreover, they may rely on socio-economic data which would not generally be admissible in a court of law (O'Neil, 2016: 20-25).

Courts of all kinds, but especially international courts, rely on legitimacy for their power to make decisions. They do not wield power themselves but rely on others to execute their judgments. In the case of the ECtHR, execution relies on the very states that it may have just judged to have violated a human right. The source(s) of this legitimacy is a field of discussion in and of itself, but few commentators would likely disagree with current President of the ECtHR, Robert Spano, that important elements include accountability, transparency, equality and non-discrimination (English and Spano, 2020). While a carefully constructed DSS might improve the safeguarding of equality and non-discrimination from noisy human judgment, they cannot be held accountable for their predictions. With regards to transparency, the problem is worse yet. First, there are striking examples of ML algorithms used in national justice systems, which keep their source code a business secret, or which worry that defendants would change their survey answers if they understood how the AI works (O'Neil, 2016: 28-30). A widely-discussed example of this, is the COMPAS predictive sentencing system which was the subject of controversy in *State v Loomis* in Wyoming (HLR, 2017), but we also have examples from Europe such as the *SyRi* case from the Netherlands (Vervloesem, 2020). Secondly, even when the code is open source, the ability to understand it rests with a small, highly educated elite. An elite that, with very few exceptions, are not those judging in court.

In the European context we are therefore facing a problem in three parts. First, judicial systems all over the continent are overloaded with work and fail to provide fair trials within a reasonable time so often that cases on the right to a fair trial are overwhelming the ECtHR. As a result, the ECtHR struggles to provide closure in human rights cases within a reasonable time as well. Second, there is a growing body of evidence suggesting that judicial systems are hampered by human noise and bias, resulting in unpredictable and discriminatory judgments and decisions lacking one of the most fundamental ingredients of justice, namely equality before the law. Third, the current usage of ML in judicial and administrative systems which might have potential to ameliorate the first and second problem, has instead attracted scathing criticism for lack of transparency and accountability along with a growing body of literature providing examples of MLs that exacerbate systemic discrimination. On this theme, UN special rapporteur on Human Rights and Poverty, Phillip Alston, has warned governments against "…stumbling, zombie-like into a digital … dystopia [with private] technology companies operating in an almost human rights free zone…"(Alston, 2019).

In this paper we argue that ML algorithms *can* play a positive role in addressing the first and second problem without creating the third problem, but it requires narrowing the gap of understanding between data-scientists and legal scholars. This includes a much greater clarity in both camps regarding what it is believed justice requires of the judge (artificial, human or in combination), as well as a better understanding by legal professionals on what goes into modelbuilding including what computers are skilled at, and what humans are better suited for. This paper presents a path for clarifying the division of labour between the human and artificial judge through a theoretical contribution utilising

Dworkin's invention of the perfect judge, Hercules. This is done with the assumption that if we want to build a better judge, we first must agree on what the perfect judge would look like. A first step entails taking the question of discretion seriously: what is it that judges do or ought to be doing in easy cases as well as hard cases where clear legal guidance runs out, and how might a DSS assist? We will provide a proof of concept for the theory in a European context where the question of discretion is particularly salient, namely the ECtHR's wielding of the margin of appreciation. A doctrine that has eluded clear definition since its first usage in 1959. The paper proceeds as follows, the first section traces the main arguments for and against using ML in a judicial context that have emerged in existing literature with a particular focus on flaws in human decision-making. The second section is the theoretical contribution. It starts with an account of Dworkin's right answer thesis and presumptions about the ideal judge Hercules, before linking this vision with new research on AI and ML. The third section provides a proof of concept for the theory in section two, briefly presenting the problem of the unclear application of the margin of appreciation before showing how the AI designed by one of the authors, the ECHR Open Data project (Quemy and Wrembel, 2020) (hereafter: ECHR-OD) may assist in cracking this particular nut on discretion.

Background and related work

To [the layman] judges should and in general do, ... find the law and not make it. The layman's respect for law is founded in large part on his view that this is a fair method of deciding controversies. The academic branch of the legal profession seems now fairly agreed that the layman is mistaken.(Dworkin, 1963: 624)

A central question for anyone wishing to automate any part of the work of the judge, is whether the judge is in fact finding or making the law. In other words, is there a correct legal answer to every question? If there is, it is the task of the judge to find that answer. If on the other hand there are legal vacuums where no law is applicable, judges are left with a law-making task; it is in the nature of courts that they must always reach a conclusion, even when they are in doubt. Positivists are generally assumed to believe that law is found, not made, but a disagreement emerged in the much-cited Dworkin-Hart debate about the reach of this claim. Both Dworkin and Hart assumed that the vast majority of cases would be easy ones where the judge would simply apply the law with no question of making it. The contention was whether this also applied to hard cases where clear guidance from the written law and precedent has run out. Here Dworkin argued that for a sufficiently skilled judge, there would always be a correct legal answer to find and no discretion of personal choice. The empirical turn in legal realist research suggests, however, that personal biases of judges including but not limited to confirmation biases, hindsight biases, political, racial, and gender biases, and biases caused by poor understandings of probability, are present in general legal practice, not just in a small group of hard cases (Chen et al., 2017, Schauer, 2009), they might even be a larger problem in 'routine' or 'easy' cases.

In addition to biases, decisions may be impacted by what Kahneman et al. have termed 'noise'. Where bias has direction, noise is the unpredictability of decisions. In some cases it is useful to differentiate between level-noise, pattern-noise, and occasion-noise (Kahneman et al., 2021)., Level- and pattern noise describe inter-judge differences, the level referring to an over-all 'harshness', 'conservatism' or 'activism' while pattern noise recognises that judges may be restrictive in one type of case and activist in another. A

particularly grave example of level noise is the refugee roulette study from the United States where judges and administrators differed enormously in their propensity to grant asylum, one admitted 88 % of applicants whereas another admitted just 5 % (Kahneman et al., 2021: 7, 73). When Chen et al. modelled the process, they found that the level noise was so high that they could predict the result of an asylum application with 71 % accuracy by incorporating only the identity of the judge and no information whatsoever from the application. However, judges also differed greatly in how predictable they were. For unpredictable judges, Chen found that they tended to hold more hearings than their predictable counter-parts, suggesting that they were not making noisy decisions based on chance, but rather that they were "more sensitive to the circumstances of the cases"(Chen, 2019: 5). That is, they were more likely to treat any case as if it were a hard case. This is important because it shows that predictability and therefore homogenous law application is not sufficient nor necessarily even a sign of the best application of the law, since a bias may be both consistent and predictable. Discrepancies between judges of this kind hurt the legitimacy of a court since it suggests the judges are acting as political actors rather than applying the law.

Long before level- and pattern noise became known by those names, judicial systems have incorporated structures attempting to limit them. One such structure is the obligation to give reasons (Schauer, 1995). Schauer explains how the act of giving reasons implies a self-check on decision-making. In Kahneman's terminology we might say that the judge is forced to engage their System 2-thinking, i.e. the slow and deliberate application of rules to the information provided, as opposed to the fast, intuition-based thinking in System 1 which for an experienced judge may come with almost no effort (Kahneman, 2011). Another approach is the use of the bench, i.e., having decisions made not by a single judge but by a college of judges with or without the inclusion of laypersons as well. This is how the ECtHR is set up. The bench may act as a forum where the intuitions of the different judges meet, as well as a catalyst for the giving of reasons. At the ECtHR, interviews with judges show that they particularly value inputs from judges from different professional backgrounds, because when different intuitions meet, more thorough treatment of the case follows (Bruinsma, 2006: 217, Dzehtsiarou and Schwartz, 2020: 628). The benefit of the bench, however, can be watered down by for example the tendency of groups to yield to seniority (Broude, 2014: 1127, 1148). There is also a string of research suggesting that groups tend to reach more extreme outcomes than individuals do on their own (Kahneman et al., 2021: 94-104, Main and Walker, 1973). Another way of nudging judges to engage their system-2 way of thinking, is by requiring a quantitative scoring. In Kahneman et al.'s analysis, an example of this is the use of the APGAR-scale in the determination of the health of new-borns. It forces doctors to give numerical values to qualitative qualities, such as grimace, crying, and muscle activity (Kahneman et al., 2021: 281-283). The scale is necessarily a simplification of the many qualities an experienced practitioner might pick up on when examining a new-born, but it systematises examination, removes noise, and enhances shareability of information and thereby the ability to discover if something is off. An equivalent use in the field of law might be Robert Alexy's 'Weight formula', a model for assessing proportionality and finding the correct solution in a case where two rights, principles or policy goals appear in conflict with one another. The model includes relative weights of 'the intensity of interference with a principle', 'the abstract importance of the principle' and 'the reliability of presumptions' for a given cause of action (Alexy, 2017: 16-19). The quantification of such abstract concepts is a somewhat artificial endeavour, but the very action of assigning

weights requires the judge or analyst in question to engage their System-2 thinking, to provide reasons.

Noise and errors are defined in Kahneman et al.'s terminology as deviation from the average (Kahneman et al., 2021: 73), and while equality before the law, and thus some amount of homogenisation is central for a just system, it is not at all unthinkable that a situation might arise in which the average judge is 'wrong' and a diverging judge is 'right'. It may for example be that the entire judicial system is systematically biased against a particular group in which case a judge who is not, cannot be said to be wrong. These caveats on level and pattern noise are not replicated for occasion noise, which any theory of justice would advocate eliminating. Occasion noise is differences in judgments depending on fleeting things such as the weather, the judge's mood or stress level, or depending on chance: there is for example a cognitive bias known as the gambler's fallacy where people underestimate the probability of a streak and start second guessing themselves when they reach the same conclusions several times in a row (Kahneman et al., 2021: 84-89). Anchoring biases that emerge by chance (Englich et al., 2006) can also result in occasion noise. The giving of reasons, the bench and potentially a DSS that reminds decisionmakers of relevant similar cases may all help in limiting occasion noise which a particular risk in System-1 or 'gut'-feeling based decision-making.

A more general problem than noise is the problem of objective ignorance – i.e., the problem of the future. Even though algorithms are in some cases better at predicting the future than human decisionmakers, they still get the future wrong in complex systems and real-life (non-toy) cases more often than not (Kahneman et al., 2021: 143-144). One problem is causation. While a computer can generally spot a correlation between temperature increase during the summer and a rise in ice cream sales much faster than a human being can, it cannot tell whether the temperature rose because of the increase in sales or the other way around. In simple situations like this, human beings often have no problems (Pearl and Mackenzie, 2018: 4-10). The problem arises in more complex situations, and situations where bias may interfere as well. It might not for example be immediately clear even to human decisionmakers whether a given geographical area has a higher rate of arrests because it has more crime, or if it is because that area is more heavily policed (O'Neil, 2016: 84-104). Another problem is that legal systems are dynamic rather than monotonic: they change over time. New rules and new precedents can erase or diminish the power of old rules and precedents. Changing is difficult for human beings, and failure to ignore old rules and principles may well be a source of noise in human decision-making, but it is impossible for AIs. Today's ML algorithms learn by incorporating as many examples as possible of a stationary and monotonic object or process, but they cannot unlearn. A pet-recognising AI that has been fed pictures of cats and dogs will get progressively better at recognising these animals as pets, but it would have to be re-trained from scratch if there was suddenly a rule-change that only cats should be recognised as pets. Meanwhile, a human child can unlearn recognising dogs as friendly pets immediately after having been bitten. Furthermore, the retrained AI would have to exclude the previous pieces of knowledge as if they did not exist, with all the problems that implies, far less data, far less accuracy and no leveraging or "connecting the dots" between the past decisions, now invalid, and the new ones.

Theoretical Framework

"it remains the judge's duty, even in hard cases, to discover what the rights of the parties are, not to invent new rights retrospectively" (Dworkin, 1981[1977]: 81)

Legal theories on judging can overall be divided in prescriptive and descriptive theories, and much confusion can be avoided by differentiating clearly between the two. Much of legal realism (Chen, 2019, Kang et al., 2011, Ryberg, 2016) is descriptive/analytical and focused on describing how the judicial system works in practice. Prescriptive/normative theories are focused instead on what law and the judicial branch of government is for and how it is supposed to work. Dworkin's right answer thesis is first and foremost prescriptive, claiming not that every legal question *is* answered correctly, but that every question *theoretically* has one and only one correct answer. Since human beings are fallible and often fail to discover this right answer, he proposed a thought experiment utilising a super-human judge by the name of Hercules (Dworkin, 1981[1977]: 105-130). Hercules has infinite knowledge of the written law and all historical caselaw. If a similar case has been decided before, he would therefore never make the mistake of missing a vertically binding or *stare decisis* precedent. It should go without saying that Hercules is of course completely unbiased and does not get hungry or grumpy.

With the 'right answer' thesis in hand, Hercules can expand interpretations in hard cases where clear legal guidance has run out by generalising principles, but not by creating new policies. The reason for this, is that the law and the principles it contains generate rights for individuals, whereas policies are societal compromises aiming at striking out a direction for society, which is in the realm of democratic politics(Dworkin, 1981[1977]: 82-85, 105). For our purposes it is important to note that predictions are a differentiating factor between policy and principles in that policies make predictions about the future. When deciding policy, legislators make more or less educated guesses about how society will react to a given change in the law. They might think that crime will become less prevalent if punishments are increased, or that lower taxes will increase the number of hours people are willing to work. But because societies are complex and because the future is uncertain, these policies might work, or they might fail. It may turn out that lower taxes instead incentivise people to work less while having the same standard of living, or that longer prison sentences make it harder for former criminals to re-enter society as law-abiding citizens. Rights and principles do not make predictions, they derive their legitimacy by what is not by what might be.

Some of the ML applications currently in use have mixed up this divide between the legal and political realm. The recidivism-predicting software COMPAS and similar applications distract judges from their task of discovering the correct legal answer by nudging them to consider policy-style predictions on potential recidivism. This is already problematic when the predictions are used for bail decisions, but it is particularly worrisome when used in the initial sentencing and when it incorporates facts about the defendant and their environment which are not legally relevant and would be inadmissible in regular court (O'Neil, 2016). Even Hercules is not envisioned with an ability to see the future. Nor is the future in fact relevant when deciding the right legal solution since in the right answer thesis the parties in a court case have a right to a particular judgment on the basis of existing law and the facts of the case. Another problem that emerges when squaring MLs with the task of Hercules, is that computers by design can only apply things that are clearly defined – this is exactly why computers are less noisy; but many of the

principles that Hercules may use to deal with hard cases are by nature vague and applied to a degree rather than either or. If the legislator had not predicted that a specific type of case could emerge, the ML application cannot be prepared for it either.

One place where an ML application could help human judges become more like Hercules, is in the application of some tasks performed by legal practitioners which require abilities for which the machine is already capable of doing better in terms of volume, such as reading more documents and estimating probability distribution. Hercules' infinite knowledge of the written law and historical caselaw would make many human lawyers and judges jealous, but digitalisation and simple searches have already improved their access to this knowledge, and more sophisticated computing may improve it even further. Another problem that legal scholars and practitioners face, is which precedents to rely on. Dworkin spends a few pages discussing how Hercules would handle a precedent decided by an imperfect judge who might not have found the right answer, but for the most part when legal scholars try to make sense of the ECtHR, the problem is rather the overwhelming amount of caselaw, some of which might point in one direction while other point in another. Here ML applications may help make sense of caselaw and categorise it in accordance with the articles in question, the types of questions engaged with, or indeed whether caselaw is moving in one direction or another. At the ECtHR this last question has received a lot of attention since it has become a focal point for an increasingly contentious debate about whether the ECtHR is becoming more deferential to states or less. It is beyond the scope of this paper to map this discussion in detail, but it is worth pointing out that States parties in recent years have called on the ECtHR to allow more subsidiarity in its interpretations (Brighton Declaration, 2012, Copenhagen Declaration, 2018) while there has been disagreements about whether the ECtHR was already doing this or not (Bates, 2020, Flogaitis et al., 2013, Spano, 2014).

In the following section we will demonstrate how an ML application for systematising ECtHR caselaw might assist in illuminating the usage of the margin of appreciation. The claim here is not that this application is a fully fledged DSS for studying and practicing human rights law at the ECtHR, but rather to provide a proof of concept of what kind of assistance an application such as this one might lend in the future.

Empirical Proof of Concept: The margin of appreciation

The ECHR-OD (Quemy and Wrembel, 2020)[4] uses a combination of natural language machine learning algorithms and traditional rule-based systems to structure ECtHR case law. The hope is that with time this can help researchers and practitioners gain access to Hercules' complete knowledge of all past caselaw. Dworkin created Hercules for his own jurisdiction in the United States, where there is a doctrine of binding precedent, while at the ECtHR there is no such doctrine. Precedents are still important, however, since the ECtHR has often reiterated its statement from Cossey v. the United Kingdom (1990, para 35) that while it is not bound by its earlier judgments, it will not depart from a previous interpretation without good reason. A good reason in this case may be that present day conditions of a social or technical nature require a new interpretation or because a European consensus has emerged.

For scholars and practitioners this means that any precedent could potentially be issuing a pull of persuasion and thus must be considered. The ECHR-OD is envisioned to enable

[4] https://echr-opendata.eu/

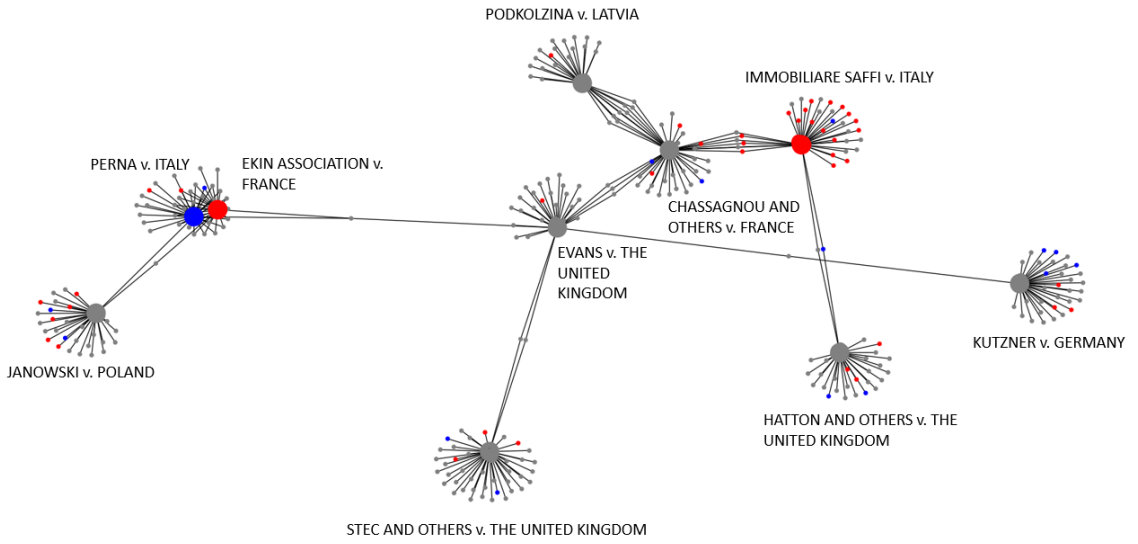
users to make more granular searches than are currently possible through the HUDOC database. This would allow retrieving cases regarding a particular sub-theme of a right where a particular doctrine has been used, such as 'deference arguments in cases on the right to family life as threatened by expulsion', or 'procedural questions in property right cases regarding pensions', it would also allow scholars to ask questions on statistics, probabilities, and interesting correlations in the caselaw, such as: 'does the court refer to the same cases when utilising the margin of appreciation as the literature does?' or 'are some states more likely to be afforded a wide margin than other states in similar cases?'. As well as legal realist-type questions on equality before the law at the ECtHR, such as, 'is there more or less state-deference in cases where the applicant belongs to a religious, ethnic, or sexual minority?'.

The uses suggested above will not change the work of either scholar or practitioner to any large degree, but it will allow for a high abstraction level overview of cases to be added to the low abstraction level knowledge of individual caselaw that is commonly associated with legal methodology. Furthermore, the uses presented here are likely recognisable to most readers as regular search queries in regular database programming. This is on purpose. It is a key point of this paper to point out that the computer at present is not capable of doing anything the human being cannot do, though oftentimes it can do it faster. The added value of the ECHR-OD does not lie in an uninterpretable black-box machine learning model with billions of parameters, but in the initial data-processing to extract knowledge and represent it in a way that is suitable both for constructing complex queries and for teaching a more complex ML algorithm to answer particular questions. In fact, the pre-processing partly uses ML algorithms, not to directly represent knowledge, but to extract information to present it in a meaningful way. Let us give a concrete example of the opposition in the two approaches: by using a natural language model, we can extract semantic entities such as implicit references to other cases, dates, places, NGOs, and persons. The model extracts information based on which one can try to answer a particular question about the Court. This approach differs from using a model to directly predict, for instance, the admissibility of an application. The difference does not lie in the technique – most likely two natural language models in this case – but in the essence of what we try to achieve: replacing the 'expert knowledge versus providing a structured information'-approach with one enhancing the capabilities of the expert knowledge. Essentially, this structuring of information could also have been undertaken by an army of research assistants, but by utilising ML it can be done faster, and new cases can be added automatically which means that the database should not become outdated the way a regular dataset does.

To illustrate how one might use this type of application, we made a query on the margin of appreciation in general and on article 6 cases. From court watching we know that the ECtHR tends to include references to previous uses of the margin of appreciation when it invokes the doctrine, and we were interested to know whether these references would be to a handful of popular cases the way it often is in the literature, or whether the court employs another logic. This is a similar approach to Frese and Olsen's study of citation networks in article 14 cases (Frese and Olsen, 2019) but with the added value that the network can be made only for the specific paragraph where the margin of appreciation mention appears.

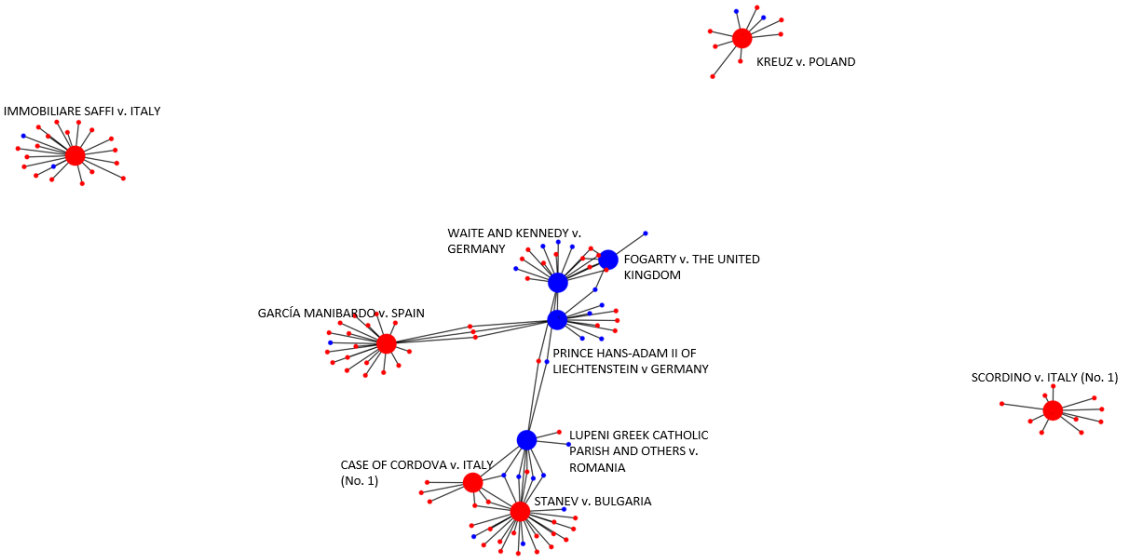
By creating a citation network of cases that are referenced in the vicinity of a mention of the margin of appreciation, we see that just some 20 cases are referenced more than 20

times while more than 300 cases are referenced only once. Readers familiar with the literature on the margin of appreciation might recognise some of the often-referenced cases such as *Chassagnou*, *Evans* and *Stec*, while other often cited cases are relatively obscure. We also see that some often-cited cases, *Kutzner*, *Hatton*, *Stec* tend to create their own universes, while others *Chassagnou*, *Evans*, *Podkolzina*, *Immobilare*, form a more robust network.



The large notes represent the 10 most cited cases near margin of appreciation mentions while the smaller notes are the citing cases, red denotes a case containing a violation of Article 6, blue a non-violation, while grey cases do not deal with Article 6.

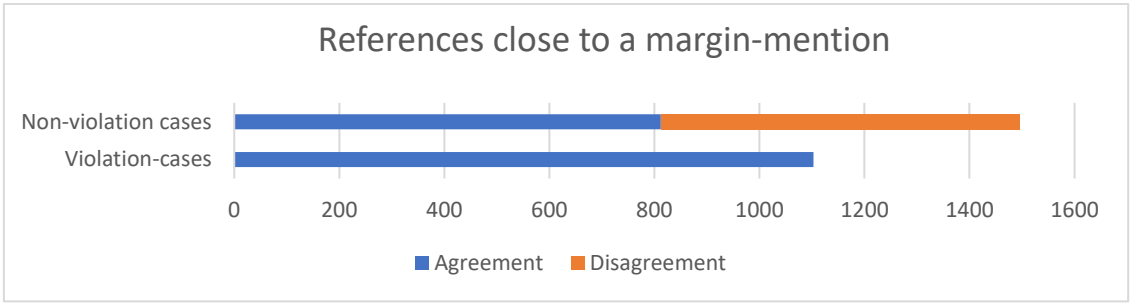
Going into a bit more detail using Article 6 cases, the results were both surprising and familiar. Surprising because the top ten most frequent references for margin of appreciation application in Article 6 cases while not unknown are not referenced all that often in the literature. Familiar because they are cases of a certain kind, namely principle-establishing cases. *Stanev v Bulgaria* for example establishes that although the state has a margin of appreciation in how to set up its system of procedural protections, it still must ensure direct access to testing in court for individuals deprived of their legal competence (Stanev v Bulgaria, 2012: para. 230). *Garcia Manibardo* similarly clarifies the essence of Article 6 to be the access to having a case tried in court, and it lays down principles the reach of states' margin of appreciation to ensure the effective functioning of their courts by limiting the number of cases that reach top courts, but maintains that the reasoning for denying appeal cannot be arbitrary or overly formal(García Manibardo v. Spain, 2000: para. 36).



The 10 most frequently referenced cases in margin of appreciation application in Article 6 cases.

The network furthermore shows a clear tendency that precedent-setting cases create their own small networks. If we take *Scordino v Italy* for example, it sets precedents regarding margin of appreciation afforded to the state with regards to determining damages and remedy in court-cases(Scordino v Italy (no. 1), 2006: para 189) which was referenced in a Russian pilot-judgment procedure and resulted in a large number of short-form violation decisions.

We also learned that there is a strong tendency that references to the margin of appreciation in judgments that ultimately result in a violation-conclusion, are much more likely to refer to other violation-decisions, whereas cases that ultimately result in a non-violation conclusion, refer to a more mixed bag of cases including both violation- and non-violation decisions.



This finding initially appeared surprising, but upon going through the most referenced cases, a pattern emerged that in cases resulting in a violation-decision which mention the margin of appreciation, such as *Stanev v Bulgaria*, there is a tendency that the only actor mentioning the margin and providing a reference for it, is the Court itself, not for example the respondent state, and in these cases, often to clarify that 'notwithstanding the margin of appreciation', the case in question touches upon the essence of a right, and therefore must result in a violation judgment. To many legal-doctrinal scholars this will not come as a surprise, but it does bring into question claims that increased mentions of the margin of appreciation are necessarily a sign of increased deference to states(Spano, 2018).

Conclusion

This brief paper aimed to demonstrate three things. First, that securing the right to a fair trial within a reasonable time for all citizens remains a serious challenge across Europe and as a result there is a strong pull towards digitalisation and the usage of AI and ML algorithms both domestically and at the level of the European Court of Human Rights. Second, that many of these current usages, which have received scathing criticism in academic commentary and in various courts, rely on data that is skewed initially or nudges courts to answer questions that they are not envisioned or equipped to answer. Furthermore, there are structural reasons, including the inability of AIs to unlearn or understand causation, and computers' inbuild avoidance of vagueness, which limit the usability of ML in a judicial setting. Third, despite these caveats, there remains a space for AI in the administration and study of justice, but utilising it requires a clearer understanding on the part of data-scientists of what a court is for and what a judge is supposed to be doing, as well as an understanding among jurists of what the AI can and cannot do. The most obvious usage outsources to computers only our most basic cognitive abilities: language (translation, entities extraction, text generation), vision (image recognition, classification, segmentation), audio (text to speech, speech to text, spectrum analysis). It will then be up to the human scholar, judge, or lawyer to make sense of the information, to balance, to analyse and draw causal links, since this is what humans are better at than computers – provided they think slow rather than fast.

Bibliography

ALEXY, R. 2017. Proportionality and Rationality. *In:* JACKSON, V. C. & TUSHNET, M. (eds.) *Proportionality: New frontiers, new challenges.* Cambridge University Press.

ALSTON, P. 2019. Report of the Special Rapporteur on extreme poverty and human rights: The digital welfare state. New York: United Nations General Assembly.

BATES, E. 2020. Strasbourg's integrationist role, or the need for self-restraint? *The European Convention on Human Rights Law Review,* 1**,** 14-21.

Brighton Declaration. 2012. *In:* MINISTERS, C. O. (ed.). Brighton: Council of Europe Member States.

BROUDE, T. 2014. Behavioral international law. *U. Pa. L. Rev.,* 163**,** 1099.

BRUINSMA, F. J. 2006. Judicial Identities in the European Court of Human Rights. *Multilevel Governance in Enforcement and Adjudication***,** 203-40.

CHEN, D. L. 2019. Machine learning and the rule of law. *Revista Forumul Judecatorilor***,** 19.

CHEN, D. L., DUNN, M., SAGUN, L. & SIRIN, H. 2017. Early predictability of asylum court decisions.

Copenhagen Declaration. 2018. *In:* MINISTERS, C. O. (ed.). Denmark: Council of Europe Member States.

Cossey v. the United Kingdom. 1990. European Court of Human Rights.

DWORKIN, R. 1963. Judicial Discretion. *The Journal of Philosophy,* 60**,** 624-638.

DWORKIN, R. 1981[1977]. *Taking Rights Seriously,* London, Duckworth.

DZEHTSIAROU, K. & SCHWARTZ, A. 2020. Electing Team Strasbourg: Professional Diversity on the European Court of Human Rights and Why it Matters. *German Law Journal,* 21**,** 621-643.

EDEL, F. 2007. The length of civil and criminal proceedings in the case-law of the European Court of Human Rights. *In:* EUROPE, C. O. (ed.) *Human Rights Files.* Strasbourg: European Court of Human Rights.

ENGLICH, B., MUSSWEILER, T. & STRACK, F. 2006. Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin,* 32**,** 188-200.
ENGLISH, R. & SPANO, R. 2020. New Strasbourg Court President on AI and the law. *In:* METZER, J. (ed.) *Law Pod UK.* United Kingdom: Crown Office Row: UK Human Rights blog.
FLOGAITIS, S., ZWART, T. & FRASER, J. 2013. *The European Court of Human Rights and Its Discontents,* Camberley, Edward Elgar Publishing.
FRESE, A. & OLSEN, H. P. 2019. Citing Case Law: A Comparative Study of Legal Textbooks on European Human Rights Law. *European Journal of Legal Studies,* 11**,** 91-131.
García Manibardo v. Spain. 2000. European Court of Human Rights.
HILL, L. L. O. H. 1998. Universality versus subsidiarity: a reply. *European Human Rights Law Review,* 3**,** 73-81.
HLR 2017. State v. Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing. *Havard Law Review,* 130**,** 1530-1537.
HOLMES, O. W. 1897. The Path of the Law. *Harvard Law Review,* 10**,** 457-478.
KAHNEMAN, D. 2011. *Thinking, fast and slow,* Macmillan.
KAHNEMAN, D., SIBONY, O. & SUNSTEIN, C. R. 2021. *Noise: a flaw in human judgment,* Little, Brown.
KANG, J., BENNETT, M., CARBADO, D., CASEY, P. & LEVINSON, J. 2011. Implicit bias in the courtroom. *UCLa L. rev.,* 59**,** 1124.
MADSEN, M. R. & SPANO, R. Authority and Legitimacy of the European Court of Human Rights: Interview with Robert Spano, President of the European Court of Human Rights. European Convention on Human Rights Law Review, 2020. Brill Nijhoff, 165-180.
MAIN, E. C. & WALKER, T. G. 1973. Choice shifts and extreme behavior: Judicial review in the federal courts. *The Journal of Social Psychology,* 91**,** 215-221.
O'NEIL, C. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy,* New York, Crown Publishers.
PEARL, J. & MACKENZIE, D. 2018. *The book of why: the new science of cause and effect,* Basic books.
QUEMY, A. & WREMBEL, R. 2020. On Integrating and Classifying Legal Text Documents. *Database and Expert Systems Applications.* Bratislava: Springer.
ROSS, A. 2019[1948]. *On law and justice,* Oxford University Press.
RYBERG, J. 2016. *Domstolens blinde øje: Om betydningen af ubevidste biases i retssystemet,* Djøf Forlag.
SCHAUER, F. 1995. Giving Reasons. *Stanford Law Review,* 47**,** 633-659.
SCHAUER, F. 2009. Legal realism. *Thinking Like a Lawyer.* Harvard University Press.
Scordino v Italy (no. 1). 2006. European Court of Human Rights - Grand Chamber.
SPANO, R. 2014. Universality or Diversity of Human Rights - Strasbourg in the Age of Subsidiarity. *Human Rights Law Review***,** 487.
SPANO, R. 2018. The Future of the European Court of Human Rights—Subsidiarity, Process-Based Review and the Rule of Law. *Human Rights Law Review,* 18**,** 473–494.
Stanev v Bulgaria. 2012. European Court of Human Rights Grand Chamber.
VERVLOESEM, K. 2020. How Dutch activists got an invasive fraud detection algorithm banned. *Automating Society Report* [Online].
ZWART, T. 2013. More Human Rights than Court: why the Legitimacy of the European Court of Human Rights is in Need of Repair and how it can be done. *In:* FLOGAITIS, S., FRASER, J. & ZWART, T. (eds.) *The European Court of*

Human Rights and its discontents : turning criticism into strength. Cheltenham [etc.]: Elgar.