THE BEST OF US OR THE WORST OF US: A CRITICAL ANALYSIS OF THE IMPACT OF AI ON THE LEGAL SYSTEM

DEWAR, SARAH
Scots Law Accelerated LLB, College of Social Sciences

ABSTRACT

The discourse around Artificial Intelligence (AI) simultaneously claims this ballooning technology to be both a harbinger of doom and an innovative boon. Proponents have increasingly cited AI as a potential driver of efficiency and analytical neutrality, encouraging its application in new areas of work and study. Crucially, AI has come to be seen as a site of evolution for the legal profession and processes, with firms investing in legal research technologies to streamline operations and courts exploring the potential for the application of AI in sentencing decisions. But even more pertinent are the aspirations present amongst AI optimists for a larger and more potent role for AI in the future, be that in providing accessible legal advice to large groups of people or in analysing large quantities of legal data to help secure optimal outcomes for those using these new programmes. However, it is the position of this paper that these aspirations are, at best, utopian (failing to reconcile certain practical limitations of the technology with the legal profession and practice that it seeks to represent) and, at worst, dystopian (exacerbating existing violations of due process, undermining accessibility of justice, and reducing judicial legitimacy). As such, legal AI must be rigorously examined in relation to the validity of its inputs, processes, and backers in order to prevent pernicious path dependencies from rearing their heads in the years to come.

INTRODUCTION

Artificial Intelligence seeks to emulate the processes and outputs of the human mind in its structures (Borg et al., 2024, p.8) and varies widely in its capabilities, approaches, and aims across models and industries. The capacity of particular software varies in line with whether it is 'Narrow' or 'General.' 'Narrow' AI processes such as Machine Learning (ML), Neural Networks (NNs), and Natural Language Processing (NLP) have grown in complexity, capability, and application in recent years, whilst 'General AI' remains beyond reach indefinitely (Engelke, 2024, p.3). Narrow AI technologies (which are the focus of this paper) have become increasingly common, infiltrating corporate and public bodies alike. Attractive due to their pace and indefatigability relative to their human counterparts, they represent cost-effective and productive alternate sites of labour across industries.

Legal administration tasks, such as the retrieval of documents, have long leant on online databases such as Lexis and Westlaw (Engelke, 2024). However, recent innovations have broadened AI's legal applications, with many bodies undertaking swift and extensive implementation of AI in the legal sphere. It now generates answers to legal questions (Kohn, 2016, p.2) and tracks key indicators like legislation usage, representation, and judicial responses to arguments (Kohn, 2016, p.2). Moreover, AI can now undertake document analysis, contract scanning, intelligence generation, document delivery, advisor support, clinical negligence analysis, and even generate outcome prediction (The Law Society, 2018, p.6). AI is omnipresent. Governments have responded with debate, consultation (Gov.UK, 2024) and, subsequently, regulation, e.g., the right to review decisions made by AI under GDPR (Engelke, 2024, p.4). Courts have similarly warmed to AI. The Chabot DoNotPay has been used to appeal parking tickets in the City of London (Park, 2020, p.47). AI has arbitrated small claims in Estonia (Park, 2020, p.46). A Colombian judge has publicly used ChatGPT to generate a decision (Alegre, 2024, p.99), and lawyers are increasingly using LLMs to generate pleadings, e.g., Varghese v China Southern Airlines Co (Alegre, 2024, p.93).

However, this paper seeks to answer the question left unexamined: can legal AI be compatible with the foundational principles of legal systems? To do so, the paper interrogates the following: firstly, what information is being fed into these systems, asking whether input bias is present and/or avoidable and what the implications of such biases could be. Secondly, even assuming the inputs are 'clean', what processes are these data points being subjected to, and is it possible to internalise legal logic and maintain principles of accountability and due process within these models? Thirdly, how are the incentives of actors within this process (owners, creators, and consumers) impacting the likely model outcomes? Are monopolistic power structures and high-paced markets conducive to positive judicial outcomes? It is the finding of this paper that, at any and each point, AI is incapable of compatibility with legal principles (both contemporarily and in an idealised form), compelling this paper to issue a stark warning.

WHAT IS THE SYSTEM BEING FED?

Without inputs upon which to train, AI cannot reason. Indeed, the greater the volume of information we 'feed' AI, the better the results (Engelke, 2020, p.6). This creates a ravenous appetite for inputs; Alpha Zero (a model trained to play the strategy board game Go) is one such example, playing 21 million games over 34 hours in order to win every future match-up (Borg et al., 2024, p.26). Alpha's inputs could be readily generated, but for most models, input generation is a labour-intensive, time-consuming, and expensive process, contingent upon arduous practices like human labelling (Guardian, 2024) and organic generation.

Even where plentiful inputs are available, progress may be hampered by the absence of high-quality inputs, which are also necessary for model success. In law, this necessitates an ability to deconstruct complex legal issues into sub-problems, rules, and premises through which information can be passed (Buchanan and Hendrick, 1970, p.7). Such deconstruction is a substantial hurdle as it 'often causes the subject of the inquiry to rethink his fundamental reasons for holding one set of principles rather than any other' (Buchanan and Hendrick, 1970, p.8) because of the non-linear and relative nature of law and legal reasoning. As such, the ready, enthusiastic, and committed participation of legal professionals is also necessary to avoid a 'bottleneck' of data availability and cleanliness (Borg et al., 2024, p.27). However, this necessitates disclosure and dissemination of career spanning (and profit-generating) expertise, which drives the success of individual lawyers and their firms. Therefore, cooperation is unlikely to be gleeful and forthcoming as these AI projects threaten lawyers with obsolescence and compel the sacrifice of comparative advantage.

Additionally, robust attempts to model 'macroscopic concepts' reflective of human perception are also necessary (and difficult); for instance, when we fill a cup, we conceptualise 'cup', 'spout', 'teapot', 'tilt', etc., all of which an AI model would have to emulate (Sloman, 1979, p.4). Moreover, these models will need to have some understanding of those aspects of human life which are so commonplace we struggle to recognise them ourselves (Sloman, 1979, p.5). Ensuring these inputs are present requires gargantuan introspective efforts, and even then, comprehensiveness may be epistemically impossible. Sourcing poses further concerns, as the need for, and drive to acquire sensitive data (often in law firms' possession e.g., client and case records) will generate substantial privacy concerns (Engelke, 2024, p.6). AI's elephantine 'inability to forget' means, once entered into a model, data or its legacy is seldom forgotten and can, in some instances, be retroactively extracted via inversion attacks (where information is gleaned by a model's response to prompts) (Borg et al., 2024, pp.91-92). As such, rights to privacy and legal professional privilege are ripe for violation, especially where firms face the apparent necessity of developing proprietary models, particularly because once violated, rectification becomes almost impossible due to entrenchment and inability to meaningfully anonymise entries (Borg et al., 2024, pp. 92-105).

AI reflects the inputs we feed it, it necessarily mirrors our behaviours, often with concerning consequences, e.g., chat-bots, which drive conversations toward sexual content whether or not the marginal user pursues it (Alegre, 2023, p.53). For law, there is a similar concern that AI, which is trained on information spanning many years and judicial/socio/political trends, will generate perverse outcomes. The product of this variety of tone, language and morality is beyond the scope of this paper, but we can estimate some potential outcomes. Outdated language, difficulty parsing the relative importance of concepts and sentiments, archaic, punitive, and carceral returns are all likely results. Therefore, the quality of recommendations made to individual clients, judges and firms will be poor as a consequence. AI inherits a status quo bias toward current structures, norms, and paradigms (McQuillan, 2022, p.43) but crucially 'know(s) nothing of history, power or meaning' (McQuillan, 2022, p.13). This is particularly dangerous where pernicious biases influence judicial outcomes, e.g., where the COMPAS system consistently 'under risked' the predicted recidivism of white offenders (Borg et al., 2024, p.121). The company argued this outcome simply reflected the pre-existing statistical landscape (Borg et al., 2024, pp.121-123) but failed to reckon with the prejudicial causes of the inequality and the pernicious impact of using such data to condemn or exonerate people into the future and the further disenfranchisement of minority communities this would create.

Statistical inputs are essential to this disenfranchisement process as they 'provide the objectifying force of algorithmic knowledge and the equitability necessary for computation' (Joque, 2022, p.88), even if such equity does not exist for the material subjects they claim to represent. Objectification (the Marxist theory of the process by which labour is alienated from its product) 'relieves us of the necessity of remembering', alienating the contents of these models from the broader landscape and making deconstruction harder (Joque, 2022). This alienation is made worse as the need for greater volume and quality of inputs grows, meaning data acquisition is increasingly venerated as a necessary means to laudable ends (Crawford, 2021, p.95). In turn, meaning models and their proprietors are increasingly perceived as being entitled to said data. This is an especially concerning trend as data is not a mere commodity; it often represents moments of vulnerability and pain for those whom it signifies. For example, an AI model trained on mugshots effectively strips images of context and their subjects of personhood in the pursuit of judicial and commercial profit, preferencing private interest above the interest of the non-consenting subjects. This moves pictures from 'image to infrastructure' (Crawford, 2021, p.93). To prevent such alienation, standpoint theory proposes 'alternative ways of knowing rooted in the lived experiences of people who are marginalised' (McQuillan, 2022, p.105), rejecting the reduction to statistics and data points that AI models encourage.

WHAT PROCESS IS THE INFORMATION UNDERGOING?

Even if inputs are perfectly clean and ethically sourced, these systems are still far from equitable and are eminently corruptible. The models are a series of 'if' and 'then' statements which codify labour. Just as the Jacquard looms of the 1800s internalised the labour of weavers, legal AI uses the rules, logic and practices of modern lawyers, clerks, and administrators (Pasquinelli, 2023) to generate outputs. AI is necessarily an imitation of existing labour, incapable of innovation much as a Jacquard Loom 'could not exceed or break the chain of reasoning that it was representing and materially embodying [...] today's algorithms for data analytics that are rebranded 'machine learning' and 'artificial intelligence' cannot creatively break the rules on which they are based and, more importantly, cannot consistently invent new ones' (Pasquinelli, 2023, p.73). This technology cannot exceed its creative ceiling as it is premised on existing knowledge and operates by reordering and sifting through information, meaning it seeks answers within the bounded limits of the data already in our possession and cannot question whether it is possible to redraw those bounds (McQuillan, 2022, p.43). Moreover, deductive AI, such as NNs, struggle to meet the specific demands of the law. The absence of pure logic or 'Truth Value' makes the application of deductive reasoning (from general to specific) difficult. Whilst it may be applicable in 'clear cases' (Susskind, 1986, p.189), marginal cases are either beyond the scope of AI or likely to return inaccurate results. The classification of the facts of cases likewise requires deductive thought, and as such, is also beyond the scope of AI, further marginalising the potential applications of this technology (Susskind, 1986, p.190).

Even the most advanced models are far away from generating definitive answers to complex legal questions. For NNs, outcomes are predicated on the artificial neuron, which receives the largest signal (McQuillan, 2022, p.17). This answer can only be a 'most likely' claim, not a definitive statement (McQuillan, 2022, p.31). 'Post-hoc guesswork' explains the pattern but not the phenomena. For example, AI suggested the discharge of asthmatic patients with pneumonia from the hospital because data showed them consistently exiting the system. All because the AI could not account for the extraneous fact that said patients were being sent to intensive care (McQuillan, 2022, p.32). AI's reasoning struggles to conceptualise the relative validity of one pattern versus another, e.g., correlation from cause and, as such, is invariably limited. However, problems are not prevented by correct pattern recognition alone, as this does not ensure the accuracy of application. Application of a general rule in a specific case may indeed lead to a violation of rights and injustice (Reed, 2018, p.3). It is insufficient to apply probabilities of guilt, generalised rules, or recidivism predictions to specific cases as a means to condemn individuals. To do so obliges the application of the general to the specific, compelling the judgement of the individual by the actions of others. It is thus antithetical to due process (McQuillan, 2022, p.37). The case of State v Loomis shows the danger of such opaque systems and how they can violate due process rights (Deeks, 2019, p.16). In this case, AI was used to assess risk in sentencing despite the defendant having no access to the programme's methodology. Consequently, the defendant was locked out of the decision-making process.

In fact, it is often the case that those subject to the whims of AI decisions have little access to its reasoning. This lack of explanation is often by design, as the 'explainability' of a model's choices comes at the expense of its accuracy (Deeks, 2019, pp.1844-1845). As such, 'Black Box' AI is incredibly popular (models whose reasoning and inner machinations cannot be retrieved) as companies and researchers have chosen accuracy over explainability. Our only route to understanding is through a kind of reverse engineering, either the 'exogenous approach' (Deeks, 2019, p.1835), where relevant information is provided to help explain how the algorithm is functioning, or through the 'decompositional approach', where we attempt to replicate the model's reasoning (Deeks, 2019, p.1835). Both of these approaches are totally insufficient, as we remain incapable of explaining or validating how individual decisions are made. Neither decision makers nor subjects can be sure that extraneous or prejudicial considerations were not made, nor can they be sure key inputs were not neglected. Without an understanding of the basis for the decision, a victim struggles to challenge its substance as they cannot point to the injustice. The procedure is unfair, and participation with and support for the judicial system likely flounders as people lose respect for it. Justice appears to forgo malleability and empathy in the pursuit of codification via technology (Wu, 2019, p.22-23).

Some AI proponents argue deficits in explainability can be traded off for AI's social benefits of efficiency and pragmatism, and losses compensated for in time, with prior explanation needed only where individual rights are at risk (Reed, 2018, pp.8-10). This is incompatible with a just and fair legal system. Take the COMPAS parole algorithm, for example, the assertion that under-risking white offenders was value-neutral and a statement of fact implies a moral equivalence between a false positive and a false negative. The subject of the former incurs 'a sense of losing control and freedom, feelings of uncertainty, separation and loss' (Harvey, 2016), loan and debt dependency, lack of access to housing, and relationship breakdowns (Dickie, 2013), while the subject of the latter is free. Such a 'mistake' within law has the potential for seismic impacts. Access to explanations of how COMPAS (and indeed any AI) arrives at its conclusions is essential in facilitating challenges to unfair decisions and ensuring fair trials for those passing through courts (Deeks, 2019). Where explainability fails to be provided, AI must not be deployed.

WHO IS MAKING THIS? WHAT ARE THEIR INCENTIVES?

Owners of AI face a hyper-competitive and accelerationist market with a clear first-mover advantage necessitating pre-emptive commitment to AI development and innovation (Marchant, 2018), often at the expense of quality control. Capital requirements are high, with significant start-up, R&D, and infrastructure spending required, meaning established Cloud Computing firms are the mainstays of the industry (Engelke, 2020, p.18). Smaller firms may build upon the work of these companies. However, they need to purchase access to products, e.g., data sets and other proprietary software, in order to do so. This makes AI (in effect) a service industry led by the likes of Microsoft, Amazon Web Services, and IBM Watson (Borg et al., 2024, p.90). As such, the development of models becomes a tangled web rapidly accelerating in a climate of disdain for and obfuscation from liability (Borg et al., 2024, p.150), where model-induced harms are seen as inevitable developmental costs. It is arguably this climate which encourages unsafe technology to be deployed before it is totally safe, e.g., the death of the pedestrian Elaine Herzberg in 2018 during a test of Uber's self-driving car (Borg et al., 2024, p.150).

The creators of AI are equally flawed. With models generally being built by those with doctoral degrees in mathematics or associated disciplines, experience with and literacy within ethics is often lacking and seldom supplemented (Borg et al., 2024, p.38). Creators are urged and pressured by employers to ensure swift production of models in order to generate returns (Tarnoff, 2022, p.121), and consequently, data has become a de facto financial asset with 'both an operational value and a speculative financial value' (Tarnoff, 2022, p.121). This entrenches the subservience of ethical considerations to fiduciary motivations in model construction. To be clear, this is not exceptional in the digital space but rather a subsequent part of 'the triumphant phase, in which the internet was remade for the purpose of profit maximisation' (Tarnoff, 2022, p.123).

The impact upon technology users is also seismic. The encouragement to objectify data and allow algorithms to 'do the thinking' for them makes the technology appear increasingly unknowable, like a day trader to whom 'what is a fully human process of production and exchange appears to the trader, as an unknowable fickle system whose whims are even less predictable than the weather' (Joque, 2022, p.87). The consumer sees an unknowable but venerable process beyond reproach, whether that be because of the AI's capabilities, or our predisposition to find meaning in language even absent intention (such as LLM outputs) (Alegre, 2024, p.22), or the encouragement we receive to trust AI implicitly (e.g., the AI guided Patriot missile system where 'operators were trained to trust the system's recommendations with reverence' despite regular failure) (Borg et al., 2024, p.62). This veneration is so great that even where a human is still 'in the loop', they will likely be overridden on the grounds of technical supremacy, pace, and efficiency despite the system generating sub-optimal outputs. Even when users are encouraged to intervene, the necessary 'assimilation of information' and 'situational awareness' needed to react in time is 'unrealistic' (Borg et al., 2024, p.62), effectively locking in errors, especially in high-paced scenarios.

CONCLUSION

At all points, either due to personal incentives, pressures from employers, or deference, it is clear that the material reality in which AI is placed encourages predation, ethical ignorance, and disempowerment, none of which can be said to best facilitate positive judicial outcomes. The inputs are fundamentally biased as they inherit the myriad faults of judicial systems, a problem that there is little to no corporate appetite to rectify. In turn, these models incentivise a new commodity fetishism where data is prized to the extent that insidious procurement is excused and encouraged. Finally, context is stripped from inputs to the detriment of understanding, diminishing the quality of model outcomes. To the extent that all of this necessitates sub-par judicial outcomes, this is not a facet of a technology that could be said to be compatible with justice. The processes are fundamentally corrupted as they are incapable of recognising their faults, presenting definitive answers where nuance is required and venerating their own conclusions by locking out or even prohibiting questioning, undermining due process and accessibility. Additionally, actor incentives are fundamentally perverse, with high-paced profit-seeking firms encouraging a climate of disdain for social, moral, and ethical considerations and producing models in their image. In so doing, we run the risk of fragmenting trust in and support for our judicial systems and undermining their validity. Crucially, these are not shortcomings that can be permitted to go unchecked as the stakes are immense. The threat of this technology being deployed as it currently exists clearly undermines myriad rights essential to judicial function and fairness. From due process to privacy and fair trial, allowing this slow creep to continue unchecked, unregulated or, more importantly, unexamined represents a threat to every one of these principles. This necessitates policy and regulation attuned to corporate incentives and algorithmic construction, a policy which foregrounds ethical and moral considerations over profit and convenience. Without such careful and rigorous review, AI can only ever reflect the worst of us.

REFERENCES

Alegre, S. 2024. *Human rights, robot wrongs*. London: Atlantic Books.

Borg, J. S., Sinnott-Armstrong, W. and Conitzer, V. 2024. *Moral AI*. London: Penguin Press.

Buchanan, B. G. and Headrick, T. E. 1970. Some speculation about Artificial Intelligence and legal reasoning. *Stanford Law Review*. **40**, pp.40-62.

Crawford, K. 2021. *Atlas of AI: power, politics, and the planetary costs of Artificial Intelligence*. Connecticut: Yale University Press.

Deeks, A. 2019. The judicial demand for explainable artificial intelligence. *Columbia Law Review*. **119**(7), pp.1829-1850.

Engelke, P. 2020. AI, society, and governance: an introduction. *Atlantic Council*. [Online]. [Accessed 25 October 2024]. Available at: https://www.jstor.org/stable/resrep29327

Gov.UK. 2024. *A pro-innovation approach to AI regulation: government response*. [Online]. [Accessed 25 Oct. 2024]. Available at: www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response

Graham, M., and Callum, C. 2024. *Meet Mercy and Anita – the African workers driving the AI revolution, for just over a dollar an hour*. [Online] [Accessed 25 October 2024]. Available at: https://www.theguardian.com/technology/article/2024/jul/06/mercy-anita-african-workers-ai-artificial-intelligence-exploitation-feeding-machine

Harvey, J. 2016. Understanding the experience of imprisonment. *The British Psychological Society*. [Online]. [Accessed 26 Oct 2024]. Available at: www.bps.org.uk/psychologist/understanding-experience-imprisonment

Joque, J. 2022. *Revolutionary mathematics: Artificial Intelligence, statistics and the logic of capitalism*. London: Verso Books

Kohn, P. E. 2016. How Artificial Intelligence is revolutionising the legal practice. *Litigation*. **43**(1), pp.12-13.

Marchant, G. and Covey, J. 2018. Robo-lawyers: your new best friend or your worst nightmare? *Litigation*. **45**(1), pp.27-31.

McQuillan, D. 2022. *Resisting AI*. Bristol: Bristol University Press,

Park, J. 2020. Your Honour, AI. *Harvard International Review*. **41**, pp.46-48.

Pasquinelli, M. 2023. *The Eye of the Master*. London: Verso Books.

Reed, C. 2018. How should we regulate Artificial Intelligence? *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*. **376**(2128), pp.1-12.

Sloman, A. 1979. Epistemology and Artificial Intelligence. In: Michie, D. ed. *Expert systems in the Microelectronic Age*. [Online]. Edinburgh: Edinburgh University Press. [Accessed 25 Oct 2024]. Available at: https://cogaffarchive.org/sloman-epist-ai.pdf

Susskind, R. E. 1986. Expert systems in law: a jurisprudential approach to artificial intelligence and legal reasoning. *The Modern Law Review*. **49**(2), pp.168-194.

Tarnoff, B. 2022. *Internet for the people: the fight for our digital future*. London: Verso Books

The Law Society. 2018. *Artificial intelligence (AI) and the legal profession*. [Online]. [Accessed 25 October 2024]. Available at: https://www.lawsociety.org.uk/topics/research/ai-artificial-intelligence-and-the-legal-profession

Wu, T. 2019. Will artificial intelligence wat the law? The rise of hybrid social-ordering systems. *Columbia Law Review*. **119**(7), pp.2001-2028.