

The Limited Potential of AI Implementation in US Courts

Eric Kim

Seoul International School, Seoul, Republic of Korea, erickim061129@gmail.com

ABSTRACT: With the recent surge of interest in the use of AI in generating writing that imitates human creativity, many have begun turning to AI as the ultimate solution to various social problems. Amidst rising concerns about discrimination in judicial decisions stemming from judges' personal leanings within the US criminal justice system, employing AI as judicial judges to eliminate these biases has been considered. Although the current stage of AI and its abilities render it impossible to replace human judges, AI technology's fast-paced development opens up future possibilities. This paper evaluates the possibility of AI, specifically GPTs, being able to act as judges in US criminal courts in the near future by assessing how effective AI can be in deciding court cases. The paper examines how AI answers currently lack validity and reliability, two key characteristics of judicial decisions, and analyzes the extent to which AI developments in the near future will be able to address these flaws so that AI may be able to produce viable judicial decisions by themselves. This research concludes that barring unforeseeably significant technological advancements, AI cannot independently act as impartial judges within a US criminal court; however, feasible developments in AI's reliability and validity in the near future would allow AI to work in a complementary capacity alongside human judges to help improve the current judicial system.

KEYWORDS: artificial intelligence, bias, ChatGPT, criminal justice, impartiality, judiciary, judgment

1. Introduction

Judicial integrity, impartiality, conscientiousness, and a sense of justice are seen as the core values that a judge in the United States is expected to uphold and apply in the realm of the law (Central Principle of Judge 2020). However, swaths of studies in the 21st century point to a general violation of these principles: a report published by the Council on Criminal Justice in 2020 found that "blacks are still serving longer sentences than whites" (Same Crime, More Time 2020) while in the same year scholar Naci Mocan found that "the arrest rate of Blacks for murder and nonnegligent manslaughter... is more than six times higher in comparison to the arrest rate of Whites for the same crime." (Mocan 2020, 97-113). Eliminating these biases is a significant concern for the American criminal justice system to maintain its integrity and, consequently, authority.

With the recent amounting interest in the successful use of Artificial Intelligence (AI) to help with everyday tasks that had previously required a human touch, possible alternative applications of AI in various fields are also being suggested (Taffel 2021, chapter 9). While a full substitution from human judges to AI ones comes as abrupt to many, it is a consideration that the general public and lawmakers are contemplating, with an unfounded trust that AI are less biased than humans (Pew Research Center 2023).

However, researchers have pointed out the lack of fundamental skills for AI, rendering them inapplicable for the high-risk decision-making roles of judges (Cremer and Narayanan 2023). They conclude that the current state of AI does not meet the qualifications of a judge, as AI lack the ability to reason morally, give logically concluded answers, and suppress prejudices (Parsons 2024). With the credibility of AI judgments coming into question, it is clear that AI in their current state cannot supplant human judges. However, the fact that AI are ill-suited for the job today does not mean they will forever remain so. It is essential to consider whether further development of AI could eventually make a system that is effective in deciding future court cases in a just manner, and if it is possible, to what extent in the US judicial system.

To answer this question, two fundamental criteria for credibility—reliability and validity—of AI will be scrutinized. As these two elements are also key components of judicial

decision-making (What Makes a Good Judge? 2024), this research paper will analyze how AI could be programmed to satisfy these elements to conclude the future suitability—or lack thereof—of AI's role as judges. In considering reliability (Collins 2008), the consistency and the creativity of AI will be considered, and for validity (Goldman 1982, 112-24), AI's morality, impartiality, and overall accuracy will be examined (Goldman 112-24). Based on this analysis, it can be concluded that AI are largely limited in regards to reliability and validity, but further development can improve, yet not entirely eliminate, these barriers that would make the role of a judge difficult for AI.

2. What is AI?

Artificial Intelligence (AI) are machines that mimic human intelligence (Xu, Liu, and Cao 2021). For the purposes of this research paper, the term “AI” will refer to a subset of these machines, called Generative Pre-trained Transformers (GPT) (Patwardhan, Marrone, and Sansone 2023). GPTs have dominated the market since 2018, following its release in 2017 (Patwardhan, Marrone, and Sansone). The vast majority of AI that are currently used by the general public are GPTs, as evinced by ChatGPT from OpenAI and Gemini from Google, formerly known as Bard; as of August 2023, these AI have amassed a total of roughly two billion users (Grant 2023).

Due to the overwhelming prevalence of GPTs, considering other AI models that were more popular in the past would be ineffective and untimely when answering the research questions. Furthermore, as GPTs are regarded as a superior version in simulating human intelligence, it is necessary to narrow down the broad definition of AI to GPTs to truly question the extent to which AI can make valid and reliable conclusions in the near future (Patwardhan, Marrone, and Sansone 2023).

With the advent of GPTs, the common capabilities of AI have expanded from convergent thinking—referring to past instances to answer a query—to divergent thinking—generating novel solutions and ideas as a response (Jagarlapoodi 2023). In other words, AI are somewhat able to generate original, reasonable answers for open-ended questions, not only close-ended ones (Davis 2024).

These improvements have resulted in significant applications of AI outside the technological industry, including in the legal realm. In the last two decades, various lawyers and law firms have begun to use AI to enhance research and produce preliminary answers for court documents (Tech Start-up Shake World 2016). However, that is not to say AI have fully replaced or can fully replace human expertise within the legal industry as of now. Raj Goyle, a Harvard Law graduate who consults legal companies regarding technology usage, comments that AI still somewhat struggle in answering open-ended questions, and that “[t]he future is coming, but it will not be as fast as some predict” (Lohr 2023). This is best evidenced by how a large portion of the legal sector and certain influential occupations, such as prosecutors and judges, have yet to use AI, even in the research process (Report on Federal Judiciary 2023). Overall, while AI is currently useful in helping complement human research, it finds difficulty in providing full legal statements with logic on par with professional lawyers, prosecutors, and judges. Thus, while those in the industry predict that the future of AI competence approaching human expertise is conceivable, it is still far from the stage of completely supplanting the human mind (Report on Federal Judiciary).

3. AI Judgments

AI can complete tedious tasks—tasks that generally require manual labor—much more efficiently than humans for the most part. While fear of automation and consequent loss of jobs are often correlated to this increasing quality and presence of AI, the implications of AI have not been as devastating in the legal industry (Smith 2017). Such anxiety regarding potential unemployment is absent, as the current state of AI is unable to compete with human lawyers (Smith 2017). Even if AI are used, they are strictly limited to researching and producing precursory suggestions (Atrey

2023). While the reason for this finite use can be partially attributed to the underdevelopment of AI that would inevitably be relieved, underlying that broad generalization are fundamental errors concerning AI judgments that may continually bar AI from extended usage (Anderson 2018).

3.1. Flaws to AI Judgments: Reliability

In terms of reliability, a major flaw of AI judgments is that the vast majority of AI-produced answers will not be consistent—even for the same question—as AI are programmed to generate creative answers via their temperature system (Allan 2023). AI are tailored toward following commands; thus, they are adept at finishing tasks that require simple manual labor or repetitive actions. However, this effectively means that it is difficult for AI to answer prompts that require more complex processes and an element of self-judgment, such as prompts that do not have an expected answer. Answers that are non-stereotypical and require a complex logical process, such as that of creative, original answers, cannot be produced by simplified, rote orders (Powering Personal Wealth Movement 2024). But in that case, why does the current state of AI seem to be able to have human-like conversations without any blaring discrepancies (Conversational AI | Google Cloud 2024)? To answer this question, one must first consider the types of answers that AI can produce. AI answers can be commonly divided into two groups based on how they are formulated. Convergent thinking refers to the ability to produce specific, fixed answers, while divergent thinking refers to the ability to create novel answers (Wigert, Murugavel, and Reiter-Palmon 2022).

The most obvious way in which convergent thinking applies to AI is with questions that have preset answers. All AI has to do here is produce a definitive answer based on their database (Yenduri et al. 2023). Alternatively, questions that seem to require creative thinking can also be answered by AI with convergent thinking (Sharma 2023). For open-ended questions, AI will simply regurgitate answers that are widely credited rather than creating a purely original answer (Dixon 2023). Thus, though the question may seem to require divergent thinking, especially when asked to other humans, for AI, many of their answers are, in truth, born from convergent thinking (Dixon 2023).

But because AI attempt to mimic human intelligence, programmers began to ponder how to incorporate creativity in AI. Creativity, in this case, is referring to the ability to produce non-stereotypical answers, and is necessary due to the possibility of unprecedented and unexpected queries; with a plethora of questions being asked each day, programmers cannot ensure that the training dataset covers all the content being inquired about. As such, while training datasets can be the starting point for generative AI, they cannot simply rely on their training datasets for all the answers (Paullada et al. 2021).

This is where temperature comes in. Temperature is a spectrum ranging from zero to effectively infinity but officially one, controlling the degree of creativity or freedom, or the extent to which an atypical answer is produced (API Reference - OpenAI API 2024). AI's extent of creativity begins with the AI assigning a token for each word, determining which is most probabilistic to come after a given word based on the context of the prompt: the higher the token, the higher the probability (Tupper 2023). Here, the probabilities are determined by the vast amount of text data the AI were trained on (Yenduri et al.) Thus, AI creativity is simply the sequencing of implausible words.

This attempt at mimicking originality via creativity does superficially seem to work. AI like ChatGPT attempt to balance consistency with creativity with OpenAI assigning a default temperature of around 0.7 (Steele 2023). Having studied massive amounts of text, AI can create answers that seem meaningful and novel by using words that sound like they should be there. However, a deeper look into the writing reveals that the superficial pomp does not have much comprehensible meaning. This phenomenon in which AI provides a plausible-sounding answer without any meaning has led to AI being called stochastic parrots (Bender et al. 2021).

In fact, the World Economic Forum describes the difference between this “AI creativity” and human creativity, or originality, as a computational, systematic one versus a more

impulsive one, respectively (Preece and Çelik 2023). While human imagination has no guideline, AI creativity follows a set of orders, closer to convergent thinking than divergent thinking. (Preece and Çelik). Given that AI creativity is systematic and has little reference to meaning, it amounts to nothing more than arbitrarily producing less likely answers. In this sense, they can only produce, not create (Buriak 2023). This means that in the attempt to produce novel answers, AI can end up producing completely nonsensical answers.

Furthermore, dialing down the temperature to produce more probable answers does not always make the AI consistent. In cases where words have tokens of the same probability, AI randomly chooses one. Recent research published by The Association for Computing Machinery exemplifies this issue as it concludes that AI's answers are produced not only without reference to meaning but also in a volatile and random manner (Bender et al.). For example, when discussing a morally controversial topic, AI is generally fed information from both perspectives. If a human were given approximately equal support on both sides, disregarding their personal opinion on the matter, they would likely claim that there are valid arguments for both and assume a neutral stance. AI trained on the same information has completely different implications. Instead of being neutral, this means that AI, 50% of the time, will produce answers supporting one side, and the other 50%, the other side (Yoo et al. 2024), because it assigns the same token value for both sides (Kalyan 2023).

This example exposes both the lack of consistency and originality in the current state of AI. The “creative” component here is that it does not have a set answer. Hence why, in AI, consistency is incompatible with originality—AI's creativity stems from its inconsistency (Tupper). When applied to the context of a criminal court, the effects are devastating. If the evidence presented by both the prosecution and defendant are calculated to have the same probability of being true, the court's decision as decided by an AI system would effectively become a game of chance (Admin 2023). Additionally, the lack of originality also affects AI's ability to replace a judge. The discrepancies in the context for a precedent case and the court case it is dealing with at the time may be overlooked by an AI, as the AI cannot determine what circumstances must be taken into account if it has no previous input on that specific aspect. However, these details may be crucial in deciding the guilt or innocence of a defendant, or the amount of sentencing given to a convicted criminal.

Unfortunately, under the context of AI as a potential replacement for judges, the lack of consistency and creativity makes AI inadequate, as judicial decisions require both. The inconsistency, which damages the reliability of the decision, thus hurts the credibility of AI judgments. And the lack of creativity inhibits complex thinking for understanding differences between cases. Therefore, AI judgments cannot be considered viable in the current state.

3.2. Flaws to AI Judgments: Validity

In addition, AI judgments are flawed in that they fail to meet the criteria that make the decisions of judges valid: morality, impartiality, and accuracy (Wendel 2012). Morality is conceptually foundational to the criminal justice system, but it is also a crucial aspect in the courts; judges are to consider the morality of the defendant's crime and take into account extenuating circumstances to decide upon the punishment via the length of the sentence (Sadurski 1987). However, AI lacks a clear moral compass, as exhibited by a study conducted by Joseph Austerweil, a psychologist at the University of Wisconsin-Madison. In the study, Austerweil asked a variety of moral dilemmas to an AI that gave responses that did not display any sort of moral pattern, jumping between a variety of answers for virtually identical questions (Metz 2021). Consequently, it can be concluded that AI is unable to effectively evaluate the extent of an action's immorality.

Another large concern when it comes to validity is the existence of biases. The datasets AI are trained on significantly affect the answers they produce. Based on their data, AI determine the next word based on probability. Thus, if an AI is trained on documents that lean toward a certain opinion, it is more probable that the AI favors one answer over the other

(Buranyi 2017). Furthermore, the mere proportion of datasets, such as the discrepancies between how many documents for each cohort AI consume, can also lead to preferences (Cook 2023). The relative underrepresentation of certain groups has been notorious for causing biases in a variety of instances, such as that of the Gender Shades facial recognition evaluation project. The project illustrated that the facial recognition systems exhibited weaknesses in detecting certain sexes and races based on their training data (Buolamwini and Gebru 2018). It concluded that a lack of training data for a specific sex or race would cause problems with accuracy and misrepresentation.

Additionally, besides systemic errors, human error regarding how AI are trained can also feed into the answers AI produce (Manyika, Silberg, and Presten 2022). Specifically, the biases that programmers have will be directly reflected in AI via the reinforcement stage (Smith, Khojandi, and Vasudevan 2023). The reinforcement stage refers to when programmers train AI to prefer certain answers or datasets over others. This enforcement has been most useful for questions that have obvious answers, such as “Murder is wrong.” However, for more controversial issues, if AI are reinforced to answer in a way that only represents one side, then the AI will continuously produce one-sided answers.

When extended to AI judgments in the legal field, these problems can be devastating. As judges, AI may deliver a verdict of guilty more often or sentence more years for certain races than others just because of unrepresentative or biased datasets. Because AI answers are based on probability, the quantitatively minor differences in how many datasets of each group the AI has been fed may lead to substantial differences in not only the final decision and sentencing but also the reasoning and explanation it produces. Thus, the bias that human judges have shown in the past will be reflected in the decisions AI judges make in the present, in the same way.

Alternatively, AI can be biased based on the framing of the question. How the question is worded may significantly affect how the AI answers a certain prompt (Popescu 2023). Instead of semantically dissecting a question, AI only considers what relation each word in the prompt has with other words. Thus, two questions that contain identical content that are only worded differently may cause AI to favor one answer over the other due to different values being assigned to the tokens of each word (Research Guides: Chatgpt AI). Thus, based on how the prompt is framed by each court in each trial, an AI system acting as judge may produce different verdicts or sentences, even when the cases mirror each other in every significant way.

These biases of AI models are further exacerbated by the assumptions surrounding accuracy. In the current state of evaluating AI performance, many value “overall accuracy”—the extent to which AI generates correct answers in general (Keylabs 2024). However, this criteria overlooks differences in sub-populations and sub-sectors (Schwartz et al. 2021). An AI can have a near-perfect overall accuracy but lack preciseness within the general group. While this does not mean AI directly reflect biases, a joint study about AI used for medical diagnosis revealed that the reliance on overall accuracy may cause bias in those that interpret the results of inter-population research (Forde et al.). For instance, the study found that despite the 94% overall accuracy of their first experimental group, Algorithm 1 or A₁, in identifying skin cancer, there were large disparities between the accuracy for men, who had 99%, and women, who had a much lower 89%. Furthermore, out of the three algorithms that showed different overall and sub-population accuracies, a significant number of those surveyed favored the algorithm with the highest overall accuracy, even though its accuracy for women was lower than others. Within a legal context, this means that an AI judge that is known to have a high rate of accuracy may have a higher rate of inaccuracy when referring to facts of certain populations in comparison to others.

The absence of an innate sense of morality, the ingrained bias within programmers and existing datasets, and the misleading nature of its reported accuracy all work to hurt the validity of AI judgments when compared to human judgments. As a whole, the problems with their reliability and validity ultimately bar AI in their current state from taking the role of judges in

a criminal court. However, knowing these flaws exist, could the rapid ongoing development of AI systems soon lead to an improved AI that has fixed all these flaws?

4. Addressing the Flaws

Although future development of AI is the main topic of discussion, the occurrence of unforeseeable breakthroughs in coding and algorithms will not be under consideration, as they cannot be predicted nor analyzed in a reliable manner by this author. Rather, this section will discuss specific and concrete changes that can be implemented in the imminent future to nullify some of the aforementioned issues.

4.1. Changing the Approach of AI Developers

A large part of the problems centering around AI can be reduced through a change in the approach taken by the programmers who develop AI. AI developers need to be able to recognize and acknowledge the issues that stem from AI and actively work to negate the harm that could be derived from them. One of the largest yet simplest problems to address is the biases that stem from the datasets the AI are trained on. The obvious solution would be to tailor the content of the dataset used to train AI so that it can be more equally representative, as well as make sure that the information that is additionally reinforced does not mislead users or perpetuate harmful stereotypes. But a more concrete solution that can be applied immediately is to go public with what the AI is trained with and how the AI is trained. Transparency with both the training data and the training method not only prevents users from blindly assuming their information is accurate, but also helps developers figure out what part of their AI training needs to be amended. A common criticism that the leading AI systems, Gemini (Wiggers 2023) and ChatGPT (Iyer 2023), receive is that they do not release the training data to the public. This means that these datasets cannot be reviewed for potential biases caused by unrepresentative training data or inappropriate reinforcement. Being transparent with the training data could facilitate more peer review to help catch these errors.

Another change necessary in the AI creators' approach is to recognize the problems associated with technological solutionism. Technological solutionism—the idea that advancements in technology can solve societal problems—permeates AI developmental settings, such as the Silicon Valley (Schwartz et al.). Combined with the tendency to treat ethics as a technical problem (Moss and Metcalf 2020), this belief fosters blind faith in AI. On the developers' end, it results in overlooking possible biases or systemic errors, and on the consumers' end, it leads to significantly lower skepticism (Moss and Metcalf 2020). Also, in the case that AI is considered a complementary source to a human judge, a proper check and balance would be difficult due to the human judge regarding the AI as invariably correct.

Because technological solutionism is an ideology, steps to curtail its influence will not lead to an immediate eradication of the belief, but will help more people acknowledge its problems. Public awareness campaigns should emphasize—especially to developers and users in high-stakes settings—how unreliable and invalid AI may be. Changing this attitude is necessary, as the persistence of technological solutionism will only result in shallow, insufficient answers being accepted as true, which would only exacerbate existing issues.

4.2. Changing AI Programming

Another flawed area to be addressed are the workings of AI systems themselves. AI can be induced to produce answers that counter problems of overall accuracy and inconsistency. Overall accuracy can be problematic aside from misrepresentation and disparities between sub-groups because the simple calculation of accuracy, which divides the number of times the AI was correct by the total number of answers generated, overlooks how different answers hold different values in the real world. For instance, in a legal setting where AI is acting as a judge, an incorrect verdict is considered a single incorrect verdict regardless of the context. However, there is a significant difference between an inaccurate verdict that sentences an innocent person to prison versus that which does

not correctly sentence a criminal. The former incurs a far higher social and moral cost than the latter (Halvorsen 2004). But in reality, AI see both cases as equivalent: it is, thus, conceivable that two AI with an overall accuracy of 90% is seen as equal, even if the first AI has a 10% inaccuracy of convicting the innocent while the latter has a 10% inaccuracy of not convicting the guilty.

One way to address this issue is to specify not only the overall accuracy of the AI system in use but to also display the components of which this overall accuracy consists of. For example, for use in the courts, AI should show how its accuracy rates for sentencing and conviction differ for different ethnicities and genders. The accuracy rates should also separately list the likelihood of pronouncing an innocent person guilty and the probability of not convicting a guilty person. With all this information available, AI users will be able to make an informed decision about how to qualify the answer produced by AI. In some settings, an AI that has a higher overall accuracy but a large disparity between assessing groups may be much more problematic than an AI that has a somewhat lower overall accuracy but produces consistent answers to all groups being assessed.

To address the problem of inconsistency in AI, users can lower the temperature of AI, programmers can reinforce specific answers, and for prompts that do not have an overwhelmingly probable answer, all possible answers could be listed for review. If the case in question has a clear-cut conclusion—meaning that the circumstances of this case mirror a precedent in all ways that are relevant—then users can lower the temperature of the AI so that it will choose the most probable answer, or programmers can reinforce the AI to produce a set verdict for those types of cases. Then, the AI will invariably generate the same answer. Alternatively, if the case in question has elements that do not match the circumstances of previous cases in the database, then developers can program AI to generate a comparative answer that displays all the results the AI is considering. Instead of providing one of the choices as an answer, displaying all the choices it considers in addition to the probability of each can help eliminate the degree of randomness and ensure that all perspectives are considered. For this method to work, AI need to be working as a complementary tool for a human judge, as the human judge will be the one to make a final decision. Having summarized much of the information and research necessary to reach a decision, the AI would assist the human judge to more thoroughly review the circumstances and make an informed judgment.

5. Conclusion

In the process of evaluating whether AI can act as judges with the goal of eliminating biases that currently plague the US criminal justice system, this paper has found that AI in their current state cannot fully replace human judges because of their lack of reliability and validity. Further development can partially address those flaws, but cannot eliminate them entirely, at least in the near future. Furthermore, regardless of the improvements suggested, AI still fundamentally lack in morality and logic and many other sectors when compared to a human judge. Even if AI does coincidentally produce an answer that successfully mimics a decision that a human judge would make, sufficient both in breadth and depth, the process of arriving at that conclusion—void of any ethical or semantic consideration—effectively nullifies its legal authority. Because these problems stem from the foundation of AI programming, they cannot be resolved unless future breakthroughs find a way for AI to mimic the whole thought process of the human brain. However, the incapability of AI to replace human judges does not preclude the possibility of AI being useful in other ways within the judicial courts. Because of the immense amount of information AI can command in a moment, human judges can use AI to complement them and make more fully informed decisions. With the aforementioned improvements, AI can produce preliminary answers that can in turn be evaluated by human judges to make a final decision. This complementary usage can still greatly improve the efficiency of court procedures—upholding a defendant’s right to a speedy, fair trial (US Court of Appeals 2021).

Due to technology being a lucrative sector, many look forward to progress and optimal functioning, seldom looking back on the societal issues the development implicates. Before more technological advancements infringe on parts of human society that are not fully prepared to deal with the fallout, more consideration of their potential is necessary, and regulations need to be set accordingly, as is shown with this case study on the judicial branch.

References

- 2023 Year-End Report on the Federal Judiciary, 2023. <https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf>.
- Admin. “The Importance of Judicial Consistency.” Mortons Solicitors, April 15, 2023. <https://www.mortons-solicitors.co.uk/the-importance-of-judicial-consistency/>.
- Allan, Kodundo. 2023. “CHATGPT: A Natural Language Generation Model for Chatbots.” ResearchGate, 2023. https://www.researchgate.net/publication/366893580_ChatGPT_A_Natural_Language_Generation_Model_for_Chatbots.
- Anderson, Janna. 2018. “Artificial Intelligence and the Future of Humans.” *Pew Research Center: Internet, Science & Tech*, December 10, 2018. <https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/>.
- Atrey, Ishan. 2023. “Revolutionising the Legal Industry: The Intersection of Artificial Intelligence and Law.” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4632440>.
- Bender, Emily M., Timnit Gebru, Angelina McMilian-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.” *ACM Conferences*, March 1, 2021. <https://dl.acm.org/doi/10.1145/3442188.3445922>.
- Buolamwini, Joy, and Timnit Gebru. 2018. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” *Proceedings of Machine Learning Research* 81 (2018): 77–91. <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.
- Buranyi, Stephen. 2017. “Rise of the Racist Robots – How AI Is Learning All Our Worst Impulses.” *The Guardian*, August 8, 2017. <https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>.
- Buriak, Jilian. 2023. “Best Practices for Using AI When Writing Scientific Manuscripts.” *ACSPublications*, <https://pubs.acs.org/doi/10.1021/acsnano.3c01544>.
- California Courts. 2013. “The Central Principle of Being a Judge.” https://www.courts.ca.gov/documents/culture_handout3.pdf.
- Collins, Paul M. 2008. “The Consistency of Judicial Choice.” *The Journal of Politics*, <https://doi.org/10.1017/s002238160808081x>.
- Conversational AI | Google Cloud.” Google. Accessed March 25, 2024. <https://cloud.google.com/conversational-ai>.
- Cook, Alexis. 2023. “Identifying Bias in Ai.” *Kaggle*, April 20, 2023. <https://www.kaggle.com/code/alexisbcook/identifying-bias-in-ai>.
- Cremer, David De, and Devesh Narayanan. 2023. “How AI Tools Can—and Cannot—Help Organizations Become More Ethical.” *PubMed Central*, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10324517/>.
- Davis, Mark. 2024. “How Does CHATGPT Understand Context? 2024 Update.” *ChatGPTDetector*, March 17, 2024. <https://chatgptdetector.co/can-chatgpt-understand-context-and-maintain-conversation-flow/>.
- Dixon, Brendan. 2023. “The Underwhelming Creativity of Ai.” *Mind Matters*, May 18, 2023. <https://mindmatters.ai/2019/05/the-underwhelming-creativity-of-ai/>.
- Engstad, Nils. 2017. “Consistency of the Case Law as a Prerequisite to Legal Certainty.” Council of Europe, 2017. <https://rm.coe.int/nils-engstad-president-of-the-council-of-europe-s-consultative-council/1680759572>.
- Forde, Jessica Zosa, A. Feder Cooper, Kweku Kwegir-Aggrey, Chris De Sa, and Michael Littman. 2021. “Model Selection’s Disparate Impact in Real-World Deep Learning Applications.” *Science and Engineering of Deep Learning Workshop*, 2021. <https://doi.org/https://doi.org/10.48550/arXiv.2104.00606>.
- GOLDMAN, SHELDON. 1982. “Judicial Selection and the Qualities That Make a ‘Good’ Judge.” *The ANNALS of the American Academy of Political and Social Science* 462, no. 1 (July 1982): 112–24. <https://doi.org/10.1177/0002716282462001010>.
- Grant, Nico. 2023. “Google Connects A.I. Chatbot Bard to YouTube, Gmail and More Facts.” *The New York Times*, September 19, 2023. <https://www.nytimes.com/2023/09/19/technology/google-bard-ai-chatbot-youtube-gmail.html>.
- Halvorsen, Vidar. 2004. “Is It Better That Ten Guilty Persons Go Free than That One Innocent Person Be Convicted?” *Criminal Justice Ethics* 23, no. 2 (June 2004): 3–13. <https://doi.org/10.1080/0731129x.2004.9992168>.

- Iyer, Prithvi. 2023. “New Study Suggests CHATGPT Vulnerability with Potential Privacy Implications.” *Tech Policy Press*, December 2, 2023. <https://www.techpolicy.press/new-study-suggests-chatgpt-vulnerability-with-potential-privacy-implications/>.
- Jagarlapoodi, Dr.Ing. Srinivas. 2023. “Generative Pre-Trained Transformer: Revolutionizing Language Generation and Creativity.” *LinkedIn*, June 30, 2023. <https://www.linkedin.com/pulse/generative-pre-trained-transformer-revolutionizing-jagarlapoodi/>.
- Kalyan, Katikapalli. 2023. “A Survey of GPT-3 Family Large Language Models Including Chatgpt and GPT-4.” *Natural Language Processing Journal*, December 19, 2023. <https://www.sciencedirect.com/science/article/pii/S2949719123000456>.
- Keylabs. 2024. “Improving Your AI Model’s Accuracy: Expert Tips.” Keylabs, January 26, 2024. <https://keylabs.ai/blog/improving-your-ai-models-accuracy-expert-tips/>.
- Lohr, Steve. 2023. “A.I. Is Coming for Lawyers, Again.” *The New York Times*, April 10, 2023. <https://www.nytimes.com/2023/04/10/technology/ai-is-coming-for-lawyers-again.html>.
- Manyika, James, Jake Silberg, and Brittany Presten. 2022. “What Do We Do about the Biases in Ai?” *Harvard Business Review*, November 17, 2022. <https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai>.
- Metz, Cade. 2021. “Can a Machine Learn Morality?” *The New York Times*, November 19, 2021. <https://www.nytimes.com/2021/11/19/technology/can-a-machine-learn-morality.html>.
- Moss, Emanuel, and Jacob Metcalf. 2020. “Ethics Owners.” *Data & Society*, https://datasociety.net/wp-content/uploads/2020/09/Ethics-Owners_20200923-DataSociety.pdf.
- Mocan, Naci. 2020. “Biases in judicial decision making”. In J. Avery & J. Cooper (Eds.), “bias in the law: A definitive look at racial prejudice in the u.s. criminal justice system” (p. 97-113). Lexington Books.
- OpenAI. 2024. “API Reference - Openai API.” <https://platform.openai.com/docs/api-reference/chat>.
- Parsons, Lian. “Ethical Concerns Mount as AI Takes Bigger Decision-Making Role.” *Harvard Gazette*, January 3, 2024. <https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/>.
- Patwardhan, Narendra, Stefano Marrone, and Carlo Sansone. 2023. “Transformers in the Real World: A Survey on NLP Applications.” *Information* 14(4): 242. <https://doi.org/10.3390/info14040242>.
- Paullada, Amandalynne, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. 2021. “Data and Its (DIS)Contents: A Survey of Dataset Development and Use in Machine Learning Research.” *Patterns* 2, no. 11 (November 2021): 100336. <https://doi.org/10.1016/j.patter.2021.100336>.
- Pew Research Center. 2023. “1. Americans’ Views on Use of Ai in Hiring.” *Pew Research Center: Internet, Science & Tech*, April 20, 2023. <https://www.pewresearch.org/internet/2023/04/20/americans-views-on-use-of-ai-in-hiring/>.
- Popescu, Roxana. 2023. “Generative AI in Creative Design Processes: A Dive into Possible Cognitive Biases.” *IASDR 2023: Life-Changing Design*, October 9, 2023. <https://doi.org/10.21606/iasdr.2023.784>.
- Preece, Chloe, and Hafize Çelik. 2023. “Ai Is a Powerful Tool, but It’s Not a Replacement for Human Creativity.” *World Economic Forum*, 2023. <https://www.weforum.org/agenda/2023/06/ai-cannot-replace-human-creativity/>.
- Q.ai - Powering a Personal Wealth Movement. 2024. “The Pros and Cons of Artificial Intelligence.” *Forbes*, February 20, 2024. <https://www.forbes.com/sites/qai/2022/12/01/the-pros-and-cons-of-artificial-intelligence/?sh=2dd5efb24703>.
- “Research Guides: Chatgpt and Generative Artificial Intelligence (AI): Potential for Bias Based on Prompt.” Potential for bias based on prompt - ChatGPT and Generative Artificial Intelligence (AI) - Research guides at University of Waterloo. Accessed March 27, 2024. https://subjectguides.uwaterloo.ca/chatgpt_generative_ai/chatgptbias.
- Sadurski, Wojciech. 1987. “Judicial Morality and the Limits of the Law.” *Jstor*, 1987. [https://www.york.ac.uk/media/law/yorklawreview/vol3/Clothier - Judicial Morality and the Limits of the Law \(1\).pdf](https://www.york.ac.uk/media/law/yorklawreview/vol3/Clothier - Judicial Morality and the Limits of the Law (1).pdf).
- “Same Crime, More Time.” Georgia State News Hub, October 12, 2020. <https://news.gsu.edu/research-magazine/spring2020/incarceration>.
- Schwartz, Reva, Leann Down, Adam Jonas, and Elham Tabassi. 2021. “A Proposal for Identifying and Managing Bias in Artificial Intelligence.” *NIST*, June 22, 2021. <https://doi.org/10.6028/nist.sp.1270-draft>.
- Sharma, Aditya. 2023. “AI Generates Responses, It Doesn’t Create: The Role of Artificial Intelligence in Human Creativity.” *LinkedIn*, September 1, 2023. <https://www.linkedin.com/pulse/ai-generates-responses-doesnt-create-role-artificial-human-sharma/>.
- Smith, Aaron. 2017. “Automation in Everyday Life.” *Pew Research Center: Internet, Science & Tech*, October 4, 2017. <https://www.pewresearch.org/internet/2017/10/04/automation-in-everyday-life/>.
- Smith, Benjamin, Anahita Khojandi, and Rama Vasudevan. 2023. “Bias in Reinforcement Learning: A Review in Healthcare Applications.” *ACM Computing Surveys* 56, no. 2 (September 15, 2023): 1–17. <https://doi.org/10.1145/3609502>.
- Steele, Colt. 2023. “Understanding Openai’s Temperature Parameter.” *Colt Steele*, June 28, 2023. <https://www.coltsteele.com/tips/understanding-openai-s-temperature-parameter>.

- Taffel, Sy. “Communicative Capitalism, Technological Solutionism, and The Ocean Cleanup.” Essay. In *Plastic Legacies*, n.d.
- “The Tech Start-up Planning to Shake up the Legal World.” BBC News, May 16, 2016. <https://www.bbc.com/news/business-36303705>.
- Tupper, Kevin. 2023. “Creatively Deterministic: What Are Temperature and TOP_P in Generative AI?” *LinkedIn*, August 12, 2023. <https://www.linkedin.com/pulse/creatively-deterministic-what-temperature-topp-ai-kevin-tupper>.
- United States Court of Appeals for the Armed Forces. “MISCELLANEOUS MATTERS: Speedy Trial: Sixth Amendment Right to a Speedy Trial.” 2021. <https://www.armfor.uscourts.gov/digest/VB4.htm>.
- Wendel, Bradley. 2012. “Impartiality in Judicial Ethics: A Jurisprudential Analysis.” *Notre Dame Journal of Law, Ethics, & Public Policy*, 2012. <https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=1122&context=ndjlepp>.
- “What Makes a Good Judge?” ACT Courts. Accessed March 24, 2024. https://www.courts.act.gov.au/_data/assets/pdf_file/0003/1911477/Activity-What-makes-a-good-judge.pdf.
- Wigert, Benjamin, Vignesh Murugavel, and Roni Reiter-Palmon. 2022. “The Utility of Divergent and Convergent Thinking In ...” *Psychology Faculty Publications*, 2022. <https://digitalcommons.unomaha.edu/cgi/viewcontent.cgi?article=1281&context=psychfacpub>.
- Wiggers, Kyle. 2023. “Google’s Gemini Isn’t the Generative AI Model We Expected.” *TechCrunch*, December 8, 2023. <https://techcrunch.com/2023/12/06/googles-gemini-isnt-the-generative-ai-model-we-expected/>.
- Xu, Yongjun, Xin Liu, and Xin Cao. 2021. “Artificial Intelligence: A Powerful Paradigm for Scientific Research.” *The Innovation*. <https://doi.org/10.1016/j.xinn.2021.100179>.
- Yenduri, Gokul, Ramalingam M, Chemmalar Selvi G, Supriya Y, Gautam Srivastava, Praveen Kumar Reddy Maddikunta, Deepti Raj G, et al. 2023. “Generative Pre-Trained Transformer: A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions.” *ArXiv.org*, May 21, 2023. <https://arxiv.org/abs/2305.10435>.
- Yoo, Byunghye, Jinwoo Kim, Seongeun Park, Changbum R. Ahn, and Taekeun Oh. 2024. “Harnessing Generative Pre-Trained Transformers for Construction Accident Prediction with Saliency Visualization.” *MDPI*, January 12, 2024. <https://www.mdpi.com/2076-3417/14/2/664>.