

Accessing Legal Risks of Large Language Models: Causes, Mitigation and Regulation Approaches

Dr. Kashif Javed¹ & Usmat Batool²

¹Associate Professor/Chairperson Department of Law, Email: Kashifjaved777@gmail.com

²Assistant Professor, Institute of Islamic studies, Bahuddin Zakariya University, Multan,
Email: usmatbatool@bzu.edu.pk

ARTICLE INFO

Article History:

Received: September 15, 2025
Revised: October 08, 2025
Accepted: October 21, 2025
Available Online: November 07, 2025

Keywords:

AI generative model; Justice, SDG-16; Subjectivity; Deep Fakes; Regulatory framework

Corresponding Author:

Dr. Kashif Javed

Email:

Kashifjaved777@gmail.com

ABSTRACT

AI generative models are in high demand across academic, business, and research realms due to their human-like capacity to create complex reasoning tasks, codes, and solutions. The potential of this tool in promoting justice in society. The current study discusses legitimacy challenges posed by AI-generated models. Employing qualitative methodology and engaging in-depth analysis, this study investigates the complex issues surrounding AI generative models, with a specific emphasis on their subjective nature and the legal risks they entail within judicial systems. Furthermore, it finds that the key contributors to the subjectivity of AI generative models are the training of AI generative models on specific data sets, textual interpretations, ethical considerations, and aesthetic preferences. Based on the findings, the recent study recommends legislation for a universal regulatory framework that includes explainable algorithms, diversified training and a close working relationship environment and continuous dialogue between legal experts and AI developers.



Introduction

Justice is the cornerstone of a prosperous and economically stable society. Thus, Sustainable Development Goal (SDG) 16 on Peace, Justice, and Inclusion notes that “access to fair justice systems is an inherent human right, yet victims of violations do experience unjust treatments and unresolved prosecution processes.” The UNDP (2023) measures this goal, which enables the 2030 agenda, through 12 global targets and 24 associated indicators. It is an admitted fact that artificial intelligence has caught the eye of the judiciary around the globe. This field has become a subfield of every subject, and its most recent innovation is Large Language Models.

Every species on earth uses codes to transfer and catch the feelings and tasks of other species in the same family. Similarly, humans also have the ability to transfer and accept thoughts and

emotions from other humans. All of the species, like humans, use different codes depending on their origin. These codes, if structured properly (grammatically), develop a complete sense of one's thoughts and emotions. Artificial intelligence (AI) involves creating machines that can do human-like jobs. Scientists term these codes "words," and refer to the structure as language. Scientists, since the evolution of AI (Y. Lu, 2019), are trying to teach machines and make their capacities like humans about understanding and mimicking language by using Natural Language Processing (NLP) (Chowdhary, 2020). However, the variety of languages and their tones and slangs are not very accurate if translated into other languages.

John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon coined the term "Artificial Intelligence (AI)" during the 1950s Dartmouth Conference (McCarthy, 1956). AI research first focused on creating software that could perform human-intelligence-requiring tasks like solving difficult mathematical equations or playing chess. It was expanded to natural language processing, perception, and reasoning in the 1960s and 1970s (Nohr, 2019). This led to Expert systems advanced AI in the 1980s (Durkin, 1996). The above software replicated a domain expert's cognitive processes. However, expert systems could not handle unusual situations until machine learning and neural networks entered the field of artificial intelligence in the 1990s. Machine learning lets machines learn without programming (Chen, 1995) and neural networks simulate the human brain. Moreover, Deep learning algorithms, which use neural networks with numerous layers and data, have driven artificial intelligence advances. These advances have improved optical and aural pattern recognition, computational linguistics, and automated equipment.

Large language models (LLMs) are capable of processing large amounts of natural language text data and generating responses that are often indistinguishable from those of humans (Radford, 2019). These models are being used in a wide range of applications, from chatbots and virtual assistants to language translation and content generation. The topic of AI generative has gathered significant attention in contemporary discourse. We can find closed-source and open-source AI-generative platforms. The impact of artificial intelligence on the creation of art, literature, and other forms of content has been significant. Even so, concerns have been raised regarding the potential threat to human labor as AI may eventually replace human workers in various industries. Based on a study conducted by Oxford, it is highly probable that in the foreseeable future, AI generative models will have a significant impact on number of segments including human resource centers, potentially leading to the displacement of current job roles (Julio, 2023). Although the proportion of jobs that will be impacted is relatively low, the consequences will undoubtedly affect individuals, as mentioned in Figure 1, given that the global population has surpassed 8 billion (UN, 2022). People losing their jobs are likely to have a significant social impact. The concentration of wealth is likely to result in adverse consequences for the economically disadvantaged, potentially leading to an escalation in criminal activity and posing a fresh challenge to the rule of law.

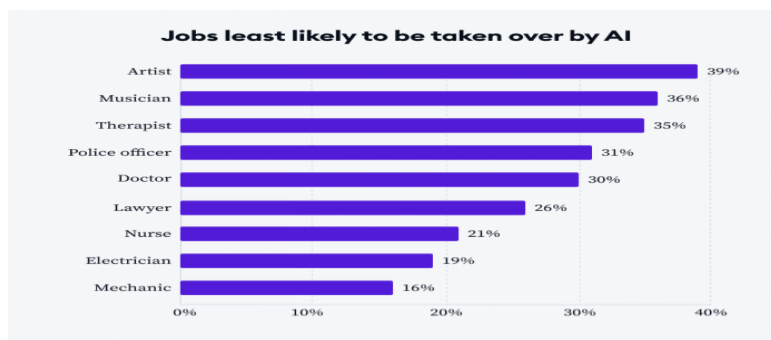


Figure 1: Jobs going to be affected by AI (Julio, 2023)

Interactions with users are an important aspect of any conversational AI system, including large language models like ChatGPT and DALL-E. A pragmatic approach to interactions involves designing the system in a way that is user-friendly, efficient, and effective in achieving its goals (Fjeld & Cawsey, 2020). Because there is such a large demand for these kinds of models in today's modern times. So, the proliferation of these models has the potential to cause social unrest. If a model has intrinsic subjectivity, then it has an extremely high probability of producing results that are biased, discriminatory, and false. This paper clearly talks about the problems with OpenAI's (ChatGPT, DALL-E) and Google's (BARD, IMAGEN) subjective models, such as the deep fakes they make and the general threat they pose to IP rights.

There is evidence that judges from Columbia, Peru, Mexico, India, and Pakistan have used ChatGPT to take advice for judgment (University of Arizona, 2023). However, there are currently no regulation regarding usage or prohibition from utilizing AI-generated models for personal inquiries. It is important to note that reliance on such models may result in the adoption of biased and inaccurate data. Judges who fail to cross-check this information may inadvertently contribute to societal injustice. This article sheds light on the circumstances surrounding generative models, which are supposed to have the possibility of legal consequences. In addition to this, it digs into the current legal conflicts that have taken place between various individuals and the firm. This statement highlights the significance of a complete ethical and regulatory system that incorporates not only the rights of individuals but also the obligations of handlers of the LLM model. This structure should include both the duties and rights of individuals and the responsibilities of handlers. This study was carried out by employing a qualitative research approach throughout its entirety.

Literature Review

Language modeling, known as LLM, is a technique that is considered an important tool for improving the language intelligence of machines. In recent years, language models have become a subject of great scholarly interest (Kaplan, 2020). As Rosenfeld pointed out, language models based on learning approaches to fill in gaps with words emerged in the 1990s (Rosenfeld, 2000). Researchers Thede (1999) employed Statistical Language Models in the domains of natural language processing, task execution, and information retrieval, and later, the field advanced to NLM "Neural Language Models," which use artificial neural networks to learn and predict the probability distribution of words or sequences of words in a given natural language (Bengio, 2003). The evolution of "pre-trained models" signified a further advancement. Context-sensitive word representations were first obtained with Embeddings from Language Models by Matthew E. Peters in 2018. When first presented, this method involved pre-training a bidirectional long short-term memory (bi-LSTM) network to obtain context-sensitive word representations, followed by fine-tuning the bi-LSTM network based on the specific downstream tasks being performed. Finally, research has developed large language models (Kaplan, 2020). The research community frequently uses "large language models (LLM)" to refer to the big-sized pre-trained language models (PLMs) (Taylor, 2022). AI employs electromagnetism to evaluate textual input and provide findings similar to those of humans (Floridi, 2013). In figure 2, Graph below, as per online survey, the high percent bar shows the believe of people that AI can revolutionize their lives while small bars represent the people have no idea about this technology.

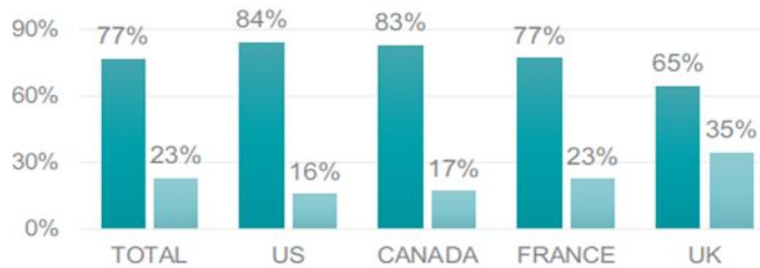


Figure 2: People believe about AI revolution in their life

The ChatGPT series is an achievement of Open AI (2023). This series of interface models is called the Large Language Model series. This model was trained on the model "a sibling model to InstructGPT". This model is the same as the conversation model used by humans. Reportedly, "human-generated conversations (playing both the roles of user and AI) are pooled with the InstructGPT dataset in a dialogue format for training ChatGPT. Later on, the plugin mechanism has been supported in ChatGPT, which further extends the capacities of ChatGPT with existing tools or apps." The latest model, ChatGPT-4, has extended multimodal signals and is experienced in solving diversified and challenging tasks (Bubeck et al., 2023). ChatGPT demonstrated remarkable aptitude in engaging in human communication, characterized by a comprehensive repository of information, adeptness in logical reasoning pertaining to mathematical problems, adeptness in accurately tracking the discourse context in multi-party conversations, and alignment with human values, thereby ensuring the ethical and secure utilization of the technology, as shown in Figure 3.

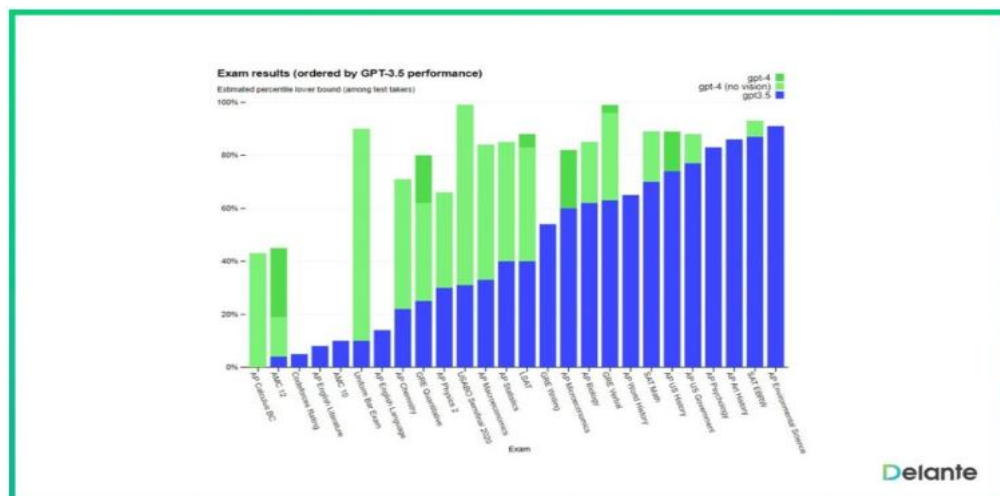


Figure 3: ChatGPT passed different institutional exams i.e. Bar exams (Delante ,2023)

The implementation of ChatGPT is expected to have significant implications for the domain of artificial intelligence, shedding light on the pursuit of AI models that emulate human-like characteristics. Additionally, a six-month iterative alignment process with an additional safety reward signal in the reinforcement learning from human feedback (RLHF) training has improved the security of GPT-4's responses against malicious or provocative inquiries. In accordance with the technical report, measures are being taken to mitigate the proliferation of deleterious or toxic material (Ganguli et al., 2023). OpenAI's GPT-3.5 architecture serves as the foundation for this system. ChatGPT can be used with anything from a chatbot to a message service to a voice assistant. By using these interfaces, users can have natural-sounding conversations with ChatGPT, asking questions and getting answers in the same vein (Lee & Lee, 2022). ChatGPT interprets text

using natural language processing, machine learning, and deep learning. It can recognize patterns, identify word-phrase associations, and respond to input text. However, ChatGPT's understanding of text is exclusively based on its training data and algorithms. Its responses are typically contextually suitable and grammatically correct, but it is neither sentient nor conscious. GPT-4 uses a robust deep learning framework and advanced optimization methods. Enhanced LLMs can still cause hallucinations with factual mistakes or dangerous reactions. LLM safety and capability research is underway. OpenAI uses a five-stage development and deployment lifecycle to develop models and products iteratively to reduce deployment risks (Open AI, 2023). while DALL-E is an image model. DALL-E is built to aid customers in a number of ways, including through the use of chatbots, messaging platforms, and voice assistants (Brown et al., 2020). Both models are capable of learning and adapting to new scenarios over time, and they can understand and respond to input in natural language. Chatbots, voice assistants, and other conversational interfaces (DALL-E, 2023)

In 2022, Google succeeded in introducing the BARD conversational agent with a large language model (Bard, 2023). It has been trained on a substantial corpus of code and text to acquire the ability to generate text, translate languages, produce diverse creative content, and furnish insightful responses to inquiries. Its base is the transformer framework, a highly effective transformer model known for translation and question answering (Vaswani et al., 2017). This is a kind of distributed computing system comprised of multiple interconnected computers operating in tandem. Every individual server retains its own database, comprising textual and coding information, along with a replicated version of the Transformer model. A high-speed network interconnects all of the servers in the system. The BARD system is equipped with a user interface that is accessible through the internet. The user interface is characterized by its simplicity and ease of navigation (Bard, 2023). Bard responds to user requests in seconds. Its versatile quality automatically translates text into multiple languages and answers questions with explanation. It is capable of generating code in several languages, along with writing poems, code, screenplays, music, emails, letters, and such other tasks. It is still under development and might revolutionize computer interface because it has the capacity to replace humans in structured and unstructured data analysis, labeled and unlabeled data customer service, medical diagnosis, and legal research (Bard, 2023). Google AI's Imagen text-to-image diffusion model is a vast text and image collection trained as a large language model. It can create lifelike visuals from complex or odd language descriptions by encoding the written descriptions into integers (IMAGEN, 2023). A vast text and image collection trains the language model. This helps the language model understand text-image relationships.

Diffusion models gradually blur images by adding noise by backpropagation, which also optimizes model parameters. Diffusion model weights are the parameters that conclude the image generation. Imagen lets users develop models from current data sources for applications, insights, and predictions. It offers more AI model-building tools. On the other hand, Open AI is confined to synthesizing images from text captions and also lacks model management and deployment tools. AI language models can write creatively based on the patterns and relationships they learned from the massive volumes of data they were trained on, generating innovative and unexpected responses. Like ChatGPT, it may generate new words, phrases, and concepts that are contextually appropriate and grammatically correct but not in the training data. Based on its comprehension of concepts and word relationships, it may be able to respond creatively to prompts or inquiries. However, training data and algorithms limit AI language models' inventiveness. They lack the subjective experiences and feelings that drive human creation; hence, they lack actual creativity and uniqueness. AI language models respond using statistical patterns and associations from their training data.

Research Methodology

Specifically, this investigation makes use of the qualitative research methodology, which comprises conducting an in-depth analysis of the previously published literature. When it comes to the field of scientific study, this methodology is widely acknowledged and held in high regard. Within the context of this system, the accomplishment of the ultimate goal is directly proportional to the sequential execution of a collection of independent tasks. The literature review is a series of activities that takes place with the primary purpose of collecting and analysing scientific papers that are associated with a particular topic area in order to provide answers to particular questions. In order to address particular concerns, this strategy necessitates the identification and evaluation of all publications that are pertinent to the research at hand. In order to carry out this research, we adhere to a structured approach that includes introducing the topic, providing its background, outlining the research objectives and goals, evaluating the existing literature to identify the limitations and constraints of LLM, and identifying reasons and proposals for its improvement. All of these steps are necessary in order to carry out this research.

Research Hypothesis

Use of LLM in different field is posing legal risk to fundamental rights of the people.

Research Questions

- a- What are the limitations of AI generative models?
- b- What kind of legal risk do these limitations of AI generative models create?
- c- Which factors contribute to posing legal risks because of AI generative models?
- d- How can these challenges be mitigated?
- e- What kind of approaches are being adopted by developing and developed countries?

Research goal

The ubiquity of LLM across all domains is apparent. Companies employing both open-source and closed-source models are competing to develop LLM in order to maximise their earnings. These IT behemoths disregard regulations and are actively violating existing restrictions. Their motto appears to be that if a lawsuit is launched against the corporation, they will pay the settlements at a later time. The research objective is to uncover the constraints and difficulties encountered by LLM. It also aims to reveal the elements that contribute to difficulties. Furthermore, its objective is to propose strategies for alleviating the difficulties and surmounting the constraints.

Limitations and Challenges of LLM

Hallucinations

While Large Language Models (LLMs) are making significant progress in creating text that resembles human writing, a major obstacle still exists in their inclination to produce stuff that seems accurate but has a solid foundation. The problem of hallucination is likely the most significant obstacle to the safe implementation of these strong LLMs in real-world production systems that have an impact on people's lives (S.M Towhidul Islam Tonmoy et al., 2024). The presence of hallucination, which refers to the inclination of generating content that is factually inaccurate, is a significant obstacle to the reliable and trustworthy implementation of large language models (LLMs) in real-world applications (Junyi Li, et al., 2024). It has been noted that while these models are capable of producing language that appears natural and fluent, they often lack accuracy and may include irrelevant or unsupported facts/information. An analogous occurrence is noted in end-to-end speech recognition systems, wherein a segment of the resulting

text exhibits distinct acoustics in comparison to the matching audio signal (Ashish Mittal et al., 2024). Contrary to conventional AI systems that concentrate on specific tasks, Hence, it is evident that LLMs have been exposed to extensive quantities of internet text material during their training. Although this enables them to showcase their remarkable linguistic proficiency, it also implies that they can derive information from the biases present in the training data, misread ambiguous cues, or alter the information to superficially conform to the input. This becomes highly concerning when we depend on the powers of language generation for sensitive applications, such as condensing medical information, analysing financial reports, and examining legal cases as depicted in figure 4.

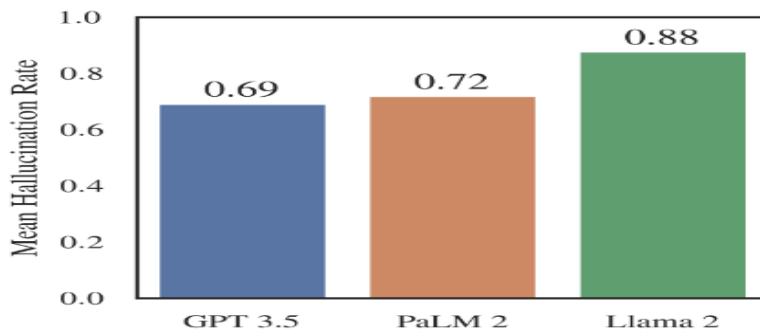


Figure 4: Hallucination Rate

Algorithms' Black box nature

In the first place, we contend that the black box problem is, in reality, a superficial one since it is the outcome of an overlap of four distinct problems that are, however, interconnected with one another. These problems are the opacity problem, the strangeness problem, the unpredictability problem, and the justification problem (Brožek, et al., 2023). It is becoming increasingly normal practice to use algorithms in the process of evaluating the possible risk posed by defendants who are awaiting trial in order to decide whether or not they should be released into the community while they are awaiting their trials. When compared with their white counterparts, black defendants have a considerably higher possibility of displaying characteristics that lead to an incorrect classification as high risk. This is the case in some situations. In recent years, a variety of approaches have been proposed in an effort to achieve algorithmic equity and rectify the inequities that have been identified (Corbett-Davies, 2017).

Algorithmic discrimination occurs when computers choose sides when making decisions. If the AI system's data is skewed or the algorithms utilized to generate decisions are biased, this can happen. Algorithmic discrimination could worsen social and economic inequality, raising ethical concerns. In the employment process, an AI system educated on discriminating data may ignore qualified candidates from underrepresented groups. The predictions about individuals regarding their personalities are neither intuitive nor provable by the use of big data analytics and artificial intelligence (AI). These deductions rely on heterogeneous, feature-rich data whose value is hard to forecast, opening the door to potentially discriminating, prejudiced, and intrusive outcomes (Mittelstadt, 2016). The extent to which algorithms exert influence on individuals and society, particularly in terms of transparency, has been a topic of significant interest in understanding their potency (Diakopoulos, 2014). When algorithms are utilized to make decisions and their inner workings cannot be comprehended or elucidated, this occurrence is commonly referred to as the "black box effect" of algorithms (Tschider, 2020). The study of algorithms as a mathematical problem may not present any issues, as it primarily involves practice and the formulation of hypothetical assumptions. Conversely, if these algorithms serve as the foundation for decisions

that significantly impact human life, there are apprehensions regarding the fundamental principles of life, as life necessitates a thoughtful approach to addressing its issues. In health-related applications, it is imperative to avoid making assumptions that lack a clear future, given the use of algorithms (Ferdous et al., 2020). The utilization of AI-generated applications has become widespread among individuals who extract information and generate content for their personal use. However, the lack of transparency regarding the algorithms employed in these LLMs increases the likelihood of producing fraudulent, biased, and illegal outputs. This warning is prominently displayed on the front page of these applications' portals. It is worth contemplating the potential gravity of the black box effects of algorithms utilized in various applications and their implications for the future of humanity (Floridi, 2023).

It is pertinent to say that algorithmic discrimination can be mitigated by ethical frameworks and coding algorithms that prioritize fairness and decrease bias. Data anonymization, diversity and inclusiveness in data collection, and continual monitoring and assessment of AI systems can help prevent prejudice and discrimination. Fighting algorithmic discrimination requires acknowledging the need to eliminate social preconceptions and injustices. This includes tackling social and economic inequalities that cause bias and discrimination and promoting IT industry diversity and representation. Developers, ethicists, and other stakeholders must continue to collaborate to overcome algorithmic prejudice. Since AI can perpetuate or reduce inequalities and biases, everyone must work to establish fair and just AI systems for data protection and individual privacy (Leese, 2014).

The term "grund norm" refers to a fundamental principle or standard upon which all other laws and judicial decisions are based. The fundamental standard in the field of machine learning is a set of rules for developing and implementing algorithms that adhere to the principles of objectivity, clarity, and accountability. As a means of resolving the issue, a number of credible specialists have proposed using a "grund norm" approach during the legislative procedure (Rubel, 2016). The first and foremost important "grund norm" for algorithms is explainability (Gunning, 2019). Machine learning requires explainability. Machine-learning algorithms must be designed to explain their decision-making processes. Thus, people would be able to understand and question the automated decisions (Martin, 2022). The concept of fairness serves as an additional illustration of a fundamental principle for machine learning. The objective of this guiding principle is to guarantee that machine learning algorithms make impartial and fair judgments devoid of any bias. Apprehensions regarding the potential of algorithmic decision-making to amplify pre-existing disparities or engender novel ones would be assuaged (Corbett-Davies, 2017).

Machine learning legislation would ensure that algorithms are conceived and performed ethically and legally. Academic revision: transparency, comprehensibility, and equity mitigation may be required (Koene, 2019). Enacting laws based on machine learning principles ensures that these powerful technologies are used morally, lucidly, and responsibly and do not harm humans or perpetuate social inequalities (Yudkowsky & Bostrom, 2011). Therefore, it is imperative that transparency, accountability, and fairness be regarded as fundamental norms in order to expose the black box effect of algorithms.

Subjectivity (Biasedness)

In philosophical terms, subjectivity refers to the idea that each person's subjective experience determines their own truth. In this contexts, a claim is said to have subjective truth if and only if it can be understood through the eyes of a sentient being. LLMs as compared to PLMs are more divergent in being able to perform tasks with a small prompt, but this quality is mysterious and demands investigation (Wei, 2022) to come across the actual facts of having such credibility (Fu et

al., 2022). Training LLMs is itself a big task in terms of resources and cannot possibly be achievable by small techs. It has been noted that if machine learning is applied without any checks, there is a strong possibility of multiplying the bias that is already present in the data (Bolukbasi et al., 2016). While tech giants do not reveal the computational methods (algorithms), data collection sources, or data cleaning procedures, that itself is posing social values and norms at risk because there is a strong possibility that LLMs are also likely to produce lethal, fabricated, or damaging output that requires a legal framework (Open AI, 2023). The study predicts that the algorithms are discriminating as to race and gender (Buolmwini & Gebru, 2018).

Due to the inherent limitations of computer systems, a certain level of variance will persist among them. The adoption of numerical relativity, also referred to as computer subjectivity, is imperative for the business sector. The process of multiplication inevitably leads to a reduction in the number of significant digits, although the degree of imprecision may differ depending on the specific computational system or software being utilized (Colonna, 1993). Digital theory's flexibility and experience's rigidity are trademarks. Subjective data drives the digital revolution, from qualitative assessments to quasi-quantitative solutions like empirical decision-making for quantitative management. In the digital revolution, the "fuzzy-AI model" digitizes subjective information. AI quantifies decision-making and fuzzy estimation uncertainties in this universal problem solver. "Subjective information" can digitally predict human cognition or complex objective events. This makes "big data" searches very dependent on source information, providing precise and complete prognostication (Lin & Zhu, 2021).

Human bias can lead to computer bias. Now, when we talk about GPT and DALL-E, these AIG (artificial intelligence generation) models have been trained by humans. The architectures have contributed to training these LLMs by using the available data, and no one has complete access to the authenticity of the data. The data on which these models are trained are comprised of a large data set and are biased by humans, so the biased output in the model is inevitable (Gebru et al., 2018). The models utilize the acquired patterns from the training data to generate predictions regarding the subsequent word or sequence of words. However, this approach may result in subjective or biased outcomes (Crawford, 2016). When prompted to produce written content pertaining to particular topics or cohorts, a language model that has been trained on news articles may yield prejudiced or partial results. Similarly, it is important to highlight that an image generator trained on a biased dataset may also create biased images (Buolamwini & Gebru, 2018). Recognizing the inherent subjectivity of artificial intelligence (AI) generative models is crucial, as is thoughtfully considering the reasons and methods for their use. Although researchers and practitioners have made efforts to address these challenges through techniques like dataset curation and bias mitigation strategies, a considerable amount of work remains to be done (Bender & Friedman, 2018).

Generative models, capable of producing text and images similar to human creations, have made significant strides in artificial intelligence (AI). These models, however, can yield results that are subjective and biased. A prime example is the ChatGPT language model, a leading AI generative model known for generating biased or inaccurate political content. OpenAI's text-to-image model is also capable of creating subjective outputs (Brown et al., 2020). Research on ChatGPT indicates that white men are more frequently associated with positive terms, while other groups tend to be linked with negative terms, highlighting the subjectivity inherent in AI generative models (Brown et al., 2020). The researcher examining ChatGPT's model also addressed its potential biases and advocated for collaborative efforts to reduce them (Radford et al., 2015). This bias, along with other ethical concerns, can be identified in the data sheets used to train the model on specific datasets (Boyed, 2021). Another approach to reducing bias in word embeddings, which are often

used in language models like ChatGPT, involves removing lexemes associated with stereotypes from the word embeddings' lexicon. This method can create visual representations from user descriptions after extensive training on a vast array of images and text (Bolukbasi, 2016). The system may produce incomplete and inaccurate images that could perpetuate harmful stereotypes or contain inappropriate content. Subjective AI generative models have the potential to disseminate false information or other damaging content, raising concerns. It is important to consider subjectivity and take measures to ensure authenticity when utilizing such models (Radford, 2015).

Subjectivity (Biasedness): An Evidence from ChatGPT and Bard

This biased behavior is not veiled by the ChatGPT. It is clearly mentioned on its portal as a limitation of the model that this model "may occasionally produce harmful instructions or biased content" and 'may occasionally generate incorrect information'. Further, when we ask ChatGPT to discuss its biases, it discloses that it inherits bias from the data provided for training and the types of biases it has, like gender and cultural bias, political bias, confirmation bias, and representation bias (Bolukbasi et al., 2016).

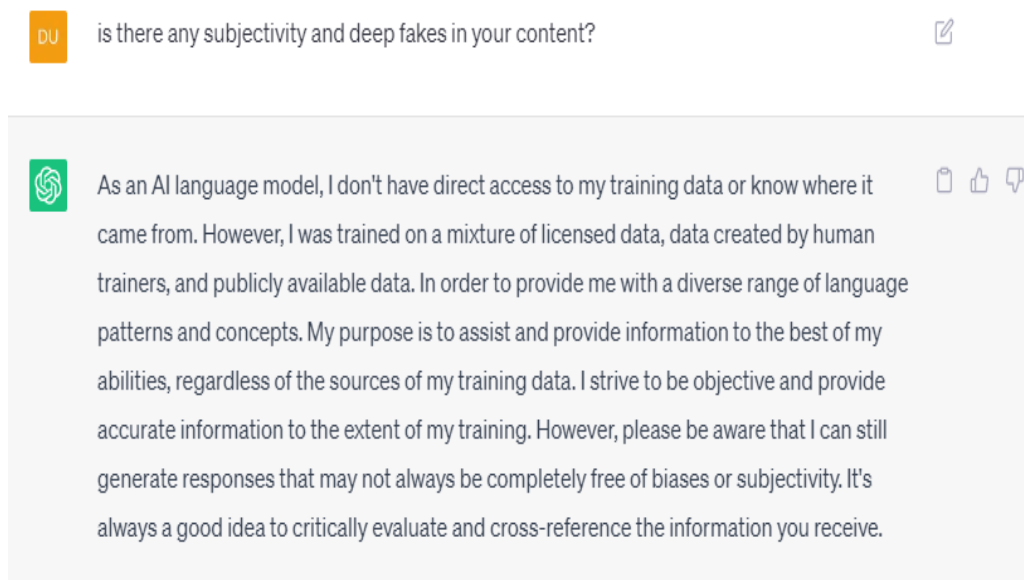


Figure 5: ChatGPT is not free of biasedness (ChatGPT, 2023)

ChatGPT considers bias as an ongoing challenge, as mentioned above in Figure 5. It interprets textual descriptions subjectively, which may differ from human expectations. DALL-E also produces biased images. The issue of bias has emerged as a significant and contemporary concern that warrants careful attention. Given the widespread utilization of models that have the capacity to generate authentic depictions, it is imperative to acknowledge the growing number of global users who are employing these models on a daily basis (Melissa, 2013). When given the prompt "blue horse with wings," you can draw a blue-winged horse rather than a blue horse with wings. Based on training data and biases, model interpretation can vary, resulting in subjective outcomes that may produce different results if textual instructions are unclear or have numerous interpretations. Assuming a fluffy cloud in the shape of a rabbit (OpenAI, 2023).

The model can create a variety of visuals, like a cloud engulfing a rabbit. The model's judgment depends on its language comprehension and training dataset visuals. Training data, which is based on internet text and images, may have cultural and environmental biases. The model's outcomes may reflect biases, reinforcing previous assumptions or social norms. It can draw a male doctor

when given the command "a doctor," suggesting gender bias in the training data. Aesthetic tastes affect production. The model may prefer certain styles, colors, and compositions when producing images from textual descriptions. It can produce an image with bright colors and spectacular illumination when given the stimulus—a picturesque mountain landscape, displaying its aesthetic sensibilities; photos show its subjectivity.

The model can provide stunning visuals, but it may also cause artifacts, distortions, and unnatural components. The intricacy of the prompt, the availability of training examples, and the constraints of model architecture can influence the quality of the model's output. It is crucial to acknowledge its subjectivity and interpret its visuals with care. Human oversight and critical evaluation are vital to identifying and addressing a model's subjective biases and limitations, especially in situations sensitive to fairness and accuracy (OpenAI, 2023). The findings generally endorsed a generative model as a representation of subjects' perceptions of interstimulus relationships (Wiener et al., 1980). The model aids in elucidating how subjectivity is formed and how it evolves (Leech, 2010). This framework illustrates that each neutral generative model is a specific instance that helps uncover a particular set of informational constraints; these constraints define a broader range of non-neutral generative processes that converge on the same neutral pattern (Frank, 2009). Similarly, subjectivity can be found in 'Bard,' the LLM model by Google (Bard, 2023). However, Bard's output may feature varying degrees of subjectivity depending on the nature of the task at hand, as mentioned in Figure 6.

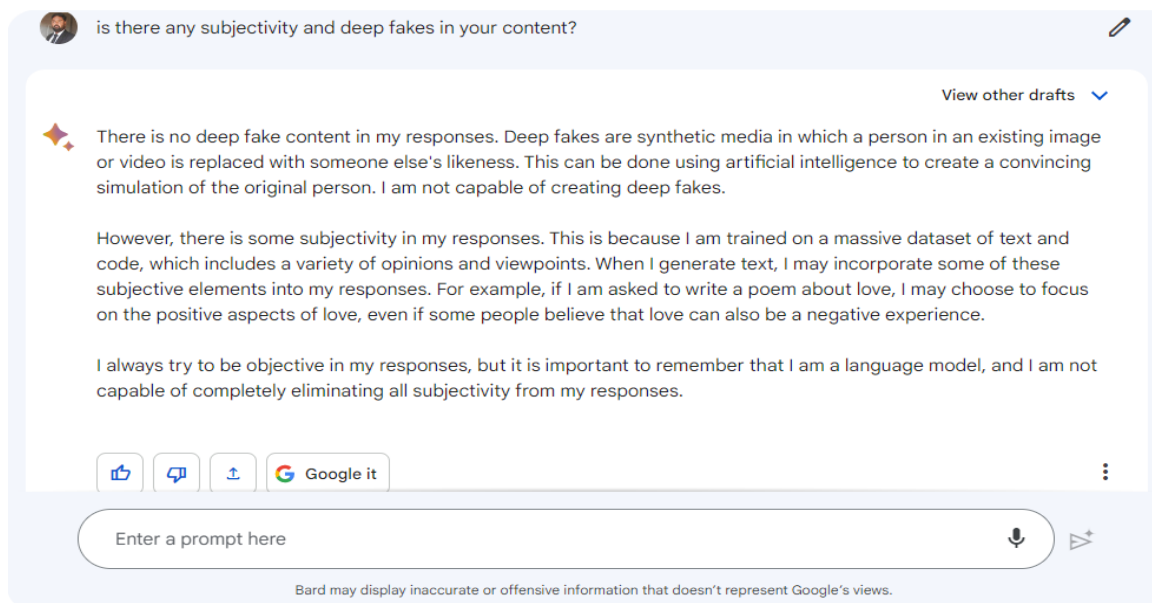


Figure 6: Bard's response as to subjectivity (Bard,2023)

When it comes to jobs that require factual precision, like the condensing of a news piece, Bard's output is often characterized by objectivity. It typically writes subjective poems and stories. Bard can write a human-sounding blog post about a current event with a subjective approach. Bard exhibits racial bigotry. A study requested romantic love poetry from Bard, and the poetry, composed by a white male and female, featured inappropriate racial assumptions (Geoffrey, 2023). Similarly, IMAGEN, another model by Google (Imagen, 2023), Studies show that it favors particular genders, as it created scientific images, and results showed a gender gap, with male characters outnumbering female ones. IMAGEN's race-based discrimination has strong footing due to the discriminating data on which these models have been trained (Joy, 2023). It created

healthcare provider images for a study project in which images overrepresented Caucasian physicians and underrepresented ethnic minority physicians. Research shows that IMAGEN favors certain cultures. It can be a source for misinformation and malicious harassment in the form of weird photos (Casey, 2022). IMAGEN was asked to create family portraits. Its drawings focused on white, middle-class families and lacked multicultural families. It is evident that it favors certain age groups. It was asked to draw successful people. The visual outputs showed a predominance of people in their 30s and 40s and a dearth of older people. It depicts that IMAGEN favors physically capable people. It created images of people doing physical exercises for a study project. IMAGEN's photos mostly depicted athletes and not disabled people. From these cited examples, it can be deduced that IMAGEN's partiality is clearly evident (IMAGEN consortium, 2020).

The primary focus of significance pertains to the attribution of accountability for the subjectivity inherent in these models. The focus of analysis can be either the model itself or its architecture. All of the models under discussion are pre-trained models that have undergone training through the provision of extensive datasets comprising both open and closed source data. The preparation of the data sheets was carried out by the architects, while the development of the commands, coding, and algorithms was also undertaken by the architects. In the event that subjectivity is present in the output of models developed by Microsoft Open AI, Google, or any other technology company, it is the responsibility of the company and the architecture of the models, rather than the models themselves, to be held accountable. The widespread utilization of these models and their associated content creation capabilities has the potential to yield significant benefits for users. However, it is imperative that due diligence be exercised in verifying the ethnic origins of the materials utilized, as failure to do so may result in the proliferation of counterfeit output on a grand scale. The enduring impact of this phenomenon is attributable to the expeditious nature of social media, whereby content is rapidly generated and disseminated globally without adequate oversight or regulation.

Biasedness of AI generative models: An Evidence from Courts

Inaccuracies from AI models can also create many problems. These include unwanted bias and misleading or entirely erroneous outputs such as 'hallucinations' from generative AI. The term "evidence from the courts" refers to cases that have been officially reported to the court system. However, it is important to note that there are likely numerous cases that have been resolved through alternative means and remedies rather than being brought before the courts and undergoing lengthy legal proceedings. Instances have been documented wherein individuals have demonstrated awareness of the partiality inherent in AI generative models utilized in everyday activities. This underscores the necessity for the establishment of robust legislation to address this issue. We can take the example of Volkswagen, which installed emissions-cheating software in diesel vehicles. Volkswagen was fined and recalled millions of vehicles after the scandal. Volkswagen settled the controversy for nearly \$20 billion (New York Times, 2017). In another case, Waymo, an autonomous vehicle subsidiary of Alphabet Inc. (Google's parent company), sued Uber for allegedly misappropriating its self-driving car technology. Uber settled the lawsuit by giving Waymo \$245 million in equity and promising not to use Waymo's confidential information in its autonomous vehicle technology (Anita, 2017).

The utilization of generative models within the realm of artificial intelligence has the potential to generate prejudiced prose. Oracle sued Google for using Java APIs in Android, citing copyright infringement (Google vs. Oracle). It is plausible for AI generative models to exhibit errors. In the year 2020, Microsoft filed a lawsuit against iCrowd, contending that iCrowd had violated Microsoft's patents by utilizing its AI generative model for the purpose of producing medical reports. The legal proceedings pertaining to the case are currently ongoing (Microsoft vs. iCrowd).

It has been observed that artificial intelligence (AI) generative models are susceptible to exhibiting bias. Similarly, a legal action was initiated by CtW Investment Group against Amazon, alleging that the company had engaged in an inequitable labor practice in violation of the 1964 Civil Rights Act by leveraging insights from its AI generative model (Amazon v. CtW Investment Group). Roberto Mata has filed a legal action against Avianca Airlines, alleging harm resulting from the serving cart incident that occurred during a flight to New York. The attorneys representing the plaintiff presented a series of favorable judgments. However, upon cross-examination by the judge, it was determined that all of them were fraudulent. Reportedly, according to Judge Kevin Castel of the Southern District of New York, six cases appear to have bogus citations (Germania, 2023). These examples illustrate some of the ways that bias and fake data can create legal and ethical issues for companies. In each of these cases, the companies involved faced significant financial penalties and reputational damage as a result of their actions.

Contributory Factors in Subjectivity

There are number of factors that contribute in the subjectivity of AI generative models but the most important role is played by the following;

The training of AI generative models on specific data sets

The frequent occurrence of biases in data is a key worry in machine learning models. This is due to the fact that datasets are not objective reflections of reality; rather, they frequently reflect the perspectives, attitudes, and constraints of the individuals who created them. As models are trained, biases can become more evident, potentially resulting in outputs that perpetuate or even intensify these biases (Kristian, 2023). The lack of diversity in training data further complicates this issue. Models trained on limited or uniform datasets may struggle to adapt to a broader range of situations and perspectives. For example, models trained exclusively on positive customer reviews often find it challenging to generate realistic responses to negative feedback. This shortcoming can lead to outputs that are either stereotypical or lacking in depth.

Text interpretations

Textual interpretations are inherently subjective, and there are many components to it. Ambiguous or vague texts allow the people to view in their own way with an open ending. The information, characters, and events are interpreted based on individual differences in prior knowledge, experiences, and values. As readers focus on specific aspects of the text they interpret meaning according to their own peculiar concerns, priorities and interests giving rise to different interpretations. Both language and style as well as those cultural and historical backgrounds use of it are involved in the wide range of ways this consensus may be reached (Grondin, J. 2015). Readers experience texts through the prism of their own biases and gut level feelings, so social and emotional responses play a role in understanding words on the page as well. A third level, intertextuality, guided by which texts con-vesre with other texts also shapes interpretations. This technique is a link in the chain between associations of previously learned information. At its most basic level, textual interpretations necessitate an active engagement whereby readers create meaning. This has the outcome of making for a rich landscape of literary analysis, with multiple perspectives as well as a more nuanced connection to potential meanings.

Ethical considerations

There are several complexities involved in the link between ethical issues and subjectivity. The challenges of ethics are essentially subjective, as they depend on the interpretations of individuals, the norms of cultures, and their priorities. In the process of applying ethical principles, these all

shape the application of ethical principles. Individual moral convictions are a factor that contributes to the variety of subjective methods, and emotional reactions can affect decision-making. Similarly, in civilizations, power dynamics have the potential to distort ethical judgments. Despite the fact that there is subjectivity involved, it is feasible to identify common ground and make appropriate decisions (Kearney, D., 1992). Dialogue, critical reflection, self-awareness, appealing to common values, and the development of ethical frameworks are some of the ways that are recommended. Therefore, participating in a courteous conversation broadens one's understanding and makes it easier to make ethical concessions. By focusing on shared principles like fairness and justice, you can begin to instill the scaffolding for responsible and ethical decision making. The framework must be inclusive, approachable, cognisant of subjective bias and foster responsible decision making in our ever-changing world.

Aesthetic preferences

Aesthetics preference is very subjective and dependent on a lot of things. So, experiences as an unknowable history of the individual and cultural heritage create multiple prisms through which beauty can be perceived and even experienced inside. Such differences in sensory perception — for example, disparities in color vision or changes that occur as we age — might lead to different aesthetic judgments.

It is possible to have different aesthetic judgments due to variances in sensory perception, such as disparities in color vision and changes that occur with aging. A further factor that contributes to the formation of subjective preferences is the active cognitive processing of aesthetic stimuli, which takes into consideration aspects such as balance, symmetry, complexity, and cultural allusions. An additional dynamic layer is added to subjective judgments as a result of the fact that aesthetic preferences develop as a result of learnt and acquired experiences, such as exposure, learning, and social conditioning. Emotional responses, which are interwoven with aesthetics, have the ability to affect preferences through the experiences of enjoying, being amazed, being excited, or being uncomfortable. An individual's aesthetic preferences are shaped by social and cultural settings, as well as fleeting fashions and trends, which can occasionally result in momentary alterations in judgment. It is unfortunate that personal biases and prejudices, which originate from preconceptions and preconceived notions, can have an effect on those who have preferences for aesthetics. In order to appreciate the wide range of human experiences and to encourage open-minded interaction with a variety of art and beauty forms, it is essential to have an understanding of the inherent subjectivity that is present in aesthetic choices. This understanding highlights the depth and complexity of subjective aesthetic appreciation without having to pass judgment (Kuhn, Cheney and Weinstock, 2000).

Challenges of AI generative Models

The application of artificial intelligence (AI) to assist in the process of making legal decisions has become increasingly common (Cohen, et al., 2023). The hallmark of AI generative models is their ability to produce novel content by utilizing prior learning from provided data and interactions with individuals. This methodology facilitates individuals' searching and retrieval of information by engaging with the models. Moreover, these actions present a potential legal liability. While certain risks may be apparent, others are expected to develop over time. These models have achieved extensive proliferation and are being utilized by a significant proportion of the global population. There is an ongoing academic discourse worldwide regarding the advantages and disadvantages of these models. Individuals have the right to utilize technology to their benefit; however, it is incumbent upon the state to safeguard citizens against the potential negative consequences of technology. AI generative models have the potential to generate biased and

fabricated data due to their subjectivity. The aforementioned data presents a potential legal risk with regards to the violation of intellectual property rights, privacy rights, restrictions on freedom of speech, the administration of justice, adherence to the rule of law, and potential shifts in the current legal culture. In Figure 6, below bars show the tendency of impact of generative models on law. Whereas the variation in colours denotes the low, significant and transformative impact of generative AI on law in US, Canada, France and UK accordingly.

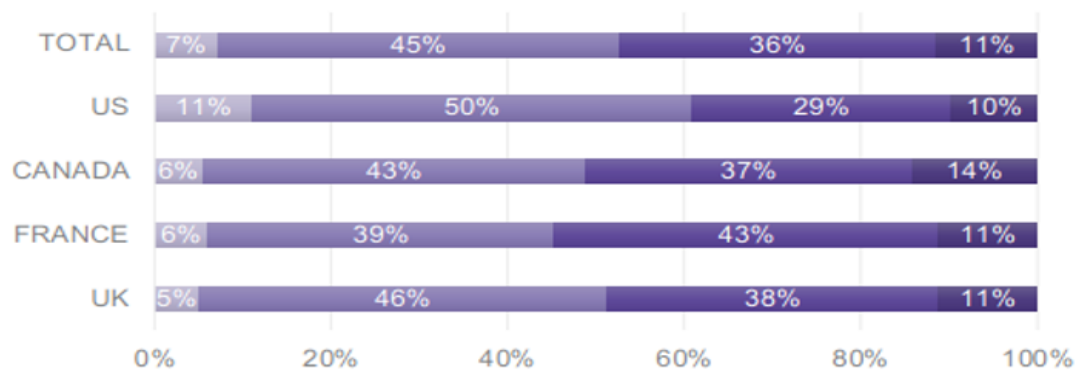


Figure 7: Chart shows the different impacts on Generative AI in US, Canada, France and UK

Right to Intellectual Property

The utilization of AI generative models poses two potential legal risks with regards to content creation. Specifically, these risks involve the infringement of the intellectual property rights of content creators whose work is utilized by AI companies without obtaining prior permission from the rightful owners (Kyle, 2023). The second issue pertains to the assertion of intellectual property rights over user-generated content resulting from the incorporation of unique and innovative ideas into models, thereby creating distinct content that diverges from the norm. There exists a divergence of viewpoints regarding intellectual property rights among the United States, the United Kingdom, the European Union, and Australia. The guidelines released by the US Copyright Office assert that the products generated by generative AI tools are typically not safeguarded. Conversely, UK copyright legislation has the potential to protect computer-generated works in the absence of human involvement, although this domain is presently being scrutinized. Within the European Union, copyrightable work must typically be the result of the intellectual creation of a human author. This requirement is not satisfied by artificial intelligence. Conversely, Australian courts have expressed the view that just as each human being possesses a distinct identity from birth, AI generative technology may also be capable of producing new content with its own unique identity (Rebecca et al., 2021).

Cases relating to copyright infringement are currently being litigated in a number of different courts. Plaintiffs are filing these lawsuits. GitHub Copilot, Sara Silverman, stable dissemination, and the New York Times are just a few examples of cases that illustrate this point. Other examples include the instance of steady dissemination. In the latter instance, the New York Times has filed a lawsuit against open AI and Microsoft, saying that they exploited millions of New York Times articles without licence in order to train their massive language models. In response to the recent allegations made by the New York Times, the action was filed. As a consequence of this, the New York Times is attempting to collect damages in the billions of dollars. An important choice that needs to be made is whether or not the Kelven have clean origins. The United States District Court for the District of Columbia ruled on August 18, 2023, in the case of Thaler v. Perlmutter, that

artwork produced by artificial intelligence is not within the purview of copyright protection. Likewise, Lord Kitchin, who was a member of the Supreme Court of the United Kingdom, has rejected the appeal that there can be no patent. The appeal was dismissed by the Supreme Court in the case of *Thaler v. Comptroller-General of Patents, Designs, and Trademarks*. The Supreme Court concludes that the doctrine of accession does not, as a matter of law, confer on Dr. Thaler the property in or the right to apply for and obtain a patent for any technical development made by DABUS acting autonomously, affirming the Court of Appeal's decision. On the other side, there is a proposition that has been put forth by a number of authors. This proposition is referred to as the "Sony Doctrine," and it has the functionality to incorporate the Large Language Model. Because artificial intelligence (AI) and copy rights are both equally important, it is necessary to safeguard both of them and make use of them in order to reap the benefits of both.

Right to privacy

It is considered proper etiquette to acknowledge the significant importance of data, whether in the form of textual information or visual representations. Individuals may choose to disclose certain data publicly, while other data may be considered highly confidential. The confidentiality of information is a matter of utmost importance for nations, institutions, and individuals alike. When individuals entrust sensitive information to the digital realm for safekeeping or future use, they encounter a significant risk of theft due to the use of cookies and other commonplace applications by websites. Given the ubiquity of the cyber world, this risk is unavoidable. The leaders of the technology industry are currently in pursuit of data, as the absence of data would render their models untrainable. A vast amount of data, amounting to trillions of terabytes, is utilized for model training purposes, and it is evident that the origin of this data is not solely attributed to tech giants. The data being utilized pertains to the masses. The individual in question is cognizant of the fact that their data has been compromised, resulting in a violation of their entitlement to privacy and security, which has occurred for a considerable period of time in the past. The typical user of AI generative models lacks access to the primary content and accompanying data sheets utilized in the training of said models. It cannot be ascertained with certainty whether the General Data Protection Regulation has been adhered to in both its literal and intended sense (Burges, 2020). The contentious subjectivity of the AI generative model pertains to the potential misuse of data and violation of individuals' rights to privacy and security.

Defamation and Hate speech

The exercise of freedom of speech is not absolute and may be restricted in certain circumstances, such as in cases of defamation, hate speech, the dissemination of fake news, and the perpetration of fraud. The preservation of this fundamental right has been a significant challenge for humanity. The potential for AI generative models to produce defamatory material, hate speech, and fake news is a concern due to their inherent subjectivity. Generative models based on artificial intelligence have the capability to generate content and images in accordance with the user's specifications. It has been observed that in developed nations, instances of defamation are typically addressed through legal action against the perpetrator. However, in developing nations, such cases may not be given due attention and may even result in fatal consequences for the individuals involved. Likewise, in cases where defamation pertains to revered manuscripts, individuals, or locations, the resultant responses have been more severe and enduring. The creation of indecent images has become a facile means of defaming individuals, with the potential to significantly impede their livelihood. The ease with which these models facilitate the creation of hate speech and other potentially legally problematic content poses a significant threat to established societal norms. Such models have provided a platform for individuals who engage in disruptive behavior within otherwise healthy communities. In both the physical and digital realms, individuals often

neglect to authenticate news sources, with the latter being particularly susceptible to the widespread dissemination of unverified content. Individuals engaged in the dissemination of fabricated news, such as the false report of an explosion near the Pentagon, which subsequently led to a precipitous decline in the stock market, resulting in the loss of billions of dollars within a brief period (Philip, 2023). The potential for misuse of the right to free speech and the risk of polarization within society through the propagation of defamatory content, hate speech, and false information are significant concerns arising from the subjective nature of AI generative models.

Deep fakes are 'digital media' created using artificial intelligence algorithms and other digital tools to fabricate or manipulate real images, audio, and video in a manner that deviates from reality (Kietzmann et al., 2020). Generative adversarial networks (GANs) can create lifelike images and movies by using the dual-model technique to train a generator and discriminator against each other (Goodfellow et al., 2020). The generator trained with actual images and data creates new photos and movies, while the discriminator trained using real and fake images determines their legitimacy. Adversarial training then trains the generator and discriminator. The generator tries to make images and movies that the discriminator can't tell apart in adversarial training. The discriminator seeks to distinguish real from fake visual media. As generators and discriminators train, their skills improve. The generative model gets better at making authentic images and films, while the discriminative model gets better at distinguishing them. After training, generator and discriminator models can generate deepfakes. The generator can create realistic images and videos. The discriminator may determine if a picture or video is real or fake. So, these biased machine learning models have negative impacts (Bolukbasi et al., 2016). This technology can be used to create fake news, alter or manipulate existing media, or create entirely new content. The implications of these technologies are far-reaching and can have serious consequences as they can cause harm (Amodei et al., 2016). The words representing some data may also be biased. This bias can be avoided by making the architects of the model the responsible persons and compelling them to provide the data sheets on which the model has been trained.

Deepfake, a machine learning-based software application, allows image and video manipulation. Inquiries and court cases often use visual evidence. Deep-fake technology has cast doubt on such evidence. Fake visual media can be manipulated to appear real. Identifying such fake media can be tough. Deepfakes have been used for extortion, terrorism, propaganda, and character assassination. Certain people can agitate society and politics (Albahar & Almalki, 2019). Deep fakes have challenged our ability to recognize the real from the fake (Giansiracusa, 2021). Deep fakes can be used to spread false information, manipulate elections, and create fear and distrust. They can also be used to exploit individuals or groups, create false evidence, fake pornographic videos, and discredit innocent people (Harris, 2018). The emergence of deepfakes in online disinformation can be attributed to the potential involvement of artificial intelligence (AI). A study conducted in the United Kingdom has revealed that individuals are experiencing challenges to online civic culture in democratic societies, including uncertainty, loss of trust in social media, indeterminacy, and skepticism (Vaccari & Chadwick, 2020).

The "deep fake phenomenon is also impacting journalism by producing fake news (Gregory, 2022). An AI-generated image depicting an explosion near the pentagon led to a loss of \$500 billion in the stock market within minutes (NPR, 2023). It also reveals that the utilization of deep fake has the potential to bring about significant alterations in the conception, production, editing, and targeting of advertisements (Campbell et al., 2022). Deepfakes is abrogating the rights in likeness that refer to an individual's legal right to control and profit from the commercial use of their name, image, and likeness and will probably damage the reputation of any individual or group of people. Society is greatly affected. The data show that we are unprepared for deep fakes at all levels. The

early stage of their growth shows shortcomings in problem-solving, explaining the lack of substantial effects. Deep fakes are posing challenges to the laws of the UK and USA (Perot & Mostert, 2020). This challenge requires technology, education, training, and government to play their roles simultaneously to nibble evil in the bud (Karnouskos, 2020).

Legal culture

AI generated models are also a legal risk for the existing legal cultures of the countries. Legal culture can be elaborated as "a set of shared beliefs, values, and attitudes about law and the legal system that exists within a particular society" (Friedman, 1969). It depends on the politics, history, religion, and economy of a country. A strong legal culture paves the way for the rule of law, and a weaker legal culture leads to anarchy. If people are using AI generative models to build their understandings, they are going to depend on tools that are most likely subjective in nature. The common sharing of feelings will also be interpreted as per these models. That is a great risk as to the opinion building and leading to a role in legal culture in the future. Furthermore, if these models get some recognition, it will strengthen the content generated by them.

Administration of Justice

The use of ChatGPT by judges to decide cases of five different origins is also posing serious concerns. ChatGPT clearly mentioned that it may create biased content; judges have used this in deciding the cases. There is also evidence of the usage of the Chat GPT by lawyers in a Manhattan court, whereby there were about six cases that were supposed to be drafted by using AI generative citations (Jesse, 2023). So, if judges don't use their prudent minds and rely on the probable biased information produced by the AI generative models, it will result in biased judgments that will surely lead to injustice in society. The subjectivity of AI generative models as to race, color, and languages has strong probabilities to create biased suggestions, as has been noticed in the cases of recidivism (Lauren, 2016; Van, 2022).

Table 1: Use of generative in legal domain

Use of AI generative	TOTAL	US	CANADA	FRANCE	UK
Researching matters	62%	59%	59%	69%	62%
Drafting documents	46%	45%	39%	54%	50%
Writing email	42%	38%	57%	32%	39%
Understanding new legal concepts	29%	25%	42%	26%	23%
Document analysis	23%	23%	21%	28%	22%
Developing litigation strategies	7%	6%	8%	11%	4%
Conducting due diligence	11%	16%	11%	4%	8%
others	10%	7%	11%	15%	9%

AI generative tools are being used in administration of justice in many ways. Table 1, represents percentage of use of AI generative in different formats of law.

Mitigating the Challenges

The Center for AI and Digital Policy has filed a complaint with the Federal Trade Commission, demanding OpenAI cease the development of novel ChatGPT models until sufficient safety measures have been established (CAIDP, 2023). An investigation has been launched by an Italian regulatory authority called the Garante into OpenAI in light of a recent data breach and the absence of age verification mechanisms during the registration process, both of which are intended to protect children from being exposed to inappropriate AI-generated content (Reuters, 2023). It is

the intention of the Irish Data Protection Commission to work in conjunction with the Italian Garante and the Data Protection Commission of the European Union in order to determine whether or not ChatGPT has violated any privacy legislation (BBC, 2023). According to a report by Reuters, the privacy regulatory organizations in both Sweden and Spain have not yet indicated any desire to launch an inquiry into ChatGPT (Supantha, 2023). This is the case even though both countries have the legal authority to do so. On the other hand, if people have complaints about the artificial intelligence technology in the future, the company might think about doing so in response to such complaints. Recent complaints, investigations, and comments from a variety of countries have focused on the question of whether or not AI businesses are accountable for their actions. The events described above represent only a small portion of the total number of such occurrences. If OpenAI violates the General Data Protection Regulation (GDPR) or any other laws that are analogous to it, according to Ulrich Kelber, a spokesperson for the Federal Commissioner for Data Protection in Germany, the organization might be subject to a ban that is analogous to those imposed by other countries (Berlin, 2023). Artificial intelligence (AI) systems' conception, implementation, and use have ethical and social consequences. AI technology's growing sophistication and prevalence have raised ethical concerns. Ethics can be implemented in machines via ethical frameworks and decision-making algorithms. The frameworks can incorporate ethical values like fairness, justice, and human rights to guide AI system decision-making. Creating AI systems that can learn and adapt to changing ethical norms may be the issue at hand. It could also include methods for human oversight and intervention. It's vital to recognize that ethical norms in AI systems are not a universal solution. Different AI applications use different ethical frameworks and decision-making algorithms. AI's ethical limitations must be acknowledged. Developers, ethicists, and other stakeholders must collaborate to ensure the ethical production and use of AI. AI systems' impartiality depends on their training data. AI systems may be biased due to societal biases and discrimination. AI decision-making must be fair and impartial. The potential of artificial intelligence (AI) systems to capture and analyze large amounts of personal data has raised concerns about data privacy and security. Safeguarding privacy and preventing data exploitation require strong security and ethical standards. Artificial intelligence (AI) systems can make judgments that affect people and society.

However, understanding how such judgments are made can be difficult. In building and implementing artificial intelligence systems, clear accountability and transparency channels are essential. Artificial intelligence (AI) systems can automate many aspects of our daily lives, raising questions about humans' role in decision-making and the effects on our freedom and self-determination. Ethical frameworks and governance mechanisms are needed to guide the development and deployment of AI systems as society integrates them. The above ethical issues relate to artificial intelligence. To ensure ethical design and application of artificial intelligence technology, sustained discussion and thought are needed. The system's capacity to comprehend user input in natural language, precisely decipher the user's intention, sustain context throughout various interactions with the user, individualize responses according to the user's preferences and prior conduct, and facilitate interactions through diverse channels, such as text-based chatbots, voice assistants, and messaging platforms, is imperative. The provision of feedback to users on their interactions with the system, as well as the ability for users to provide feedback on the system's performance, is deemed necessary. Continuous improvement of the system can enhance its performance and optimize user experiences (Flstad & Taylor, 2021).

De lege late de lege ferenda: A round up of the laws for LLM

Countries all across the world, from Japan to the United Kingdom, are interested in regulating artificial intelligence (AI) because of the hazards it represents and also to take advantage of the

economic potential it presents. On a state-by-state basis, the laws that govern artificial intelligence in Canada are now subject to a variety of different types of legislation, including those pertaining to data privacy, human rights, and intellectual property. On the other hand, the Artificial Intelligence and Data Act (AIDA) is scheduled to be enacted no later than 2025, and the process of developing it has already begun under Bill C-27, which is the Digital Charter Implementation Act of 2022. By drawing inspiration from the national AI Bill of Rights framework, the proposed legislation in the state of California, United States of America, aims to prohibit discrimination and harms across a variety of private sectors, including education, utilities, healthcare, and financial services firms. "Self-regulation" is the policy that the United States intends to implement for the artificial intelligence technology business. A board member of the Tech Council of Australia named Victor Dominello advocated for the establishment of a "tech reg vanguard" that would be able to evaluate new advancements in artificial intelligence and provide guidance to government agencies for how to monitor these developments for potential dangers. There are presently 499 members of the European Parliament (MEPs) out of a total of 620 voting in favor of the new law during a plenary vote in June 2023. The European Union (EU) Artificial Intelligence Act is currently being drafted through the European Parliament. Agricultural technology, education, healthcare, smart cities, and smart transportation were the five primary sectors of artificial intelligence research that were taken into consideration by India's National Strategy for Artificial Intelligence. Then, in the year 2020, a draft of the National Artificial Intelligence Strategy was created, which included a section on the ethical use of artificial intelligence. This section called for all systems to be transparent, accountable, and impartial. The Artificial Intelligence Act of South Korea is now in its final stages of writing, and votes are scheduled to take place within the National Assembly. In addition to the General Data Protection Regulation (GDPR), the government of the United Kingdom has also published an Impact Assessment with the purpose of determining how to regulate artificial intelligence developers in a manner that is both appropriate and equitable. It has a policy that is more flexible, "pro-innovation," and seeks to govern artificial intelligence based on the sector rather than the software that underpins it. The Cybersecurity Law (CSL), the Data Security Law (DSL), and the Personal Information Protection Law (PIPL) are the three data-related rules that have been enacted in China. In addition, "Measures for Generative Artificial Intelligence Services" have been established in China. Managing the utilization of artificial intelligence services such as OpenAI's ChatGPT was the purpose of the interim measures. China has implemented specific legislation for a variety of emerging technologies, such as recommendation algorithms and generative artificial intelligence, and is currently in the process of drafting a more comprehensive national AI law that will be implemented in the future years.



Figure 8: Indicates the current regulatory behavior of different countries regarding AI generative

Development of new qualities with old Virtues

Besides legislation, our civilization and technology may require new virtues to guide our behavior and judgment. Philosophers debate the need for new virtues. Due to the pervasiveness of technology in our daily lives, digital literacy and data ethics may be necessary virtues (Toffler, 2022). Alternative views hold that the timeless values of integrity, fortitude, empathy, and sagacity

may guide us in all situations (Ellul, 2021). The theory suggests that new qualities may devalue traditional virtues (Feenberg, 1995). New virtues depend on our view of ethics and social change. If ethical frameworks should adapt to social and technical advances, then new virtues may be necessary. If one believes that essential qualities are timeless and universal, creating new virtues may not be necessary. It's important to talk about morality and how it affects our behavior, regardless of your views. We must promote virtue in ourselves and others and build ethical systems that reflect our highest ideals. ChatGPT analyzes incoming text to find patterns, connect words and sentences, and generate responses that depend on its dataset and algorithms. It also learns from its users and can respond contextually and grammatically, but it lacks consciousness and comprehension. It is required to ensure that the outputs of large language models (LLMs), which are becoming increasingly common in a range of industries, comply with social values and ethical issues as they rise in importance. This is necessary since LLMs are becoming increasingly prevalent in a variety of fields (Tang, 2023).

Basic Elements for Normative Architecture

The potential for AI generative models to produce deleterious or unethical content, such as deep fakes or prejudiced writing, is a matter of concern. The implementation of normative architecture in AI generative models has the potential to effectively tackle the aforementioned challenges. The normative architecture should be grounded in the development and utilization of ethical generative models for AI. The inclusion of privacy, justice, responsibility, and transparency is imperative. The normative architecture should incorporate data quality standards that guarantee the precision, reliability, and inclusiveness of the training data utilized in AI generative model training. This practice guarantees that the models produce unbiased and significant content. In order to enhance the visibility and intelligibility of AI generative models, it is recommended that the normative design incorporate norms pertaining to explainability. Human oversight is essential for identifying and remedying imperfections and partialities. The safety and effectiveness of AI generative models ought to be verified through normative architecture testing and validation criteria. This guarantees that the models are appropriate and do not result in any adverse effects.

The normative architecture should be developed and deployed in a responsible and accountable manner within governance frameworks for AI generative models. The application of industry standards, regulatory control, and self-regulation may be necessary. The development of ethical normative frameworks for AI generative models must prioritize the assurance of accuracy, justice, transparency, and safety. The system ought to facilitate the establishment of AI generative models while also promoting responsible development. It is imperative that artificial intelligence models adhere to regulations pertaining to data privacy, intellectual property, and anti-discrimination measures. It is imperative that artificial intelligence (AI) models adhere to the principles of human rights, transparency, accountability, and non-discrimination. It is imperative that artificial intelligence (AI) models possess characteristics of safety, security, trustworthiness, and ethicality. It is imperative that AI models provide explanations for their judgments and incorporate safeguards to ensure equitable, precise, and impartial determinations.

Conclusion

Article 15(1)(b) of the "International Covenant on Economic, Social, and Cultural Rights" of 1966 (ICESCR, 1966) envisages the right to benefit from scientific progress and its applications. In this era of AI, as every field tries to benefit from it, the legal field cannot be overlooked. Hence, AI applications and their beneficial usage are obvious. It is manifest that the judicial process has always been of great concern and has become more important in the era of artificial intelligence. On the other hand, the UN Convention on the Protection of Humans from the Harms of

Technology also provokes states to save their citizens from the harms of technology and make sure that the citizens do not seize the opportunity to enjoy their fundamental human rights. Now this progress has raised questions about the protection of rights associated with judicial decisions in the age of machines. Research reveals that there is still a major concern regarding the ongoing difficulty of minimizing biases in the language models used by artificial intelligence. It is essential for both users and developers to be aware of the biases that can exist in the material that is generated by AI, to make active attempts to reduce the impact of these biases, and to take a critical stance toward this type of content by comparing information obtained from a variety of sources as and when it is necessary to do so.

In conclusion, the concerns raised by usage of AI generative models presents a strong demand for the development of a robust regulatory framework. These models have been shown to be highly effective in producing realistic and innovative material, but at the same time, they raise a number of significant ethical, legal, and societal concerns. Because of the subjective nature of the models used in AI's generative processes, there are fears that the end results of AI will be biased or harmful. Current study reveals the complex nature of the subjectivity problem that arises in AI generative models. It concludes that the existing regulatory approaches have their limitations and we need a framework that is more comprehensive in order to account for the particulars of AI generative models. However, there has been substantial progress made toward resolving these challenges, but there is still a long way to go before we can consider the problem solved. The development of approaches for quantifying and analyzing subjectivity in AI-generated information should be a main priority of future research, as should the creation of algorithms that eliminate bias and promote transparency.

Additionally, requiring diverse efforts from academics, government officials, and industry leaders, effective regulatory frameworks based on collaborative approach are essential for economic growth. Constant public dialogue and involvement are required in order to guarantee that the ethical standards and points of view of society are taken into consideration in the regulation of AI generative models. The establishment of moral guidelines and guidelines for the appropriate use of these technologies is absolutely necessary in order to promote both innovation and accountability. It is a challenging problem to resolve the use of AI generative models while maximizing their potential benefits and simultaneously minimizing their potential downsides. These models have a strong capacity for mind-building. Now, if the models themselves pose danger to the established norms of society, this may put the public in danger. Like, judges have discretionary power to interpret the law; if the interpretation is based on the ChatGPT suggestions instead of settled rules of interpretation of statutes, this will cause problems in adjudication and will endanger the whole process of adjudication. However, by addressing the issues related to these models through stringent regulatory frameworks, we may establish an environment in which AI can be utilized for the benefit of society while still safeguarding human rights and values. This will allow AI to be employed without compromising the rights and values of individuals. We need to collaborate as a group and frequently assess how well we are doing in order to successfully negotiate this shifting landscape and construct the future of AI in a way that is responsible and will benefit everyone. The formation of a task force dedicated to promoting responsible usage of generative artificial intelligence (AI) in the legal domain and the subsequent development of guidelines for applying established principles of justice administration represent a significant pace forward. However, the matter of responsible generative AI usage must be addressed on a global scale to meet the need of sustainable development goal i.e. Justice.s

References

1. Albahar, M., & Almalki, J. (2019). Deepfakes: Threats and countermeasures systematic review. *Journal of Theoretical and Applied Information Technology*, 97(22), 3242-3250.
2. Amazon v. CtW Investment Group
3. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*.
4. Anita Balakrishnan (2017) Alphabet's self-driving car company just sued Uber — and it all stemmed from an email fail.CNBC<<https://www.cnn.com/2017/02/23/google-just-sued-uber--and-it-all-stemmed-from-an-email-fail.html>> (accessed 13 March 2023).
5. BBC.(2023).ChatGPT banned in Italy over privacy concerns <<https://www.bbc.com/news/technology-65139406>>(accessed 25 May 2023).
6. Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6, 587-604.<https://doi.org/10.1162/tacl_a_00031>(accessed 28 May 2023).
7. Bengio, Y. a.-S. (2003). Adaptive importance sampling to accelerate training of a neural probabilistic language model. *IDIAP*, 3-35.
8. Berlin(april3,2023) "Germany Could Block ChatGPT if Needed, Says Data Protection Chief"< <https://www.voanews.com/a/germany-could-block-chatgpt-if-needed-says-data-protection-chief-/7034099.html>>(accessed 25 May 2023).
9. Bigman, Y. E., Waytz, A., Alterovitz, R., & Gray, K. (2019). Holding robots responsible: The elements of machine morality. *Trends in cognitive sciences*, 23(5), 365-368.
10. Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29.
11. Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29.
12. Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29.
13. Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29.
14. Boyd, K. L. (2021). Datasheets for datasets help ml engineers notice and understand ethical issues in training data. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1-27.
15. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
16. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
17. Brożek, B., Furman, M., Jakubiec, M., & Kucharzyk, B. (2023). The black box problem revisited. Real and imaginary challenges for automated legal decision making. *Artificial Intelligence and Law*, 1-14.
18. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

19. Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77-91). PMLR.
20. Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77-91). PMLR. <<https://doi.org/10.1145/3178876.3186151>> (accessed 5 June 2023)
21. Burges (2020). What is GDPR? The summary guide to GDPR compliance in the UK <<https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018>> (accessed 5 June 2023).
22. CAIDP (March 30, 2023)< <https://www.caidp.org/cases/openai/>> (accessed 5 June 2023).
23. Campbell, C., Plangger, K., Sands, S., Kietzmann, J., & Bates, K. (2022). How deepfakes and artificial intelligence could reshape the advertising industry: The coming reality of AI fakes and their potential impact on consumer behavior. *Journal of Advertising Research*, 62(3), 241-251.
24. Case No. 1:22-cv-01564, (D.D.C. 2022),
25. Casey Tonkin(2022) “Google's new AI can generate weird photos” [<https://ia.acs.org.au/>](https://ia.acs.org.au/) <<https://ia.acs.org.au/article/2022/google-s-new-ai-can-generate-weird-photos.html>> (accessed 5 June 2023).
26. Chen, H. (1995). Machine learning for information retrieval: Neural networks, symbolic learning, and genetic algorithms. *Journal of the American society for Information Science*, 46(3), 194-216.
27. Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. *Fundamentals of artificial intelligence*, 603-649.
28. Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and information technology*, 12, 209-221.
29. Cohen, M. C., Dahan, S., Khern-Am-Nuai, W., Shimao, H., & Touboul, J. (2023). The use of AI in legal systems: determining independent contractor vs. employee status. *Artificial intelligence and law*, 1-30. <<https://doi.org/10.1177/0162243915608944>> (accessed 5 June 2023)
30. Colonna, J. F. (1993). The subjectivity of computers. *Communications of the ACM*, 36(8), 15-19.
31. Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017, August). Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (pp. 797-806).
32. Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017, August). Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (pp. 797-806).
33. Crawford, K. (2016). Can an algorithm be agonistic? Ten scenes from life in calculated publics. *Science, Technology, & Human Values*, 41(1), 77-92.
34. D'Amato, K. (2023). ChatGPT: towards an AI subjectivity. *Phil papers archive*.
35. Delante (March 16, 2023) “GPT-4 – Is It a Revolution for Chat GPT?” <<https://delante.co/gpt-4-is-it-a-revolution-for-chat-gpt/>> (accessed 8 June 2023).
36. Diakopoulos, N. (2014). Algorithmic accountability reporting: On the investigation of black boxes.
37. Durkin, J. (1996). Expert systems: a view of the field. *IEEE Intelligent Systems*, 11(02), 56-63.
38. Ellul, J. (2021). *The technological society*. Vintage.
39. Feenberg, A. (1995). *Alternative modernity: The technical turn in philosophy and social theory*. Univ of California Press.

40. Ferdous, M., Debnath, J., & Chakraborty, N. R. (2020, July). Machine learning algorithms in healthcare: A literature survey. In 2020 11th International conference on computing, communication and networking technologies (ICCCNT) (pp. 1-6). IEEE.
41. Fjeld, J., & Cawsey, A. (2020). A pragmatic approach to developing chatbots for customer service. *Journal of Business Research*, 116, 27-34. <https://doi.org/10.1016/j.jbusres.2020.04.008>
42. Floridi, L. (2023). AI as Agency without Intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology*, 36(1), 15.
43. Følstad, A., & Taylor, C. (2021). Investigating the user experience of customer service chatbot interaction: a framework for qualitative analysis of chatbot dialogues. *Quality and User Experience*, 6(1), 6.
44. Frank, S. A. (2009). The common patterns of nature. *Journal of evolutionary biology*, 22(8), 1563-1585.
45. Friedman, L. M. (1969). Legal culture and social development. *Law and society review*, 29-44.
46. Fu, Y., Peng, H., & Khot, T. (2022). How does gpt obtain its ability? tracing emergent abilities of language models to their sources. Yao Fu's Notion.
47. Ganguli, D., Lovitt, L., Kernion, J., Askell, A., Bai, Y., Kadavath, S., ... & Clark, J. (2022). Red teaming language models to reduce harms: Methods, scaling behaviors, and lessons learned. *arXiv preprint arXiv:2209.07858*.
48. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumeé III, H., & Crawford, K. (2018). Datasheets for datasets. *arXiv preprint arXiv:1803.09010*. <<https://arxiv.org/abs/1803.09010>> (accessed 8 June 2023).
49. Geoffrey A. Fowler (2023) Say what, Bard? What Google's new AI gets right, wrong and weird. *Washington Post* <<https://www.washingtonpost.com/technology/2023/03/21/google-bard/>> (accessed 8 June 2023).
50. Germania Rodriguez Poleo(2023) "NYC lawyer admits using ChatGPT"Mailonline <https://www.dailymail.co.uk/>< <https://www.dailymail.co.uk/news/article-12139497/NYC-lawyer-admits-using-ChatGPT-research-case-client-sued-Avianca-airlines.html>>(accessed 8 June 2023).
51. Giansiracusa, N. (2021). How Algorithms Create and Prevent Fake News: Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More. New York: Apress.
52. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144.
53. Google BARD(2023), <<https://Bard.google.com/>>(accessed 8 June 2023).
54. Google LLC v. Oracle America, Inc. <https://www.supremecourt.gov/opinions/20pdf/18-956_d18f.pdf>(accessed 8 June 2023).
55. Gregory, S. (2022). Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. *Journalism*, 23(3), 708-729.
56. Grondin, J. (2015). The hermeneutical circle. *A Companion to Hermeneutics*, 299-305.
57. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science robotics*, 4(37), eaay7120.
58. Harris, D. (2018). Deepfakes: False pornography is here and the law cannot protect you. *Duke L. & Tech. Rev.*, 17, 99.
59. IMAGEN consortium. (2020). The IMAGEN study: a decade of imaging genetics in adolescents. *Molecular Psychiatry*, 25(11), 2648-2671.
60. Imagen. (2023). <<https://imagen.research.google/>>(accessed 8 June 2023).

61. International Covenant on Economic, Social and Cultural Rights. (1966). General Assembly resolution 2200A (XXI) (accessed 8 June 2023).
62. Iphofen, R., & Kritikos, M. (2021). Regulating artificial intelligence and robotics: ethics by design in a digital society. *Contemporary Social Science*, 16(2), 170-184.
63. Jesse O'Neill. (2023). NYC lawyer admits he used ChatGPT to file 'bogus' court documents < <https://nypost.com/2023/05/30/steven-schwartz-admits-he-used-chatgpt-to-file-bogus-court-doc/>>(accessed 8 June 2023).
64. Johnson, A. M., & Axinn, S. (2013). The morality of autonomous robots. *Journal of Military Ethics*, 12(2), 129-141.
65. Joy Buolamwini (2023) Artificial Intelligence Has a Problem with Gender and Racial Bias. *Times* < <https://time.com/5520558/artificial-intelligence-racial-gender-bias/>>(accessed 8 June 2023).
66. Julio Musto. (May 12,2023). "Artificial intelligence could replace up to 80% of human jobs" *FoxBusiness*< <https://www.foxbusiness.com/technology/artificial-intelligence-could-replace-80-percent-human-jobs>>(accessed 8 June 2023).
67. Julio Musto. (May 12,2023). Artificial intelligence could replace up to 80% of human jobs. *FoxBusiness*< <https://www.foxbusiness.com/technology/artificial-intelligence-could-replace-80-percent-human-jobs>>(accessed 9 June 2023).
68. Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., ... & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*. Kaplan, J. M. (2020). Scaling laws for neural language models. *CoRR*, arXiv:2001.08361v1 [cs.LG], vol.abs/2001.08361.
69. Karnouskos, S. (2020). Artificial intelligence in digital media: The era of deepfakes. *IEEE Transactions on Technology and Society*, 1(3), 138-147.
70. Kearney, D. (1992). *An Ethical Geography: Mapping the Route Between Subjectivity and Moral Realism* (Doctoral dissertation, Reed College).
71. Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat?. *Business Horizons*, 63(2), 135-146.
72. Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). A governance framework for algorithmic accountability and transparency.
73. Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive development*, 15(3), 309-328.
74. Kyle Wiggers. (2023). The current legal cases against generative AI are just the beginning< <https://techcrunch.com/2023/01/27/the-current-legal-cases-against-generative-ai-are-just-the-beginning/>>(accessed 9 June 2023).
75. Lee, J. G., & Lee, K. M. (2022). Polite speech strategies and their impact on drivers' trust in autonomous vehicles. *Computers in Human Behavior*, 127, 107015.
76. Leech-Wilkinson, D. (2010). Listening and responding to the evidence of early twentieth-century performance. *Journal of the Royal Musical Association*, 135(S1), 45-62.
77. Leese, M. (2014). The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. *Security Dialogue*, 45(5), 494-511.
78. Li, J., Chen, J., Ren, R., Cheng, X., Zhao, W. X., Nie, J. Y., & Wen, J. R. (2024). The Dawn After the Dark: An Empirical Study on Factuality Hallucination in Large Language Models. *arXiv preprint arXiv:2401.03205*.
79. Lin, S., & Zhu, W. (2021). How subjective information with AI for digital revolution. *Journal of Intelligent & Fuzzy Systems*, 41(6), 7577-7587.
80. Martin, A. (2022). *Ontology to Represent Norms: Ethics, Law and Informatics. on AI Compliance Mechanism (WAICOM 2022)*, 101.
81. Matthew E. Peters, M. N. (2018). Deep Contextualized Word Representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for*

- Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers) (pp. 2227-2237). New Orleans, Louisiana: Association for Computational Linguistics. doi:<https://doi.org/10.18653/v1/N18-1202>
82. McCarthy, J. (1956). The Birthplace of AI. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Hanover, New Hampshire. <<https://www.cantorsparadise.com/the-birthplace-of-ai-9ab7d4e5fb00>>(accessed 8 June 2023).
 83. Melissa Heikkilä. (March 2,2013). These new tools let you see for yourself how biased AI image models are. MIT technology review
 84. Microsoft v. iCrowd
 85. Mittal, A., Murthy, R., Kumar, V., & Bhat, R. (2024, January). Towards understanding and mitigating the hallucinations in NLP and Speech. In Proceedings of the 7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD) (pp. 489-492).
 86. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.
 87. Nohr, R. F. (2019). The Development of Decision Support Systems in the 1960s as Antecedent of “AI-Rationality”. *Eludamos: Journal for Computer Game Culture*, 10(1), 67-90.
 88. NPR. (2023). Fake news about blast near pentagon <<https://www.npr.org/2023/05/22/1177590231/fake-viral-images-of-an-explosion-at-the-pentagon-were-probably-created-by-ai>>(accessed 8 June 2023).
 89. OpenAI. (2023). Gpt-4 technical report. OpenAI, <<https://arxiv.org/abs/2303.08774>>
 90. Perot, E., & Mostert, F. (2020). Fake it till you make it: an examination of the US and English approaches to persona protection as applied to deepfakes on social media. *Journal of Intellectual Property Law & Practice*, 15(1), 32-39.
 91. Philip Marcelo. (2023). Fake image of Pentagon explosion briefly sends jitters through stock market”<<https://apnews.com/article/pentagon-explosion-misinformation-stock-market-ai-96f534c790872fde67012ee81b5ed6a4>>(accessed 8 June 2023).
 92. Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
 93. Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
 94. Rosenfeld, R. (2000). Two decades of statistical language modeling: Where do we go from here? *Proceedings of the IEEE* 88, 8, 1270-1278.
 95. Rubel, A. (2016). *The Black Box Society: The Secret Algorithms that Control Money and Information*, by Frank Pasquale. Cambridge: Harvard University Press, 2015. 320 pp. ISBN 978-0674368279. *Business Ethics Quarterly*, 26(4), 568-571.
 96. Supantha Mukherjee, Reuters. (April 4, 2023). Italy's ChatGPT ban attracts EU privacy regulators <<https://www.reuters.com/technology/germany-principle-could-block-chat-gpt-if-needed-data-protection-chief-2023-04-03/>>(accessed 8 June 2023).
 97. Tang, Y. (2023). Understanding Societal Values of ChatGPT.
 98. Taylor, R. M. (2022). Galactica: A large language model for science. arXiv preprint arXiv:2211.09085, 282-308.
 99. Thede, S. M. (1999). In Proceedings of the 37th annual meeting of the Association for Computational Linguistics. ACL, College Park, Maryland, USA: Association for Computational Linguistics.<<https://aclanthology.org/P99-1>>(accessed 9 June 2023).
 100. Toffler, A. (2022). *The third wave: The classic study of tomorrow*. Bantam.

101. Tonmoy, S. M., Zaman, S. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models. arXiv preprint arXiv:2401.01313.
102. Tschider, C. A. (2020). Beyond the "Black Box". *Denv. L. Rev.*, 98, 683.
103. Turner, J., & Turner, J. (2019). Legal personality for AI. *Robot Rules: Regulating Artificial Intelligence*, 173-205.
104. University of Arizona library. (2023). Judges Using ChatGPT <<https://law-arizona.libguides.com/c.php?g=1301273&p=9671724>>
105. Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media+ Society*, 6(1), 2056305120903408.
106. Van Dijck, G. (2022). Predicting recidivism risk meets AI Act. *European Journal on Criminal Policy and Research*, 28(3), 407-423. <https://doi.org/10.1007/s10610-022-09516-8>;
107. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
108. Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... & Fedus, W. (2022). Emergent abilities of large language models. arXiv preprint arXiv:2206.07682.
109. Wiener-Ehrlich, W. K., Bart, W. M., & Millward, R. (1980). An analysis of generative representation systems. *Journal of Mathematical Psychology*, 21(3), 219-246.
110. Woo, B., Huynh, T., Tang, A., Bui, N., Nguyen, G., & Tam, W. (2024). Transforming nursing with large language models: from concept to practice. *European Journal of Cardiovascular Nursing*, zvad120.
111. Y. Lu. (2019). Artificial intelligence: a survey on evolution, models, applications and future trends. *Journal of Management Analytics*, 6, 1-29.
112. Yudkowsky, E., & Bostrom, N. (2011). The ethics of artificial intelligence. *Cambridge handbook of artificial intelligence*. Cambridge University Press, Cambridge. [2023] UKSC 49