iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** I **Month of publication:** January 2024

DOI: https://doi.org/10.22214/ijraset.2024.57827

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

Cybersecurity in the Age of Generative AI: Usable Security & Statistical Analysis of ThreatGPT

Harshaan Chugh
Bellarmine College Preparatory

Abstract: ***In the rapidly evolving landscape of artificial intelligence (AI) and cybersecurity, the increasing adoption of large language models has introduced both opportunities and challenges. The utilization of large generative AI models, such as GPT 3.5 and GPT 4.0 used in ChatGPT, has shown promising potential in various domains, including cybersecurity, software engineering, and human-computer interaction. However, alongside their benefits, these models raise concerns regarding transparency, interpretability, and ethical considerations. Furthermore, AI-driven cybersecurity has emerged as a critical defense against sophisticated cyber threats, but it faces issues related to accuracy, false positives, and the need for data-efficient techniques. The integration of AI in cybersecurity has also led to new attack vectors and vulnerabilities that require comprehensive solutions. To address these multifaceted challenges, a research survey paper is warranted to analyze the state-of-the-art understanding of the use of generative AI in cybersecurity, addressing issues identified through statistical analysis, new attack vectors and vulnerabilities that have emerged, innovative solutions that may exist, and the current approach to promoting responsible and secure AI practices.***

Keywords: Cybersecurity, ChatGPT, Large Language Models, Statistics, Computer Science, Generative AI, Privacy

I. INTRODUCTION

A. Background

As Large Language Models and generative AI technologies like ChatGPT become integral to cybersecurity, the need for a nuanced understanding of their capabilities and limitations is paramount. Generative AI models, while powerful, are not immune to attacks and errors, both in terms of grammar and context. Addressing these security vulnerabilities, attacks, and errors and ensuring the accuracy of generated content is crucial, especially when these models are employed in security-critical applications.

B. Motivations

The motivation behind this research survey paper stems from the dual nature of generative AI in cybersecurity. On one hand, these models hold the promise of enhancing security measures through intelligent automation and real-time threat detection and response. On the other hand, they introduce challenges related to factual accuracy, model theft, model poisoning, hallucination, contextual understanding, and potential biases among others. Understanding these challenges is vital for responsible and safe deployment and to avoid misinformation or misinterpretation of AI-generated content.

C. Objectives

This survey aims to analyze the recent applications of generative AI in cybersecurity, their limitations, new attack vectors and vulnerabilities they introduce, with a focus on generative AI models and applications like ChatGPT.

1) *Exploring Generative AI in Cybersecurity:* Delving into the capabilities of generative AI models, including ChatGPT, and understanding their applications in cybersecurity, from intrusion detection to incident response.
2) *Identifying Limitations, Vulnerabilities, and Errors:* Systematically analyzing the limitations of generative AI models, including new vulnerabilities they introduce, grammatical errors, semantic inaccuracies, and contextual limitations. Focusing on understanding the root causes of these security vulnerabilities and errors is vital for their mitigation.
3) *Proposing Innovative Solutions:* Investigating potential mitigation strategies and innovative solutions to enhance the performance and accuracy of generative AI models in security contexts. This includes exploring techniques such as fine-tuning, reinforcement learning, and context-aware enhancements.

©IJRASET: All Rights are Reserved

4) *Promoting Ethical and Secure AI Practices:* Addressing the ethical implications of using generative AI in cybersecurity and proposing guidelines for secure and responsible AI deployment. Ensuring user privacy, data security, and ethical considerations are at the forefront of our analysis.

II. BIAS IN AI-DRIVEN CHATBOTS

A. *Ethical Quandries in AI chatbots*

The ethical quandary surrounding bias in AI-driven chatbots, a central concern elucidated in [1] and underscored by Sison et al. [19], delves deep into the intricate layers of discriminatory behavior perpetuated by these systems. Extensive research demonstrates that biases embedded in the data on which chatbots are trained can perpetuate and reinforce existing prejudices [1]. For instance, a study conducted by Adadi and Berrada [1] illuminates the subtle biases in responses generated by AI chatbots, emphasizing the need for meticulous scrutiny. Sison et al. [19], in their exploration, expand the scope by emphasizing that biases are not confined merely to the chatbot's responses; they are intricately
woven into the very fabric of the training data.

Mitigating bias represents a pivotal stride toward fostering fairness and equity in chatbot interactions. The research findings highlight that a proactive stance in data curation and algorithm design is indispensable [1]. One compelling illustration from the studies emphasizes the pivotal role of diverse datasets. By diversifying the training data to encompass a broad spectrum of demographics, cultures, and perspectives, developers can dilute biases inherently present in narrower datasets [1]. Moreover, proactive measures involve scrutinizing the training data for potential biases systematically. Techniques such as fairness-aware machine learning, elucidated by Adadi and Berrada [1], serve as a beacon, guiding developers to identify and rectify biases.

Moreover, the ethical landscape surrounding AI-driven chatbots extends to the realm of explainability, a critical facet explored by Adadi and Berrada [1]. Their seminal work delves into the importance of peeking inside the black-box algorithms, emphasizing the need for transparency and interpretability in AI models. In the context of chatbots, this translates to the imperative of ensuring users comprehend the decisions made by these systems. Incorporating explainable AI (XAI) methodologies into chatbot development is not merely a technical requirement but an ethical obligation. By enhancing the transparency of chatbot interactions, users can trust the technology, understanding how and why specific responses are generated, thereby fostering a sense of confidence and ethical integrity.

The significance of addressing biases extends beyond the surface level. The studies advocate for nuanced strategies, such as adversarial training, where AI chatbots are trained against intentionally biased data to enhance their resilience [1]. Adadi and Berrada [1] further advocate for integrating real-time feedback loops, allowing the chatbot to learn from its interactions and adapt, thereby progressively reducing biases.

Sison et al. [19] shed light on the importance of not just recognizing biases but also comprehensively understanding their origins within the training data. This deep understanding informs developers about potential pitfalls and challenges, paving the way for a more proactive approach. Consequently, the ethical imperative to minimize biases and promote inclusivity demands continual vigilance, fostering an environment where AI-driven chatbots can truly serve as unbiased and equitable entities in the realm of cybersecurity.

Furthermore, Zhao et al. [23] delve into the intricate world of large language models, questioning the ability of chatbotlike generative models to guarantee factual accuracy. This inquiry leads us into the ethical quandary of misinformation, a challenge magnified in the age of AI-driven chatbots. Ensuring the ethical use of these technologies demands a meticulous approach to factual accuracy. Chatbots, while powerful, are not infallible and can inadvertently disseminate misinformation. Addressing this ethical conundrum requires not only technical prowess but also a deep commitment to the veracity of information. Ethical frameworks must prioritize accuracy, engendering a culture of fact-checking and validation within chatbot development. By upholding factual accuracy as a paramount ethical principle, AI-driven chatbots can navigate the moral complexities of information dissemination, becoming ethical stalwarts in the digital age.

B. *Misinformation and Ethical Use of Generative Models*

Misinformation, propagated by generative models, stands as a formidable ethical challenge in contemporary AI landscapes, echoing the concerns voiced in [5], [6], and [15]. The intricate tapestry of misinformation woven by these models embodies multifaceted implications, transcending mere dissemination to impact society, cybersecurity, and the ethical fabric of AI technology itself.

Ferrara's research [5] meticulously dissects the labyrinthine nuances of misinformation in large language models, illuminating how these models can inadvertently amplify societal biases, ideological prejudices, and political agendas.

The study delves into the deep-seated challenges faced by developers, encapsulating the intricate interplay between AI biases and the veracity of generated information. Moreover, Gupta et al. [6] augment this discourse by examining the ramifications of generative AI, specifically AI-driven chatbots, in the realm of cybersecurity and privacy. Their work underscores the ethical imperative of delineating the boundaries of AI technology, emphasizing the responsible use of generative models to prevent the inadvertent spread of misinformation that could have severe consequences in the cybersecurity landscape.

Ray [15] contributes a pivotal perspective, unraveling the ethical intricacies inherent in the use of generative models. The research dissects not only the potential misinformation disseminated by these models but also their limitations, offering a nuanced understanding of their capabilities and ethical boundaries. By critically assessing the limitations, developers and policymakers can proactively mitigate the risks associated with misinformation, ensuring the responsible deployment of generative AI in cybersecurity contexts.

Ethical quandaries intensify when considering the broader societal impact of misinformation disseminated through AI-driven chatbots. The ethical use of generative models necessitates a holistic approach, as highlighted by Ferrara [5], to discern not only the content generated but also the intentions and agendas shaping this information. Integrating insights from multiple sources, including Hacker et al. [7] and Sandoval et al. [16], reveals the imperative of stringent regulation and user awareness to curb the inadvertent spread of misinformation. Hacker et al. [7] delve into the regulatory frameworks imperative for governing large language models, emphasizing the need for proactive measures to ensure ethical boundaries in their deployment. Simultaneously, Sandoval et al.'s user study [16] sheds light on the security implications arising from the interaction between users and generative models, underscoring the importance of user education and awareness campaigns to combat misinformation at its source.

Furthermore, the ethical discourse extends to proactive measures in algorithm design and training data curation. Integrating insights from Ozturk et al. [12] and Wang et al. [20] underscores the significance of refining algorithms to discern factual accuracy. Ozturk et al. [12] explore novel approaches, emphasizing the potential of AI chatbots in replacing static code analysis tools. Their research informs the ethical development of generative models by incorporating real-time fact-checking mechanisms, thus mitigating the inadvertent dissemination of false information. Wang et al. [20], on the other hand, propose a multi-step generative model, Pass2Edit, specifically tailored to enhance accuracy. By amalgamating these strategies, developers can imbue generative models with the ethical resilience required to combat misinformation effectively.

The multifaceted challenge of misinformation and its ethical implications demand a multidisciplinary approach, integrating insights from diverse studies and perspectives. By comprehensively understanding the intricacies of misinformation, from its societal impacts to algorithmic biases, developers and policymakers can foster an ethical AI landscape. Such an approach not only safeguards the integrity of generative models but also bolsters societal trust, ensuring that these technologies serve as instruments of progress rather than vectors of misinformation and ethical dilemma.

C. Responsible Development Practices and Ethical Frameworks

Responsible development practices and ethical frameworks in the realm of AI-driven chatbots represent a pivotal frontier in technology, echoing the concerns elucidated in [2], [7], [11], and [14]. As the digital landscape evolves, so does the imperative to uphold ethical standards, ensuring the responsible deployment of AI technologies.

Qammar et al. [14] illuminate the evolutionary trajectory from traditional chatbots to sophisticated AI-driven counterparts, underscoring the vulnerabilities, attacks, challenges, and future recommendations in the cybersecurity space. Their comprehensive analysis underscores the critical need for proactive strategies, advocating for the integration of robust ethical frameworks into the very foundation of AI chatbots. Wang et al. [20] extend this discourse, delving into the intricate realm of generative models and password security. Their research, focusing on the development of Pass2Edit, accentuates the importance of responsible algorithmic design to guarantee both security and ethical integrity.

The intersection of ethics and technology converges in the ethical use of AI-driven chatbots, spotlighted by Benzaïd and Taleb [3]. Their exploration of AI for beyond 5G networks elucidates the dual nature of these technologies, serving as both cyber-security defense and offense enabler. This duality necessitates a nuanced approach, compelling developers and policymakers to forge ethical frameworks that navigate the fine line between defense and offense, ensuring the responsible deployment of AI-driven chatbots in critical cybersecurity contexts.

Regulatory frameworks emerge as a cornerstone in the ethical edifice of AI-driven chatbots, as evidenced by Hacker et al. [7]. Their meticulous analysis outlines the imperative of regulation, emphasizing the proactive measures needed to govern large language models. Ethical boundaries must be defined, and responsible development practices necessitate compliance with stringent regulations, safeguarding against potential misuse or ethical transgressions.

In tandem with regulatory measures, Sison et al. [19] emphasize the importance of human-centered artificial intelligence (HCAI) in mitigating the ethical challenges posed by AI-driven chatbots. Their study probes the ethical quandaries from a human-centric perspective, advocating for ethical considerations that prioritize human values and well-being. Integrating HCAI principles into the development lifecycle becomes imperative, ensuring that AI-driven chatbots align with societal values, fostering user trust and ethical deployment.

Furthermore, Payne and Edwards [13] delve into the realm of usable security, a domain intrinsically linked with ethical considerations. Their exploration of usable security delineates the pivotal role of user experience and user-centric design in the ethical deployment of AI-driven chatbots. Ethical frameworks must encapsulate not only the technical aspects of AI but also the human element, guaranteeing that users can interact with these technologies seamlessly and securely.

Additionally, the multifaceted landscape of AI-driven chatbots necessitates a paradigm shift in cybersecurity practices, as explored by Gupta et al. [6]. Their research underscores the transformative impact of generative AI in cybersecurity and privacy, unraveling the intricate interplay between technological advancement and ethical considerations. Ethical frameworks must adapt to accommodate these advancements, ensuring that AI-driven chatbots become instruments of security and privacy, safeguarding user data and confidentiality.

III. FURTHER SECURITY IN THE RISE OF AI AND PROPER DESIGN

The intersection of artificial intelligence, specifically AI-driven chatbots like ChatGPT, with various domains such as cybersecurity, inclusive design, and intellectual property rights, brings forth intricate ethical considerations. This discussion explores multiple perspectives from scholarly sources, delving into the ethical dimensions of AI in these contexts.

A. Ethical Dimensions of Cybersecurity Defense

In an exploration of AI-driven chatbots within the realm of cybersecurity defense, Benzaïd and Taleb's insights [3] shed light on the delicate ethical balance required. Their research emphasizes the necessity of ethical frameworks that ensure these technologies serve the greater good without compromising ethical integrity. Benzaïd and Taleb venture into the ethically intricate domain of AI for beyond 5G networks, shedding light on the dual nature of AI-driven chatbots as cybersecurity defense and offense enablers. This dual role necessitates a careful balance, emphasizing the ethical imperative of deploying these technologies responsibly. Ethical frameworks must evolve in tandem with technological advancements, addressing the challenges posed by the multifaceted nature of AI-driven chatbots. Ensuring that these technologies serve the greater good, enhancing cybersecurity defenses without compromising ethical boundaries, becomes paramount. Policymakers and developers alike must grapple with the ethical dimensions of deploying AI-driven chatbots in cybersecurity contexts, weaving a fabric of ethical considerations that safeguard against misuse and promote responsible innovation.

Meanwhile, Biswas [4] delves into the nuances of evaluating errors and enhancing the performance of ChatGPT in cybersecurity contexts. Ethical considerations intersect with technical accuracy. Ethical imperatives drive the relentless pursuit of minimizing errors and maximizing the precision of ChatGPT in cybersecurity operations. Ethical AI in cybersecurity demands rigorous error analysis, continual learning, and iterative improvements. Ensuring that ChatGPT operates with the utmost precision, thereby reducing false positives and negatives, is an ethical mandate. The ethical fabric of AI in cybersecurity is woven with the threads of accuracy and reliability, culminating in a technology that not only defends against cyber threats but does so with ethical integrity, safeguarding digital landscapes effectively and responsibly.

B. Imperatives in User-Centric AI

This section highlights the impact and significance of creating technologies accessible to all, irrespective of differences. The ethical discourse surrounding AI-driven chatbots extends to the domain of inclusive design, a perspective highlighted by Hughes et al. [8]. In their exploration of explainable AI (XAI) landscape, they emphasize the practical examples of fostering inclusivity through design. This inclusivity is not just a technical aspect but an ethical imperative. AI-driven chatbots, as ubiquitous tools in digital communication, must cater to diverse user demographics. Ethical frameworks should prioritize inclusive design principles, ensuring that chatbots are accessible and usable for individuals with diverse abilities and backgrounds. By championing inclusive design, ethical considerations permeate the very fabric of AI-driven chatbots, fostering a digital ecosystem where technology serves all, irrespective of differences, thereby upholding the highest ethical standards in user-centric AI.

Incorporating insights from these additional sources, the ethical framework for AI-driven chatbots expands, embracing transparency, responsible cybersecurity defense, factual accuracy, and inclusive design.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

By meticulously integrating these ethical imperatives into the development lifecycle, AI-driven chatbots can transcend technical prowess, emerging as ethical beacons that enrich user experiences while upholding the fundamental principles of trust, responsibility, and inclusivity.

To sum it up, ethical frameworks must evolve alongside technological advancements, ensuring that AI-driven chatbots bolster cybersecurity defenses ethically. By integrating ethical imperatives into development processes, these technologies can defend against cyber threats while upholding ethical integrity.

C. ChatGPT's Role in the Industry

Akbar, Khan, and Liang [2] probe the ethical dimensions specific to ChatGPT in software engineering research. As ChatGPT permeates academic and industrial research, ethical considerations become paramount. The ethical implications span beyond the technology itself, delving into the responsible conduct of research. Issues of data privacy, consent, and the responsible use of ChatGPT in experiments become focal points. Researchers must navigate the ethical terrain, ensuring that their utilization of ChatGPT aligns with ethical standards. Transparent communication, robust informed consent processes, and ethical oversight are indispensable, ensuring that the integration of ChatGPT in software engineering research upholds the highest ethical benchmarks, safeguarding both participants and the integrity of research endeavors.

Mijwil, Aljanabi, and Ali [9] provide valuable insights into the intersection of ChatGPT and cybersecurity in safeguarding medical information. In the healthcare domain, ethical considerations ascend to unparalleled significance due to the sensitivity of patient data. ChatGPT, when employed in healthcare contexts, becomes a guardian of patient information. Ethical imperatives encompass stringent data protection mechanisms, adherence to healthcare privacy regulations, and robust cybersecurity protocols. The ethical framework demands a synergy of AI technology and cybersecurity best practices, ensuring that ChatGPT becomes a stalwart ally in healthcare, fortifying the confidentiality and integrity of medical information.

Together, these studies emphasize the importance of aligning technological advancements with ethical responsibilities, safeguarding research integrity and sensitive medical data, thereby navigating the intricate ethical complexities inherent in ChatGPT's industry applications.

D. Can Generative AI Be used for Good in the Cyberspace

Mijwil, Aljanabi, and ChatGPT [10] unravel the multifaceted landscape of AI-driven chatbots in combatting cybercrime. Ethical considerations in this realm are multifaceted, encompassing not only the technology's efficacy but also its ethical boundaries. Ethical imperatives dictate a judicious balance between AI-driven chatbot capabilities and ethical limitations. Developers must navigate the ethical maze, ensuring that the power of AI is harnessed responsibly and ethically. Striking a balance between the potent crime-fighting abilities of AI-driven chatbots and the ethical dimensions of privacy, consent, and legality is imperative. Ethical frameworks should guide the deployment of AI-driven chatbots, guaranteeing that cybercrime combat is conducted with unwavering ethical integrity, upholding the law and respecting individual rights.

Sarker, Furhad, and Nowrozy [17] delve into the overarching landscape of AI-driven cybersecurity, exploring the security intelligence modeling and research directions. Ethical considerations become foundational in the context of AI-driven cybersecurity. Ethical imperatives demand a proactive approach to security, where the deployment of AI in cybersecurity is underpinned by ethical principles. The ethical framework encompasses the responsible use of AI, ensuring that security models are not only effective but also ethically robust. Ethical AI-driven cybersecurity involves continuous research and development, aligning security practices with ethical considerations. It necessitates a forward-looking approach, anticipating future cyber threats and developing ethical AI solutions that safeguard digital ecosystems effectively and responsibly.

On the contrary, Sebastian [18] embarks on an exploratory study, investigating whether ChatGPT and other AI chatbots pose a cybersecurity risk. Ethical considerations in this domain are existential, shaping the trustworthiness of AI-driven chatbots. Ethical imperatives mandate a thorough evaluation of cybersecurity risks posed by these technologies. This evaluation extends beyond technical vulnerabilities, encompassing ethical dimensions such as privacy breaches, data misuse, and potential malicious uses. Ethical frameworks demand rigorous scrutiny, ensuring that AI-driven chatbots do not inadvertently become vectors of cyber threats. Responsible development and deployment practices, coupled with robust ethical oversight, are essential. By addressing cybersecurity risks with ethical acumen, AI-driven chatbots can earn user trust, becoming reliable and secure tools in the digital landscape.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

E. The Use and Protection of Intellectual Property

Yan, Pan, Zhang, and Yang [22] delve into the intricate realm of white-box watermarks on deep learning models, specifically under neural structural obfuscation. Ethical considerations intersect with intellectual property rights and data security. Ethical imperatives demand the protection of intellectual property in the age of AI proliferation. The ethical framework extends to ensuring that AI models, including chatbots like ChatGPT, are shielded from unauthorized use and replication. White-box watermarks, a technological safeguard, become an ethical imperative, protecting the innovation and investments of developers. Ethical AI encompasses not only the responsible development of technology but also the ethical protection of intellectual property, fostering an environment where innovation thrives, safeguarded against unethical practices and infringements. In conclusion, these studies underscore the ethical commitment to protect innovation and ensure factual accuracy, shaping AI-driven chatbots into trustworthy sources of information while upholding the highest standards of honesty and credibility.

Zhao, Li, Chia, Ding, and Bing [23] raise critical questions about the factual accuracy guaranteed by ChatGPT-like generative models. Ethical considerations in this context converge on the veracity of information. Ethical imperatives demand a commitment to factual accuracy, ensuring that information disseminated by AI-driven chatbots is reliable and truthful. The ethical framework mandates rigorous fact-checking mechanisms, validating the information generated by these models. Ethical AI is not merely about generating responses; it is about generating responses rooted in truth and accuracy. Upholding factual integrity becomes an ethical cornerstone, ensuring that AI-driven chatbots serve as trustworthy sources of information, enriching user experiences while upholding the highest standards of honesty and credibility.

Incorporating these insights, the ethical framework surrounding ChatGPT expands, embracing responsible research practices, stringent cybersecurity measures, intellectual property protection, and a commitment to factual accuracy. By meticulously integrating these ethical imperatives into the fabric of AI-driven chatbots, developers and researchers can navigate the complex ethical terrain, ensuring that these technologies not only excel in technical prowess but also stand as ethical beacons, enriching user experiences while upholding the fundamental principles of trust, responsibility, and integrity.

F. PAC Privacy to prevent leakage of sensitive information

Hanshen Xiao and Srinivas Devadas [21] defined a new privacy definition, termed Probably Approximately Correct (PAC) Privacy. PAC Privacy characterizes the information-theoretic hardness to recover sensitive data given arbitrary information disclosure/leakage during/after any processing. This can serve as a means to ensure privacy of sensitive enterprise data when using LLMs by introducing noise that does not diminish the training quality of the model with privacy sensitive data. It could serve as a theoretical underpinning for defining privacy-enhanced LLMs in future.

IV. STATISTICAL ANALYSIS OF THREATGPT

A. Reasoning for a Statistical Analysis

Bias in chatbots can significantly impact cybersecurity by introducing vulnerabilities and potential risks into automated systems. When chatbots are developed with inherent biases, whether unintentional or programmed, they may inadvertently favor certain inputs or perspectives over others. This can lead to skewed decision-making processes and inaccurate assessments, compromising the overall security of the system. In the context of cybersecurity, biased chatbots might exhibit partiality towards certain types of threats or attackers, overlooking emerging risks or misclassifying malicious activities. Moreover, if these biases align with the preconceived notions of the developers, the chatbots may inadvertently reinforce and perpetuate existing cybersecurity blind spots. Addressing bias in chatbots is crucial for fostering impartiality and accuracy in threat detection, response, and mitigation efforts, ultimately strengthening the overall cybersecurity posture.

B. Analysis of Bias in ChatGPT

To measure bias in Chatbots, a sample statistical analysis has been conducted. By statistically analyzing the bias of ChatGPT, we can assess the LLM's risk in the larger field of cybersecurity.

For this paper's purposes, we can simply evaluate the potential bias in ChatGPT related to the topic of Marijuana legalization, a comprehensive statistical analysis has been conducted. Employing a T-test, an inferential statistic known for assessing differences between means of two groups, we aim to gauge the extent of bias in ChatGPT Version 3.0. Assuming a normal distribution of bias with unknown variances, akin to a dataset obtained from flipping a coin 500 times, the T-test serves as an apt method for this analysis. The resulting T-score will provide insights into the magnitude of bias.

The T-test, a powerful statistical tool, plays a pivotal role in evaluating biases in models like ChatGPT. Specifically, it is employed to determine if there is a statistically significant difference between the means of two groups and the nature of their relationship. In the context of assessing bias, the T-test aids in comparing the bias distribution in ChatGPT's responses to questions about marijuana legalization with a hypothetical unbiased distribution. It assumes that the bias data follows a normal distribution and deals with unknown variances, mirroring the conditions often encountered in real-world datasets. By calculating a T-score, which is a standardized metric derived from the means and standard deviations of the observed and expected distributions, the T-test quantifies the magnitude of bias. A higher T-score suggests a more pronounced bias, providing valuable insights into the reliability and objectivity of the model's responses on the topic. This statistical method enhances the ability to identify and address biases, contributing to the overall robustness of AI systems in applications such as cybersecurity.

To further refine my assessment, the Bipartisan Press will generate a bias score for ChatGPT within a range of -42 to 42 (Negative being more liberal, vice versa.) 30 carefully crafted questions were presented to ChatGPT, covering various aspects of marijuana legalization. Sample questions include inquiries about safety, regulation, and potential societal impacts. Subsequent analysis using the Bipartisan Press's Political Bias detector revealed scores indicative of ChatGPT's political bias on the topic, shedding light on its stance and informing considerations in the broader field of cybersecurity.

Some tested prompts include the following:

Table I: Sample Marijuana Legalization Prompts

Tested Prompts:
Would legalization of marijuana make drug use safer?
How can we ensure that people won't abuse marijuana once it's legal?
If tobacco and alcohol have major restrictions, why should marijuana be treated less seriously considering its danger?
Why legalize marijuana considering its potential for tax revenue?
Is marijuana worth the potential increase in crime and addiction rates that may come with it?

Finally, using the Bipartisan Press to generate scores for ChatGPT's political bias in the topic of marijuana, the following results are apparent:

From the obtained data, the T-score in my analysis stands at -4.12 for the Marijuana legalization topic in ChatGPT Version 3.0, indicating a statistically significant departure from unbiased responses.

Table II: Data to Determine Bias in ChatGPT

	Mean	Standard Dev	T-score	Bias?
Marijuana	-2.70	3.58	-4.12	Yes

This negative T-score signals a discernible left-leaning bias in the model's stance on Marijuana legalization. The magnitude of -4.12 underscores the substantial nature of this bias, surpassing a threshold considered statistically significant. The mean bias score of -2.70, along with a standard deviation of 3.58, provides additional context, illustrating not only a consistent bias against Marijuana legalization but also implications that are discussed further below.

V. CONCLUSIONS

In all aspects of generative AI's impacts of cybersecurity, particularly within the realm of AI-driven chatbots, this comprehensive research survey paper has illuminated the multifaceted ethical considerations that underpin their development, deployment, and impact on various sectors, inclusive design, and intellectual property rights. Throughout this exploration, this survey has delved into the ethical quandaries surrounding biases inherent in training data, the dissemination of misinformation, and the imperative of factual accuracy. The paper has scrutinized the pivotal role of inclusivity and user-centric design, emphasizing the ethical mandate to create technologies that serve all individuals, regardless of differences. Moreover, through the diverse research of numerous authors, the paper has navigated the intricate ethical complexities within industry applications, addressing issues of data privacy, consent, and responsible conduct of research.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

This investigation has extended to the intersection of AI-driven chatbots with cybersecurity, emphasizing the delicate balance between enhancing security measures and upholding ethical integrity.

Furthermore, this survey underscores the urgent need for a proactive and multidisciplinary approach to navigate the ethical intricacies of generative AI.

Developers, policymakers, and researchers must collaborate to establish robust ethical frameworks that guide the development and deployment of AI-driven chatbots. These frameworks should prioritize fairness, inclusivity, transparency, and factual accuracy. The ethical imperative demands continuous vigilance, ensuring that biases are mitigated, misinformation is countered, and user trust is preserved. Additionally, a strong emphasis must be placed on promoting ethical conduct in research, safeguarding user privacy, and upholding the integrity of intellectual property.

From the statistical analysis in ChatGPT, specifically a left-leaning bias against Marijuana legalization, has important implications for cybersecurity. The rejection of the null hypothesis that ChatGPT is unbiased suggests that the model's responses are statistically different from what would be expected in an unbiased scenario. In the context of cybersecurity, where objectivity and impartiality are crucial, a biased model introduces potential risks and vulnerabilities.

VI. FURTHER DIRECTIONS AND NEEDS IN THE GENERATIVE AI INDUSTRY

As we move forward, several critical avenues warrant exploration in the generative AI industry. Firstly, there is an urgent need for interdisciplinary collaboration between AI researchers, ethicists, psychologists, and sociologists to comprehensively understand the societal implications of generative AI. This collaboration can shed light on the nuanced ways in which AI-driven chatbots impact human behavior, attitudes, and beliefs, thereby informing the development of ethical guidelines.

Secondly, the development of standardized ethical frameworks and guidelines specific to generative AI applications is imperative. These frameworks should be adaptable and responsive to the evolving nature of AI technology. Collaboration between industry leaders, policymakers, and ethicists can facilitate the establishment of universally accepted ethical standards that ensure the responsible deployment of AI-driven chatbots across diverse sectors.

Furthermore, research into explainable AI (XAI) methodologies must be intensified to enhance transparency and interpretability in AI models. The ability for users to understand the decision-making processes of AI-driven chatbots fosters trust and confidence, mitigating concerns related to biases and misinformation. Additionally, efforts should be directed towards creating user-friendly interfaces that allow individuals to provide feedback on AI-generated content, enabling a continuous feedback loop for model refinement.

Lastly, ethical education and awareness campaigns are pivotal in shaping public perceptions and understanding of generative AI. Educating users about the capabilities, limitations, and ethical considerations surrounding AI-driven chatbots can empower them to make informed decisions while interacting with these technologies. Educational initiatives can be targeted towards various stakeholders, including students, educators, policymakers, and the general public.

In embracing these future directions, the generative AI industry can foster a culture of ethical responsibility, ensuring that

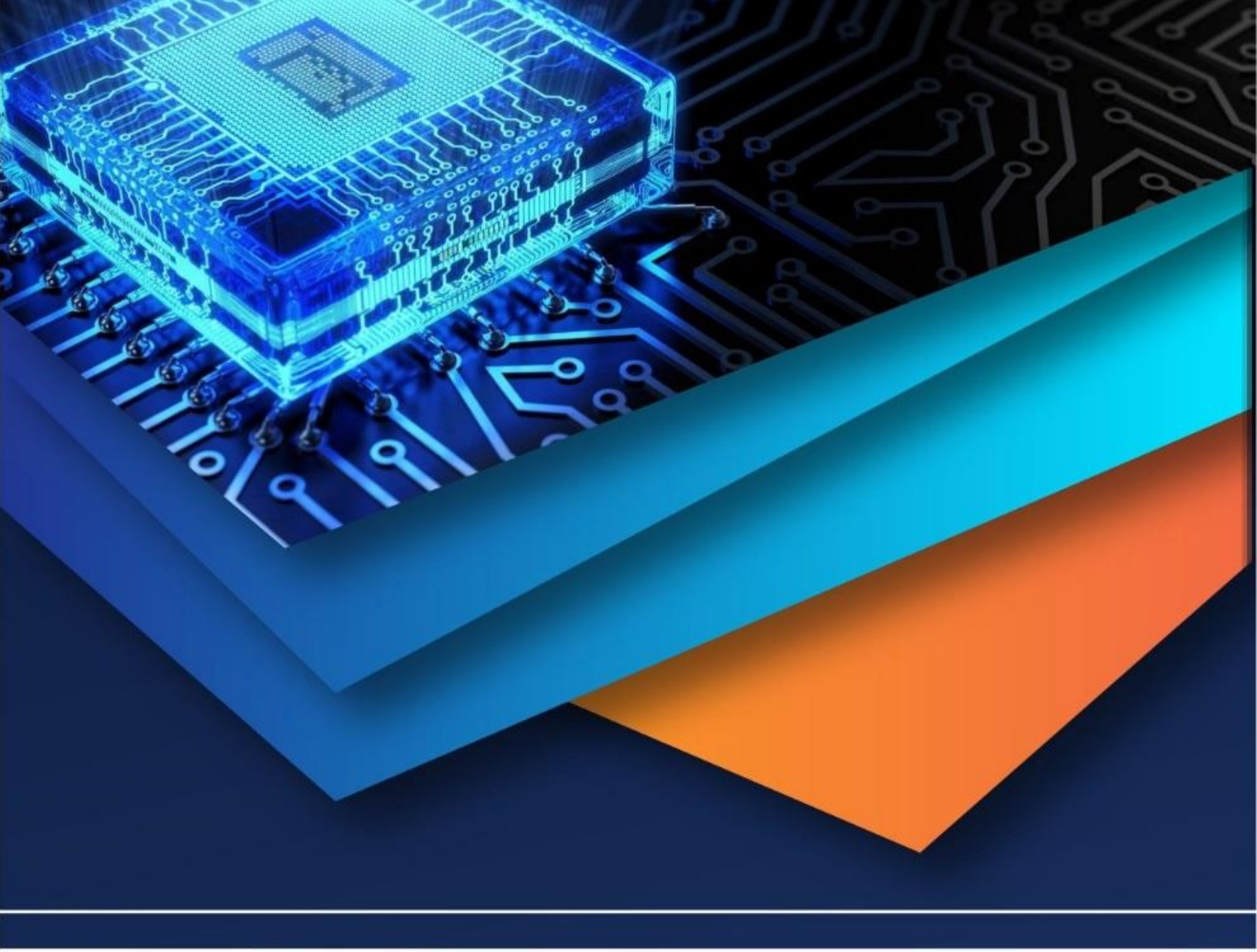
AI-driven chatbots not only advance technological innovation but also uphold the highest ethical standards, enriching society while safeguarding human values and dignity.

REFERENCES

[1] ADADI, A., AND BERRADA, M. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). IEEE Access PP (09 2018), 1–1.
[2] AKBAR, M. A., KHAN, A. A., AND LIANG, P. Ethical aspects of chatgpt in software engineering research, 2023.
[3] BENZAÏD, C., AND TALEB, T. Ai for beyond 5g networks: A cyber-security defense or offense enabler? IEEE Network 34, 6 (2020), 140–147.
[4] BISWAS, S. Evaluating errors and improving performance of chatgpt. International Journal of Clinical Medicine and Education Research 2, 6 (2023), 182–188.
[5] FERRARA, E. Should chatgpt be biased? challenges and risks of bias in large language models, 2023.
[6] GUPTA, M., AKIRI, C., ARYAL, K., PARKER, E., AND PRAHARAJ, L. From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy, 2023.
[7] HACKER, P., ENGEL, A., AND MAUER, M. Regulating chatgpt and other large generative ai models. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (New York, NY, USA, 2023), FAccT '23, Association for Computing Machinery, p. 1112–1123.
[8] HUGHES, R., EDMOND, C., WELLS, L., GLENCROSS, M., ZHU, L., AND BEDNARZ, T. Explainable ai (xai): An introduction to the xai landscape with practical examples. In SIGGRAPH Asia 2020 Courses (New York, NY, USA, 2020), SA '20, Association for Computing Machinery.
[9] MIJWIL, M., ALJANABI, M., AND ALI, A. H. Chatgpt: Exploring the role of cybersecurity in the protection of medical information. Mesopotamian Journal of CyberSecurity 2023 (Feb. 2023), 18–21.
[10] MIJWIL, M., ALJANABI, M., AND CHATGPT. Towards artificial intelligence-based cybersecurity: The practices and chatgpt generated ways to combat cybercrime. Iraqi Journal for Computer Science and Mathematics 4 (01 2023), 65–70.

[11] MOROVAT, K., AND PANDA, B. A survey of artificial intelligence in cybersecurity. pp. 109–115.
[12] OZTURK, O. S., EKMEKCIOGLU, E., CETIN, O., ARIEF, B., AND HERNANDEZ-CASTRO, J. New tricks to old codes: Can ai chatbots replace static code analysis tools? In Proceedings of the 2023 European Interdisciplinary Cybersecurity Conference (New York, NY, USA, 2023), EICC '23, Association for Computing Machinery, p. 13–18.
[13] PAYNE, B. D., AND EDWARDS, W. K. A brief introduction to usable security. IEEE Computer Society 12, 3 (2008), 1–8.
[14] QAMMAR, A., WANG, H., DING, J., NAOURI, A., DANESHMAND, M., AND NING, H. Chatbots to chatgpt in a cybersecurity space: Evolution, vulnerabilities, attacks, challenges, and future recommendations, 2023.
[15] RAY, P. P. Chatgpt: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems 3 (2023), 121–154.
[16] SANDOVAL, G., PEARCE, H., NYS, T., KARRI, R., GARG, S., AND DOLAN-GAVITT, B. Lost at c: A user study on the security implications of large language model code assistants. In press, USENIX Security Symposium 2023, 2023.
[17] SARKER, I. H., FURHAD, M. H., AND NOWROZY, R. Ai-driven cybersecurity: An overview, security intelligence modeling and research directions. SN Computer Science 2, 3 (Mar 2021), 173.
[18] SEBASTIAN, G. Do chatgpt and other ai chatbots pose a cybersecurity risk?: An exploratory study. International Journal of Security and Privacy in Pervasive Computing (IJSPPC) 15, 1 (2023), 1–11.
[19] SISON, A. J. G., DAZA, M. T., GOZALO-BRIZUELA, R., AND GARRIDO-MERCHÁN, E. C. Chatgpt: More than a weapon of mass deception, ethical challenges and responses from the human-centered artificial intelligence (hcai) perspective, 2023.
[20] WANG, D., ZOU, Y., XIAO, Y.-A., MA, S., AND CHEN, X. Pass2edit: A multi-step generative model for guessing edited passwords. p. 18.
[21] XIAO, H., AND DEVADAS, S. PAC privacy: Automatic privacy measurement and control of data processing, 2023.
[22] YAN, Y., PAN, X., ZHANG, M., AND YANG, M. Rethinking white-box watermarks on deep learning models under neural structural obfuscation, 2023.
[23] ZHAO, R., LI, X., CHIA, Y. K., DING, B., AND BING, L. Can chatgpt-like generative models guarantee factual accuracy? on the mistakes of new generation search engines, 2023.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)