BEYOND THE CODE: A REVIEW OF SOCIETAL CHALLENGES AND OPPORTUNITIES IN CONVERSATIONAL AI

CHALKE, A.[1*] – CHENG, B. L.[1] – LEE, T. H.[2] – TENK, M. T. T.[2] – YEONG, H. Y.[1]

[1] *Department of Marketing Strategy & Innovation, Sunway University, Selangor, Malaysia.*

[2] *Faculty of Business, Economics and Accounting, HELP University, Kuala Lumpur, Malaysia.*

**Corresponding author*
e-mail: anujac[at]sunway.edu.my

(Received 23rd January 2025; revised 16th April 2025; accepted 26th April 2025)

Abstract. Conversational AI technologies, such as ChatGPT, are reshaping human-computer interaction by introducing more natural, intuitive, and personalized modes of communication. These systems hold transformative potential across diverse sectors, from education and healthcare to customer service, law and content creation. However, the rapid advancement of such technologies also raises critical ethical and societal concerns such that warrant deeper exploration. This paper delves into the evolving landscape of ChatGPT, discussing its origin and subsequent iterations, including the emergence of multimodal applications, the personalization of AI-human interactions, and the ethical challenges surrounding transparency, bias, and misinformation. While ChatGPT opens new avenues for interaction, it must operate within frameworks that ensure fairness, accountability, and trustworthiness. The responsible advancement of conversational AI requires both internal governance; through ethical design, robust training, and monitoring and external oversight, including regulatory policies and industry-wide standards. As these technologies become increasingly embedded in daily life, collaborative efforts among developers, regulators, and stakeholders are essential to promote responsible innovation. This paper argues for a balanced approach that nurtures innovation while addressing societal risks, advocating for standardized ethical principles and governance structures to guide the sustainable and equitable deployment of conversational AI.
Keywords: *conversational AI, ChatGPT, ethics, societal implications, governance*

Introduction

Conversational artificial intelligence (AI) has revolutionized human-machine interactions across industries. Being at the forefront of technological innovation, a variant as developed at OpenAI known as ChatGPT on the transformer model architecture is unparalleled in its text-generating ability and vast applicability. In particular, the technology's comprehension of context, its ability to generate coherent and compelling text and intuitive creativity have piqued the interest of various industries. This is evident through ChatGPT's contribution to ongoing debates on philosophy, translation of poetry across languages and the formulation of short stories with unexpected twists. After continuous refinement of a series of preceeding models such as GPT-1, GPT-2 and GPT-3; OpenAI launched ChatGPT in the third quarter of 2022. This article serves as a comprehensive guide, aiming to shed light on the evolution of ChatGPT, its functionalities and the ethical dimensions that surround its integration into society. ChatGPT is classified as a virtual social robot that is able to perceive and react to prompts from the environment, has its own sovereignty, can associate with human beings through engaging (virtual) conversations and has a basic understanding of social norms (in most circumstances) owing to its programming (Abadie et al., 2024). Social robots can be defined as independent entities that can

participate in significant social exchanges with human beings. Their association style is influenced by their applications and uses, their defined roles within social contexts and compliance to accepted social norms. ChatGPT has gained tremendous popularity for executing tasks related to natural language processing like translations, generating texts and answers to a range of complex questions given its superior analytical and arithmetical capabilities.

Extant research suggests the immense applicability of ChatGPT as a customer support chatbot to curate emotionally rich experiences (Larivière et al., 2017). Research indicates that ChatGPT is largely received positively by managers as it aids research, ideation and drafting memos and reports, thus enhancing efficiency and internal communication (Cardon et al., 2023). However, it has been met with scepticism by academic as it contributes to issues related to student malpractice and even dilutes the learning process. Yet, it is shown to contribute positively to student engagement and motivation; and also, teachers' productivity by allowing them to focus on high-level tasks (Iqbal et al., 2022).' Overall, a recent study highlights that a consortium academics consider ChatGPT as a disruptive piece of technology, possessing great potential to transform financial systems and humanity (Dwivedi et al., 2023). In addition to easing daily tasks for momagers, the use of ChatGPT also multiplies risks of plagiarism, privacy breaches, misinformation, offensive content and damage to reputation (Abadie et al., 2024). Recent research demonstrates that it is difficult for researchers to distinguish a scientific document drafted by GPT-3 from content written by a peer, and in fact it is AI itself which is able to detect such AI generated texts (Gao et al., 2022). Users of ChatGPT are shown to underestimate how the AI rool impacts their decision making and may even be persuaded to follow amoral recommendations (Krügel et al., 2023). Apart from harbouring risks in terms of random AI failures, ChatGPT also shows signs of potential negative influence on organisational operations and employee welfare (Balagopalan et al., 2023). These AI failures could take form of circulation of hate speech or misinformation (Cao et al., 2021).

Since ChatGPT relies on data from past interactions to improve its performance, there is a risk that it could unintentionally divulge confidential user communication. The need for ethical exploration is driven by an understanding that ChatGPT runs the risk of being used to amplify massive data theft, spread of misinformation, bring damage to reputation, present discriminatory content or otherwise perpetuate harmful conducts (Abadie et al., 2024). Additionally, primary attention has been placed on the importance of investigating existing biasness in the training data and output of ChatGPT (Gonzalez, 2023). Identifying these biases and preventing them from being echoed throughout the technology is critical in enabling a fair and equitable usage of ChatGPT and other conversational AI technologies across all demographics. Thus, this paper aims to undertake a close examination of ChatGPT's architecture, with a sharpened focus on its applications to highlight ethical treatments which can enable comprehensive exploitation of innate potential within conversation AI for an all-inclusive and healthy societal development.

Literature review

The emergence and development of conversational AI like ChatGPT have represented a significant milestone for the future advancement of technologies with capability to process languages. The following review offers an overview of the historical background, transformative journey and subsequent progressive iterations of

ChatGPT in giving rise to exploratory attention concerning relevant aspects and areas on workable potentials of such AI technologies.

ChatGPT development and contributions

History surrounding the development of ChatGPT fundamentally traced the key paradigm as highlighted in the Innovative Transformation Model by Vaswani et al. (2017) through a paper called "Attention is All You Need" which adopted sequence processing mechanism to revolutionize a system's ability in natural language processing (NLP). Such transformation efficiently aligned text sequence processing in a parallel setting which markedly improved its ability to recognize and achieve promising results and outperformed the overly complicated and less efficient recurrent neural network (RNN) (Orrù et al., 2023). Continuous pursuit of similar direction further developed noteworthy success in the form of OpenAI's GPT series. The initial introduction of GPT-1 in the year 2018 highlighted the performance potentials of large-scale generation-based models (Radford et al., 2018). Whereas, the remarkable GPT-2 as introduced in the year 2019 significantly implemented a higher-quality text generation model, whilst concurrently facing a hurdle on public application due to its potential for malicious and criminalized usage (Ray, 2023). Encouraged by the wide acceptance of GPT-3 by the community, OpenAI persisted with its research and development initiatives, to develop ChatGPT, which is grounded in the GPT-4 model architecture.

What sets ChatGPT apart from other conversational technology is its exceptional understanding of context, as shown through the ability to maintain a conversation by generating appropriate and contextually aware responses (Cascella et al., 2023). As popularly documented through the education and content-creating fronts, the technology's capability to provide insights and occasional generation of useful, yet, previously unknown information is stemmed from internal synthesizing of extensive intelligence via the training data (Radford et al., 2018). Moreover, the technology's creative competencies within areas of poetry, skits and more advanced applications of data-based coding and programming have superseded the limited ingenuity of underdeveloped conventional AI technologies (Curtis, 2023).

Evolutionary trajectory of ChatGPT

OpenAI has continuously made strategic enhancements to improve the system's conversational capabilities in mitigating prevailing challenges (Sallam et al., 2023). OpenAI's periodic disclosures offer an insight into the strategic enhancements undertaken over the years to gradually transform said technology into a more sophisticated version of conversational AI. All things considered, the evolutionary milestones of ChatGPT systematically include:

Stage 1-Initial Release: ChatGPT's initial release marked a milestone in the field of conversational AI. This version was critical to demonstrate the maximum extent on effectiveness of Reinforcement Learning from Human Feedback (RLHF) in developing conversational agents that could produce more cohesive dialogues. As such, the initial release formed the basis of all subsequent enhancements, while establishing new standards for conversations with AI that are more naturally fluent (Zheng et al., 2021).

Stage 2-InstructGPT: This model was an attempt to introduce a version of ChatGPT that would minimize the risk of generating harmful or misleading content by artificial intelligence. As some researchers have analyzed, the goal of InstructGPT was to create

the technology to be as close to human instructions as desirable by its users. Therefore, this version of the system has been aimed at developing trust among fellow users on the trustworthiness of created contents (Ouyang et al., 2022).

Stage 3-GPT-3.5/GPT-4 Series: Evolution of ChatGPT has been largely tied to the improvement and advancement of underlying models and processes. GPT-3.5 and GPT-4 are versions of the ChatGPT that are powered by the models of GPT-3.5 and the recently launched GPT-4. GPT-3.5 possesses fewer number of layers as compared to GPT-4 which advanced a simplified version of the current ChatGPT. While GPT-3.5 is freely accessible by users (Ajevski et al., 2023), GPT-4 is considerably ideal for users who undertake more sophisticated tasks and applications.

Future Iterations: Future iterations of ChatGPT are likely to focus on increasing size of the architecture, with incorporating more diversified training data. These improvements will inherently enhance the system's linguistic ability and the generation of more realistic responses. Additionally, focus will also be placed in enhancing the technology's security mechanism in ensuring the safety of AI usage as at-tributed to users' answers and feedbacks. The future version is likely to be more advanced with enhanced capabilities throughout the technical trajectory as per previously mentioned. The evolutionary trajectory of ChatGPT is a shining example of OpenAI's relentless efforts to explore the full potential of conversational AI. By making a series of intelligent modifications, each iteration of ChatGPT incorporates the accomplishments and insights of its predecessor while becoming even more capable of partaking in natural language conversations, all while maintaining the safety of users and the ethicality of corresponding standard of the current society. This evolutionary process is designed to maintain the technology's cohesion to a cutting-edge technological advancement, amidst maintaining its relevance upon facing the constantly changing challenges and opportunities of artificial intelligence.

Critical consideration and ethical implications

While ChatGPT signifies a breakthrough in the landscape of conversational AI through its transformer-based architecture, the model presents several limitations that necessitate critical consideration. Primarily, the model remains vulnerable to generating responses laden with factual fallacies or outdated information (Semmler and Rose, 2017), despite its strong foundation in sophisticated linguistic ability. Such vulnerability stems from the vast, yet, static datasets of text from which the model learns to adapt and responds by inaccurately encoding some irrelevant or incomplete information within its reasoning process. In addition, the model has also encountered challenges in responding to questions with intricate reasoning or domain-specific inquiries, to which further development and innovation is needed to elevate its psychological capabilities. ChatGPT also indicates the possibility of endorsing and transmitting biases through its training data in light of the diverse, yet, imperfect resources from the Internet where the system's training has been based (Bolukbasi et al., 2016). Such reasoning, thus, sheds lights on an existing concern on bias during AI-human interaction and the training process, where the latter must be cautiously structured and guided. Bias minimization through enhanced datasets and model training remains equally critical to realize unbiased and fair interaction of ChatGPT (Dwivedi et al., 2023).

Similarly, ethical considerations emerging from utilization of ChatGPT are significantly widespread and often concern the risk of misuse via crafting disinformation, fostering unreal stereotypes or promoting deceptive conversations (Fui-

Hoon Nah et al., 2023; Zhuo et al., 2023). Such concerns outline the relevance of a balanced, ethical and transparent approach to technological implementation by advocating extensive governance and informational accountability throughout the AI's lifecycle (Shin, 2020). Engaging a vast range of stakeholders in this discussion can potentially furnish invaluable insights regarding the sociocultural impact of conversational AI, before maneuvering greater accountability of such systems and their applications. As such, the rhetoric around ChatGPT should transcends its operational capacities to cover its cultural integration within the current societal settings. A harmonious relationship between both digital advancement and regulatory framework would, therefore, be substantial in steering AI development forward in a responsible manner.

ChatGPT architecture and capabilities

ChatGPT, one of OpenAI's renowned projects, represents a revolutionary development in the field of conversational artificial intelligence. This has been largely attributed to the system's architecture and core capabilities. Several algorithmic elements are known to be critical in ensuring the technology's informed and sound conversational undertakings with a high level of fluidity, contextual sensitivity and interactivity which benchmarked the current standard of modern AI interaction. These architecture and enhanced functionalities include:

Contextual Understanding: ChatGPT's powerful contextual understanding ability is at the core of its foundation as facilitated by the transformer-based architecture model which understands the long-range dependencies that exist within text. Such ability essentially means that the system can practically understand a conversation and its enveloped context to further generate responses with immense accuracy and seamless compatibility to the specific conversation (Du et al., 2023).

Human-Aligned Responses: the system's training architecture through continuous learning from users' feedbacks signifies a radical approach to modern AI development to become increasingly human-centric. RLHF successively "makes the system's outputs more human-like when provided with feedback" in the case of ChatGPT, which essentially makes conversations feel more natural and appealing to the users (Chen et al., 2023).

Adaptability: Incorporation of an adaptable architecture in the development of ChatGPT has allowed the technology to consistently comply with a variety of different contexts. In other words, the system can be adapted to multiple functions including to respond like a gaming companion, a financial advisor, a wedding planner and a storytelling agent, among other applications (Martinez-Arellano et al., 2016).

Information Synthesis: Given that ChatGPT was trained using extensive text datasets, its architecture possesses the ability to generate and convey informative insights, making it a particularly remarkable tool for knowledge discovery and exploration (Ali et al., 2023). Therefore, the technology is also invaluable for the areas of creative research, academic work, content creation and any other domains which involves the need for an in-depth researching skill.

Creative Potential: ChatGPT has integrated a creative potential that is unparalleled in the field of AI technologies (Hussain et al., 2024). This is due to the system's potential to generate insights without prior preparation, such as the generation of scripts, codes and poetic contents.

These key attributes collectively support ChatGPT's advanced functionality against its predecessors in setting a transformative baseline for the development of conversational AI. With further modular development, these features are anticipated to be perfected in advancing its applicability and interactional quality for the evolution of AI-based communication. At ChatGPT's core lies a framework strengthened through the ground-breaking transformer-based architecture, redefining natural language processing (Vaswani et al., 2017). Unlike conventional models which undergo text processing in a linear order, transformer-based architecture processes an entire sequence of text in a single run. With this, ChatGPT can generate text that is more cohesive and relevant in light of enhanced understanding and apprehension. Moreover, ChatGPT has been empowered by OpenAI's implementation of RLHF beyond its base architecture. Such adjustments advance its algorithm, as the system's responses are conditioned based on feedback received from human reviewers. In return, the innovation allows the generation of text with appropriate specific context, as well as alignment to human values and the logic of said conversations. Paramount to multiple applications, some examples of the technology's applicability can be seen in educational tools, customer servicing and creative assistance. ChatGPT has distinguished itself by the ability to work across a variety of domains and undertakings. Ranging from creating creative contents such as poetries and narratives to performing technical supports and answering obscure questions, the different uses within ChatGPT's caliber underscore its complexity and the large scope of its training information. In summary, development undertaken from the primary product of InstructGPT through subsequently revisions of GPT-3.5 and the GPT-4 has delivered a clear message on OpenAI's efforts for tireless improvement of ChatGPT. Each recent update has brought about an increased capability to the technology in the learning of human languages, updating datasets for increased contextuality, reducing biasness and fortifying information security. On the journey of bettering the relationship between artificial intelligence and its users, advancement of ChatGPT as an interactional AI is yet to be completed. Research and design are ongoing to enable the system's continuous acquisitions of additional architectural dimensions and a comprehensive set of data-driven features. The latter will, nonetheless, increase its capability to comprehend certain language compositions for a more profound and distinctive human-AI discussion. ChatGPT will always be a foundation in transpiring upcoming development of similar, if not, more advanced AI-based technologies.

Application horizon of ChatGPT

ChatGPT's far-reaching implications can be attributed to its excellent context awareness, text fluency and context adaptability. The discussed capabilities and competencies could, therefore, assume key roles in enhancing the system's compatibility across numerous general and more industry-specific applications. Among these main applications include ChatGPT's contributions to industries which require flexible, specific and adaptable knowledge-based assistance, as per the case of medicine and healthcare in an-lysing patients' data and providing personalized health recommendations that guide error-free diagnostics and treatments (Tudor Car et al., 2020). The technology's ability to apprehend formal languages further confirm its argumentative capability within the legal field, whilst providing needed assistance on case preparation (Choi et al., 2021). This is besides its ability for data learning through collection and appraisal of extensive scholastic and scientific works towards stipulation of hypotheses or backing of breakthrough discoveries. Combining these abilities as part

of its adaptable architecture fundamentally demonstrate positive outlook in addressing industry- specific issues, while providing relevant solutions. The array of core capabilities that ChatGPT possesses encompass conversational interface, content creation, education, technical assistance and information synthesis (Hussain et al., 2024). Nevertheless, main focus has been allocated on the system's interactive ability, where natural human-AI conversation is posited towards the development of effective chatbots, virtual assistants and customer support in general. Its uncanny prowess for information storage from previous interactions, whilst generating acceptable responses hereby foster both appealing and personalized interactions. More advanced setting was seen on the system's generation of text across various contexts and styles to fit specified factual and creative requirements of writers, marketers and other content generators who demand novel subjects or information in written form (Wahid et al., 2023). This has been apparent via tasks such as summarizing of given text, answering of questions from contradictory viewpoints and the employment of creative writing cues like metaphors.

Not to mention, ChatGPT holds the caliber of customizing content and interactions, and clarification of queries to ensure a more engaging learning experience. System's adaptability has, yet, come into play on adoption of various knowledge areas, employment of real-life exemplifications and the execution of interactive exercises (Annamalai et al., 2023). The technology is also commanded on its virtual assistance in aiding repair, responding to technical questions, with the offering of precise help to professionals across designated sectors through data-driven guidance on working manuals, interpretation of explained errors and delivering of linear instructions. Additionally, ChatGPT's ability on simultaneously collecting and processing of expansive datasets would be utmost helpful within the researching field on reviewing of previous publications, drawing of assumptions and analysis patterns and the discovering of contradicted outlooks (Susarla et al., 2023). Within complex and specialized sectors like healthcare, ChatGPT shines on the revolutionizing of patient interactions by offering streamlined services such as medical responses and mental health screening and support, alongside the personalization of tailored health insights and simplifying complex medical jargon (Cascella et al., 2023). Such aptitude establish a reliable virtual companionship to doctors and medical practitioners in navigating chronic conditions, offering pre-diagnosed symptom checking (Gunawan, 2023). The area of customer servicing then sees the employment of ChatGPT for quick issue resolution and chat-driven troubleshooting, all through its integration to other communication platforms like chatbots to offer a round-the-clock support (Nazir and Wang, 2023).

Besides addressing customer inquiries, the system would excel in clarifying product descriptions, providing quotations, facilitating specific applications and managing returns or claims for warranty (Davenport and Ronanki, 2018). ChatGPT is similarly proven invaluable within the legal context for investigations, drafting of legal documents, conducting basic case analyses, as well as the forecasting of verdicts (Ajevski et al., 2023). Likewise, it helps fellow researchers in the development of research interviews, generating new exploratory themes and analyzing result patterns to the point of drafting a comprehensive scientific paper with detailed materials and applied methodology (Susarla et al., 2023). With its remarkable versatility and adaptability, ChatGPT emerges as a powerful agent of change, while laying the groundwork for a novel approach in multiple spheres. System's application for the modernization of daily practices and transformation of highly specialized processes have verified a borderless potential for enhanced efficiency and performance across

multiple sectors beyond examples given in healthcare, customer servicing, law and research (Davenport and Ronanki, 2018). As research centering around natural language processing and machine learning gain significant momentum, mindful interactivity of AI technologies such as ChatGPT is likely to gain more impressive capacities which enable rapid reshaping in virtue of technology-driven communication and workflow optimization. Upcoming adjustments as undertaken to the technology would likely redefine its functional feasibility and propensity to replicate and surpass limitations of its human users.

Ethical landscape and social dynamics of conversational AI

The application of conversational AI technologies like ChatGPT in almost every aspect of daily life has unveiled a complex tangle of ethical and social considerations. In our fast-changing technological world, such advanced systems open up new opportunities for efficiency, creativity and connections that are unprecedented. Yet, they reveal complex challenges that necessitate close scrutiny and proactive regulations. As we approach a new era in which conversational AI dominates, it is essential to explore the delicate balance between technological breakthrough and moral duty. While integration of such advanced forms of AI into our daily lives creates opportunities for unmatched efficiency, creativity and connectivity, such dramatic shift in technological phenomenon has indisputably shined the light on a broad range of ethical concerns and societal implications. It is in this context that privacy, misinformation, unemployment, social misconduct, bias need to be explored in progressively developing a practical framework which harvests the technology's potential without foregoing current human values and wellbeing. As such, *Table 1* comprehensively summarizes the main landscape and social dynamics of conversational AI.

Table 1. *Main landscape and social dynamics of conversational AI*.

Category	Description
Privacy and data integrity	Given that ChatGPT's operations are primarily grounded in using large and externally sourced datasets with a mix of publicly available and private text data (Siau and Wang, 2020), it is evident that privacy and data security should be a top priority. Specifically, stringent data handling measures must be put in place to avoid the accidental extraction of sensitive information as datasets for AI training (Dwork and Roth, 2014). Attention is strongly placed on the transparency in source and origin of the extracted database to develop a trusted environment for reliable interaction between users and AI. Corporations should prioritise user awareness about ethical concerns, including the risk of trade secret leakage, and provide guidelines on sharing information with generative AI. Additionally, implementing government regulations and policies is crucial to safeguard information privacy and security (Fui-Hoon Nah et al., 2023). Some policies include anonymizing data, implementing strong encryption, and offering users control over their data. With the advent of differential privacy techniques, it is possible to collect valuable data for AI training while preserving individual privacy (Dwork and Roth, 2014).
Countering misinformation	The potential of ChatGPT in developing logical and persuasive text makes it an influential tool for content creation. Nevertheless, its negative potential for criminalized goals such as the dissemination of misinformation and numerous malicious scams compel a stern need for information protection (Bazarkina and Pashentsev, 2020). Understandably, reliability of such technological and AI platforms should be built on the incorporation of advanced security setups such as accurate information validation and the educating of digital users on existing threats within the refractory virtual platforms and cyberspace (Flyverbom, 2016).
Mitigating employment disruptions	Employment of conversational AI has paved a contemporary route on automation which reshaped various functions as previously handled by human capital, such as customer servicing and virtual content creation (Brynjolfsson and McAfee, 2014). An increased demand for personnel within the AI management and technological centric domains, yet, reflected a possible loss of employment opportunities in other technologically dominant sectors (Zarifhonarvar, 2024). To ensure human adaptation to such phenomenon, strategies for specified reskilling among working individuals to work with AI in a collaborative capacity is essential. Industries must meticulously strategize the integration of generative AI, invest in employee training for adaptation to the changing landscape. Broad and equitable access to education and training related to AI will ensure there exists equal opportunity for each employee to up-skill and bridge any knowledge gap (Fui-Hoon Nah et al., 2023).

Balancing technological dependency with human connection	An increased dependence on conversational AI for a wide range of services such as organizational inquiries have given rise to crucial issues pertaining its effect on human interaction (Towers-Clark, 2020). Continuous verbal and written familiarity to machines over people could denote a fall in emotional intelligence towards handling the conventional human communications. As such, a positive digital environment and implementations which promote human communications would be utmost crucial in justifying the interactive nature of AI technologies within the traditional social setup (Fui-Hoon Nah et al., 2023).
Ethical dilemma	Given ChatGPT's global adoption, it is important to consider cultural sensitivities in different regions to prevent biases (Dwivedi et al., 2023). When AI plays a role in decision-making throughout various stages of employment, biases and opacity can be prevalent (Chan, 2024). Stereotypes related to gender, race, sexual orientation, or occupations may be present in recommendations generated by generative (Fui-Hoon Nah et al., 2023). Equality should be promoted with minimal bias in the digitalized environment. With the understanding that conversational AI technologies such as ChatGPT incorporates training data as key source for decisive evaluation, a more diversified and inclusive database should be extracted within the algorithm to reduce unnecessarily erroneous bias (Gonzalez, 2023). A popular practice to efficaciously achieve such results, thus, observed the incorporation of robust bias evaluation techniques within the AI's development life cycle (Chen et al., 2021).
Aligning AI development with societal values	Speedy integration of conversational AI technologies within the current society has directly affected the existing social values and its norms. It then places responsibility on technology developers, lawmakers and related authorities to route the products' righteousness and genuine advantages to the society (Hacker et al., 2023). Likewise, key concern is specifically allocated to the ethical consideration for both protection and actionability of human rights and integrity (Jobin et al., 2019).

Ethical considerations

As the implementation of ChatGPT progresses, the importance of ethical considerations needs to be consistently elevated. As per discussed, transparency and trackability are critical to build trust, while fostering the responsible deployment of AI technologies. Herewith, Explainable AI (XAI) techniques can help demystify the decision-making pattern of ChatGPT, where mechanism in which the AI reaches a specific conclusion can be tracked by fellow users to identify the existence of biases or ethical shortfalls (Chan, 2024). Likewise, one of the major ethical challenges as currently addressed is the need to mitigate risks associated with misinformation and falsified disseminations (O'Brien, 2023). As ChatGPT becomes more sophisticated in generating text and media, the risk of misuse to spread fake news, propaganda or explicit content would also increase (Bazarkina and Pashentsev, 2020). Countering these threats, thus, calls for the mandatory incorporation of content verification tools, digital literacy campaigns and algorithmic auditing to support the data-driven environment and its integrity (Chauhan and Palivela, 2021). Finally, the challenge of addressing biasedness in AI's judgments persists. Despite rapid advancement in algorithmic equality and bias reduction, creating sound and righteously informed outputs and AI-generated contents will demand continuous efforts and oversights (Dwivedi et al., 2023; Ntoutsi et al., 2020). As such, stakeholders including researchers, ethicists, policymakers and industry representatives should actively shape a respectable, ethical dialogue regarding AI development and its far-reaching implementations (Kroll, 2015).

Societal implications

The proliferation of conversational AI technologies such as ChatGPT carries numerous societal implications across multiple sectors, including employment, education, healthcare and interpersonal relationships, among others. To address these implications proactively, however, calls for a comprehensive understanding of the intricate nexus between technology and human behavior. In terms of employment, the expansion of AI-driven automation presents both opportunities and challenges. On one hand, ChatGPT and comparable technologies can help to streamline work processes,

enhance productivity and create new avenues for job opportunities. On the other hand, there is a growing fear of job redundancy and displacement following upcoming adoption of AI approaches within various industries (Pavlik, 2023). Therefore, reskilling, upskilling and workforce responsiveness are required among employees and active jobseekers to secure an equal opportunity within the changing economy. Notably, the influence of conversational AI within education is transformational in which it provides tailored learning capabilities, besides specified tutorage and adaptable educational resources (Firat, 2023). It is necessary to consider pedagogic theories, concerns of data confidentiality and the role of human instructors in enabling significant learning exposure amid the incorporation of AI technologies.

The same can be seen in the healthcare sector where conversational AI offers new ways to transform its fundamental provisions. Besides offering virtual assistance on people-centered activities and tasks, AI-powered solutions enable improved healthcare results with enhanced patient experience (Cascella et al., 2023). Nevertheless, moral reasoning around data protection, algorithmic partiality and regulatory constraints must be settled to assure responsible medical usage of such technology, considering that interactive human elements and relationships would be directly influenced. Although AI relationships and digital assistants provide comfort and companionship, they bring up worries regarding social separation, psychological addiction and criminalized deceits (Gunawan, 2023; Siau and Wang, 2020). To foster healthy digital relationship and users' experience, a balance between innovation and human-centered design is conclusively imperative.

Governance and responsibility

Rising integration and familiarity of conversational AI technologies within the society has inevitably pushed a critical requirement for an effective governance framework and digital-centered regulations to retain ethicality, protect users' integrity and ensure accountability. Notably, cross-border collaborations would be mandatory to establish globalized guidelines on numerous aspects of artificial intelligence (Jobin et al., 2019). The technological sector which would comprise of the initial stakeholders within AI development should contemporaneously develop ethical baselines, data transparency and risk mitigation for their products (Fui-Hoon Nah et al., 2023). Similarly, ethical committees should endorse the algorithmic and regulatory aspects of artificial intelligence to innovatively enhance the ethical commitment among technological employees amid their controlled digital usage (Rader et al., 2018). Guidance would then be important to the public in empowering the righteous manners in which AI technologies should be adopted, alongside the inclusivity and transparency of the technology's database towards reaching of sound decisions that endorse crucial moral and social values (Taeihagh, 2021).

Conclusion

Conventional AI, not limiting to the commonly known ChatGPT, possesses immense potential in the upcoming future on a mutual direction of societal benefits and scientific advancement. The vast extent of benefits as brought forward by such digitalized advancement is highly remarkable and noteworthy. Borderless interactivity as enabled by the technology through both emotional intelligence of mankind and rationality of innovative machines could ease the addressing of international issues, whilst promoting

strengthened global relations through a shared worldview on innovation and sustainability. In conclusion, the emergence of conversational AI platforms such as ChatGPT ushers a brave new world of human-machine interaction which has been laced with potentials and challenges. Our exploration on the platform's technological accomplishments, ethical dilemmas, societal consequences and the call for governance further uncovered a complex tapestry of potentials and caveats that come with its imposition within the contemporary society. ChatGPT's continuous evolving multimodality, improved reasoning capabilities, and interactive procedures highlight potential to enhance numerous fields beyond education and medicine. However, the technological know-how ought to be coupled with stringent strategies to combat ethical apprehensions such as dishonesty, inequality and misinformation. Societal impacts of AI-infused automation, possible shifts in human interaction, creativity and authorial prowess rounded out the proposal for advanced and sustainable plans on workforce re-skilling, renewed regulations, and bias mitigation. Cross-border cooperation and autonomous sector-focused regulation will play a significant influence in establishing ethical rigors to guarantee accountability and morale amid constructing users' confidence. A comprehensive tactic that equalizes technical invention to human values and community welfare rewards will facilitate the platform for a potential future in which ChatGPT and conversational AI promote genuine reforms which benefit individuals' wellbeing. With this in mind, collective dedication towards administering the increased influence of AI would be necessary in shaping a forthcoming, digitalized collaboration that posits life enrichments with principled governance and protected welfare.

Acknowledgement

This research is supported by HELP University Internal Research Grant Scheme (IRGS) 2023. Project number: 23-10-002.

Conflict of interest

The authors confirm that there are no conflicts of interest associated with any parties involved in this research.

REFERENCES

[1] Abadie, A., Chowdhury, S., Mangla, S.K. (2024): A shared journey: Experiential perspective and empirical evidence of virtual social robot ChatGPT's priori acceptance. – Technological Forecasting and Social Change 201: 23p.
[2] Ajevski, M., Barker, K., Gilbert, A., Hardie, L., Ryan, F. (2023): ChatGPT and the future of legal education and practice. – The Law Teacher 57(3): 352-364.
[3] Ali, S.R., Dobbs, T.D., Hutchings, H.A., Whitaker, I.S. (2023): Using ChatGPT to write patient clinic letters. – The Lancet Digital Health 5(4): e179-e181.
[4] Annamalai, N., Ab Rashid, R., Hashmi, U.M., Mohamed, M., Alqaryouti, M.H., Sadeq, A.E. (2023): Using chatbots for English language learning in higher education. – Computers and Education: Artificial Intelligence 5: 9p.
[5] Balagopalan, A., Madras, D., Yang, D.H., Hadfield-Menell, D., Hadfield, G.K., Ghassemi, M. (2023): Judging facts, judging norms: Training machine learning models to

judge humans requires a modified approach to labeling data. – Science Advances 9(19): 14p.

[6] Bazarkina, D., Pashentsev, E. (2020): Malicious use of artificial intelligence. – Russia in Global Affairs 18(4): 154-177.
[7] 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: 25p.
[8] Brynjolfsson, E., McAfee, A. (2014): The second machine age: Work, progress, and prosperity in a time of brilliant technologies. – WW Norton & Company 86p.
[9] Cao, G., Duan, Y., Edwards, J.S., Dwivedi, Y.K. (2021): Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making. – Technovation 106: 15p.
[10] Cardon, P.W., Getchell, K., Carradini, S., Fleischmann, C., Stapp, J. (2023): Generative AI in the
[11] Workplace: Employee Perspectives of ChatGPT Benefits and Organizational Policies. – SocArXiv Papers 17p.
[12] Cascella, M., Montomoli, J., Bellini, V., Bignami, E. (2023): Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. – Journal of Medical Systems 47(1): 5p.
[13] Chan, G.K. (2024): AI employment decision-making: integrating the equal opportunity merit principle and explainable AI. – AI & SOCIETY 39(3): 1027-1038.
[14] Chauhan, T., Palivela, H. (2021): Optimization and improvement of fake news detection using deep learning approaches for societal benefit. – International Journal of Information Management Data Insights 1(2): 11p.
[15] Chen, H., Yuan, K., Huang, Y., Guo, L., Wang, Y., Chen, J. (2023): Feedback is all you need: from ChatGPT to autonomous driving. – Science China Information Sciences 66(6): 1-3.
[16] Chen, R.J., Lu, M.Y., Chen, T.Y., Williamson, D.F., Mahmood, F. (2021): Synthetic data in machine learning for medicine and healthcare. – Nature Biomedical Engineering 5(6): 493-497.
[17] Choi, J.H., Hickman, K.E., Monahan, A.B., Schwarcz, D. (2021): ChatGPT goes to law school. – Journal of Legl Education 71: 14p.
[18] Curtis, N. (2023): To ChatGPT or not to ChatGPT? The impact of artificial intelligence on academic publishing. – The Pediatric Infectious Disease Journal 42(4): 1p.
[19] Davenport, T.H., Ronanki, R. (2018): Artificial intelligence for the real world. – Harvard Business Review 96(1): 108-116.
[20] Du, H., Teng, S., Chen, H., Ma, J., Wang, X., Gou, C., Li, B., Ma, S., Miao, Q., Na, X., Ye, P. (2023): Chat with ChatGPT on intelligent vehicles: An IEEE TIV perspective. – IEEE Transactions on Intelligent Vehicles 8(3): 2020-2026.
[21] Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., Baabdullah, A.M., Koohang, A., Raghavan, V., Ahuja, M. (2023): "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. – International Journal of Information Management 71: 63p.
[22] Dwork, C., Roth, A. (2014): The algorithmic foundations of differential privacy. – Foundations and Trends® in Theoretical Computer Science 9(3-4): 211-407.
[23] Firat, M. (2023): How chat GPT can transform autodidactic experiences and open education? – Anadolu University 5p.
[24] Flyverbom, M. (2016): Transparency: Mediation and the management of visibilities. – International Journal of Communication 10(1): 110-122.

[25] Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., Chen, L. (2023): Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. – Journal of Information Technology Case and Application Research 25(3): 277-304.
[26] Gao, C.A., Howard, F.M., Markov, N.S., Dyer, E.C., Ramesh, S., Luo, Y., Pearson, A.T. (2022):
[27] Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. – BioRxiv 18p.
[28] Gonzalez, W. (2023): How businesses can help reduce bias in AI. – Forbes Web Portal 9p.
[29] Gunawan, J. (2023): Exploring the future of nursing: Insights from the ChatGPT model. – Belitung Nursing Journal 9(1): 1-5.
[30] Hacker, P., Engel, A., Mauer, M. (2023): Regulating ChatGPT and other large generative AI models. – In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency 2p.
[31] Hussain, K., Khan, M.L., Malik, A. (2024): Exploring audience engagement with ChatGPT-related content on YouTube: Implications for content creators and AI tool developers. – Digital Business 4(1): 14p.
[32] Iqbal, N., Ahmed, H., Azhar, K.A. (2022): Exploring teachers' attitudes towards using chatgpt. – Global Journal for Management and Administrative Sciences 3: 97-111.
[33] Jobin, A., Ienca, M., Vayena, E. (2019): The global landscape of AI ethics guidelines. – Nature Machine Intelligence 1(9): 389-399.
[34] Kroll, J. (2015): Accountable algorithms. – Princeton University 248p.
[35] Krügel, S., Ostermaier, A., Uhl, M. (2023): The moral authority of ChatGPT. – ArXiv Preprint 20p.
[36] Larivière, B., Bowen, D., Andreassen, T.W., Kunz, W., Sirianni, N.J., Voss, C., Wünderlich, N.V., De Keyser, A. (2017): "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. – Journal of Business Research 79: 238-246.
[37] Martinez-Arellano, G., Cant, R., Woods, D. (2016): Creating AI characters for fighting games using genetic programming. – IEEE Transactions on Computational Intelligence and Ai in Games 9(4): 423-434.
[38] Nazir, A., Wang, Z. (2023): A comprehensive survey of ChatGPT: advancements, applications, prospects, and challenges. – Meta-Radiology 1(2): 12p.
[39] Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E. (2020): Bias in data-driven artificial intelligence systems-An introductory survey. – Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 10(3): 14p.
[40] O'Brien, M. (2023): Chatbots sometimes make things up. Is AI's hallucination problem fixable. – AP News 7p.
[41] Orrù, G., Piarulli, A., Conversano, C., Gemignani, A. (2023): Human-like problem-solving abilities in large language models using ChatGPT. – Frontiers in Artificial Intelligence 6: 13p.
[42] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. (2022): Training language models to follow instructions with human feedback. – Advances in Neural Information Processing Systems 35: 27730-27744.
[43] Pavlik, J.V. (2023): Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. – Journalism & Mass Communication Educator 78(1): 84-93.
[44] Rader, E., Cotter, K., Cho, J. (2018): Explanations as mechanisms for supporting algorithmic transparency. – In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems 13p.

[45] Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. (2018): Improving language understanding by generative pre-training. – OpenAI Web Portal 12p.
[46] Ray, P.P. (2023): ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. – Internet of Things and Cyber-Physical Systems 3: 121-154.
[47] Sallam, M., Salim, N.A., Ala'a, B., Barakat, M., Fayyad, D., Hallit, S., Harapan, H., Hallit, R., Mahafzah, A., Ala'a, B., Fayyad Jr, D.J. (2023): ChatGPT output regarding compulsory vaccination and COVID-19 vaccine conspiracy: a descriptive study at the outset of a paradigm shift in online search for information. – Cureus 15(2): 16p.
[48] Semmler, S., Rose, Z. (2017): Artificial intelligence: Application today and implications tomorrow. – Duke Law & Technology Review 16: 15p.
[49] Shin, D. (2020): User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. – Journal of Broadcasting & Electronic Media 64(4): 541-565.
[50] Siau, K., Wang, W. (2020): Artificial intelligence (AI) ethics: ethics of AI and ethical AI. – Journal of Database Management (JDM) 31(2): 74-87.
[51] Susarla, A., Gopal, R., Thatcher, J.B., Sarker, S. (2023): The Janus effect of generative AI: Charting the path for responsible conduct of scholarly activities in information systems. – Information Systems Research 34(2): 399-408.
[52] Taeihagh, A. (2021): Governance of artificial intelligence. – Policy and Society 40(2): 137-157.
[53] Towers-Clark, C. (2020): Human skills will be most important in the digital future of work. – Forbes Magazine 5p.
[54] Tudor Car, L., Dhinagaran, D.A., Kyaw, B.M., Kowatsch, T., Joty, S., Theng, Y.L., Atun, R. (2020): Conversational agents in health care: scoping review and conceptual analysis. – Journal of Medical Internet Research 22(8): 21p.
[55] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I. (2017): Attention is all you need. – Advances in Neural Information Processing Systems 30: 15p.
[56] Wahid, R., Mero, J., Ritala, P. (2023): Written by ChatGPT, illustrated by Midjourney: generative AI for content marketing. – Asia Pacific Journal of Marketing and Logistics 35(8): 1813-1822.
[57] Zarifhonarvar, A. (2024): Economics of chatgpt: A labor market view on the occupational impact of artificial intelligence. – Journal of Electronic Business & Digital Economics 3(2): 100-116.
[58] Zheng, X., Zhang, C., Woodland, P.C. (2021): Adapting GPT, GPT-2 and BERT language models for speech recognition. – In 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), IEEE 7p.
[59] Zhuo, T.Y., Huang, Y., Chen, C., Xing, Z. (2023): Exploring ai ethics of chatgpt: A diagnostic analysis. – ArXiv Preprint 10(4): 17p.