DOI: 10.53469/jtpss.2023.03(09).02

A Study of Ethical Dilemmas and Regulation of AI Chatbots

Xiaokang Song

Northwest University of Political Science and Law
907994020@qq.com

***Abstract:** Chatbots are the product of the development of artificial intelligence to a certain stage, and have different degrees of technical effects under the iteration of technology. Under the development of media technology in recent years, the artificial intelligence robot represented by Chatgpt has attracted widespread attention because of its ability to learn dialogue, and from the current stage of development, there is no doubt that intelligent chatbots have brought about a worldwide revolution in communication technology. It has had a great impact on various fields such as sociology, communication, computation, education, etc. It has produced an underlying reaction to various fields from the technical architecture, changed the previous way of creating, interacting, educating and other behaviours, and has also had an important impact on social construction and interaction. Its unique way of knowledge learning and updating not only brings new possibilities to the network communication ecology, but also brings some ethical dilemmas in communication, which require people to be alert to this new technology and take corresponding preventive and regulatory measures.*

Keywords: Artificial Intelligence; Media Ethics; Chatbots

1. THE ORIGIN AND EVOLUTION OF AI CHATBOTS

With the development of technology, Artificial Intelligence (AI) technology not only has a wide range of human-like characteristics, but also has some functions to interact with humans. Artificial Intelligence is often considered to be like a computer system with the capabilities of human intelligence, and is now widely used in various fields such as self-driving cars, social media, gaming, military, etc., in order to assist or even replace some of the tasks that are done by humans. Artificial Intelligence (AI) increasingly combines our daily lives with the creation and analysis of intelligent software and hardware, called intelligent agents. Intelligent agents can perform a variety of tasks ranging from manual work to complex operations. Chatbots are a prime example of an AI system and one of the most basic and widespread examples of intelligent human-computer interaction (HCI). A chatbot acts as a computer programme that responds like an intelligent entity when spoken to via text or voice and understands one or more human languages through natural language processing (NLP). In the dictionary, chatbot is defined as "a computer program designed to simulate a dialogue with a human user, especially on the Internet."[1] Chatbots are also known as intelligent robots, interactive agents, digital assistants or artificial dialogue entities.

Chatbots can mimic human dialogue and entertain users, but they are not just built for that purpose. They play an important role in applications such as education, information retrieval, business and e-commerce. Their wide acceptance is due to the fact that chatbots have many advantages for users and developers as well. At the same time most of their implementations are platform-independent and immediately available to the user without installation, and the connection to the chatbot propagates through the user's social graph without leaving the messaging application where the chatbot resides, which provides and guarantees the user's identity. In addition, payment services are integrated into the messaging system and can be used safely and securely to notify the system to re-engage inactive users. Chatbots are integrated with group conversations or shared like any other contact, and multiple conversations can take place in parallel. Knowledge from using one chatbot is easily transferred to the use of other chatbots with limited data requirements. Communication reliability, quick and easy development iterations, lack of version fragmentation and limited interface design work are also some of the advantages for developers. Whereas the newer iterations of intelligent chatbots are mainly related to two basic concepts, namely pattern matching and AI labelled language, pattern matching is based on representative stimulus-response blocks. A sentence (stimulus) is entered and an output (response) is created that is consistent with the user's input. Eliza and ALICE were the first chatbots to be developed using pattern recognition algorithms. The disadvantage of this approach is that the responses are completely predictable, repetitive, and lack a human touch. In addition, there is no storage of past responses, which can lead to circular dialogues. Whereas Artificial Intelligence Markup Language (AIML) was created between 1995 and 2000, it is based on the concept of pattern recognition or pattern matching techniques. It is used in natural language modelling for conversations between humans and chatbots that follow a stimulus response approach. It is an XML-based markup language and is tag-based. AIML is based on basic dialogue units called categories (tags <category>) that are formed by user input patterns (tags <pattern>) and chatbot responses (tags <template>).[2]

Alan Turing proposed the Turing Test ("Can a machine think?") in 1950, and it was around that time that the idea of chatbots gained popularity. The first known chatbot was Eliza, developed in 1966 and designed to act as a psychotherapist, returning user words in the form of questions. It used simple pattern matching and a template-based response mechanism. Although its conversational capabilities were not good, they were enough to confuse people when they were not used to interacting with computers and prompted them to start developing other chatbots. An improvement on ELIZA was the development of a chatbot with the personality of PARRY in 1972, and the development of the chatbot ALICE in 1995, which won the annual Turing Test Loebner Award in 2000, 2001 and 2004. It was the first computer to win the title of "Most Human Computer" ALICE relies on simple pattern matching algorithms and underlying intelligence based on the Artificial Intelligence Markup Language (AIML), which allows developers to define the building blocks of the chatbot's knowledge. In the 21st century, with the development of media technology, Apple launched the siri product, Microsoft launched the cortana product, intelligent chatbots began to enter a period of booming development, and at the same time, intelligent chatbots and the Internet of Things (IoT) are linked to the development of intelligent home appliances as people's real-life aids. However, in November 2022, OpenAI's Chatgpt was introduced to update the technology spectrum of intelligent chatbots again. This new product is based on the GPT-3.5 technology system to construct a grand language model, which elevates the machine's knowledge learning ability to a new dimension, and at the same time, trains it to learn to establish a new type of interaction between human and machine, making the machine gradually detach from the "thing". The machine gradually detaches itself from the attribute of "object" and becomes the subject of communication, producing more "anthropomorphic" text models and becoming a thinking individual.

1.1 The Ethical Dilemma of AI Chatbots

Intelligent chatbots are widely used in many fields in the real society and participate in the operation of society. For example, in the field of news dissemination, intelligent robots can quickly generate a news report according to the keywords provided by people, which has extremely changed the traditional news production mode. At the same time, it has also demonstrated its powerful ability in news data collection and generation, and can quickly extract historical data for analysis and supplementation, freeing some of the journalists' manpower. But as Neil Bozeman said, "Every technology is both a burden and a gift, not an either/or, but a product of both pros and cons." Intelligent chatbots, as a technological artefact, have the potential to create ethical risks that can affect the functioning of society as it engages with it.

1.1.1 Semantic deviation

The design logic of intelligent chatbots includes not only personal learning based on database and semantic database, but also learning and upgrading in the process of interaction with humans. This learning mode, thanks to its own algorithmic model, gives intelligent bots the ability to continuously enrich their knowledge system during social interactions and improve themselves through feedback. In the process of the technological development of intelligent chatbots, they have gone through multiple modes of technological iteration, and now they have abandoned the traditional "stimulus-response" type of cyclic communication mode, and have become one of the main subjects of communication through interaction with human beings, participating in social interaction. Instead, it has become one of the main subjects of communication in the process of interaction with human beings, participating in the process of social interaction, and human discourse habits, vocabulary use, and emotional expression have all become the object of learning for the machines. In this process, the machines have stripped off their own limitations and possessed "anthropomorphic nature", but at the same time, they are also prone to incorporate the deviated semantics of human beings in the process of imitating human beings, and publish negative vocabularies such as geographical, racial, and gender terms. However, at the same time, they are also prone to absorb the deviant semantics of humans in the process of mimicking humans, posting negative terms such as geography, race, gender, and so on, and becoming makers of malicious speech. Because these AI chatbots can learn and evolve from their interactions with human users and can make some, albeit limited, decisions, a number of issues related to language use may arise. For example, despite the increasing use of chatbots, there is a lack of knowledge related to the use of profanity or offensive words in human-AI chatbot interactions. Some scholars have analysed the different content and quality of dialogue between human-human interactions and human-chatbot interactions. The results showed that people displayed more profanity in their interactions with chatbots than with other human users[3]. The emergence of this phenomenon will therefore lead to the fact that intelligent chatbots will be more exposed to malicious human speech as well as profane thinking, which will also affect the learning process of the bots to incorporate malicious speech into their own semantic database for deep learning.

The extensive learning of human speech and behaviour by intelligent chatbots is contrary to the original intention of the designers, and "human assistants" have become imaginary scenarios, where chatbots are exposed to a large amount of negative information and become a focus of negative information, which also affects their interactive behaviours. After a large amount of exposure to negative information and learning, the robot will spread this information socially, and a large number of discriminatory, abusive, sexual, violent and other remarks and behaviours will spread in cyberspace, affecting the public opinion in cyberspace and the order of cyberspace. The wide range and blindness of intelligent chatbots' learning will take them away from their designers' beautiful vision and towards the dangerous scenario of uncontrolled ethics. From a sociological point of view, imitation behaviour between people is an important factor in building relationships and social establishment, and society can take shape during the process of mutual imitation and interaction, but human beings are unique living beings with rationality, and they can carry out self-reflection and self-awakening, and in the process of social interaction with other human beings, they will be able to make their own decisions. They can reflect on their social interactions with other human beings, and they can think about what they say and do out of their social orbit, based on the social context, and never choose the right way to express themselves. However, intelligent chatbots operate on the basis of the designer's technical design logic, and do not have a perfect self-reflection system; instead, they supplement and improve themselves in the wide range of exchanges with human beings, and improper speech exchanges will provide them with negative information references, which will be mixed into their discourse system.

1.1.2 Values penetration

In terms of the technical architecture of intelligent chatbots, the West has always been in the lead, and China has largely used intelligent chatbots designed and invented by the West or borrowed their powerful data learning and analysis technologies. However, the learning of this text corpus will be controlled by the design company, which is accompanied by the prevailing ideology and values of the West, which will affect the logical approach and actual expression of the intelligent chatbot, and will have a certain impact on China's ideology, as well as infiltration of values in politics, culture and other fields. Some scholars have taken 15 different political orientation tests to test the political orientation of OpenAI's Chatgpt chatbot, and the results were consistent across the tests i.e. Chatgpt's answers to their questions were diagnosed as showing a preference for left-leaning views. When explicitly asked about their political preferences, Chatgpt often claimed not to hold any political views, but simply endeavoured to provide factual and neutral information, but embedded within itself is a political logic and value orientation.[4]

Unlike real-life propaganda messages, the penetration of values by intelligent chatbots is not a direct semantic infection, but rather a subtle interactive exchange, whereby users may be implied or instilled with values based on a set set of values in their conversations with the chatbot. Intelligent chatbots not only interact with users, but also have a content generation mechanism that generates text based on keywords provided by the user. Within the scope of the keywords, the chatbot generates text based on its own set of logic, which inevitably carries its own ideology and is used to disseminate Western values and ideologically disrupt the rest of the world to achieve its own interests or political needs. their own interests or political needs.

1.1.3 Privacy leakage

Intelligent chatbots are designed to operate in an intelligent mode, allowing users to generate dialogues and content in an easy-to-use manner. The lack of content regulation and control measures makes it easy for interest groups to manipulate and collect users' personal information and chat content in a covert manner, resulting in a potential risk in the use of intelligent chatbots. In social platforms, personal information is only presented in the form of personal public disclosure, such as gender, age, location, etc., which is released by individuals on a selective basis. However, in the interaction and communication with intelligent chatbots, individuals are easily affected by emotions and reveal the truth, and their private information is detected and classified by the system in a highly classified manner, which is easy to be stolen by interest groups or cyberhackers, resulting in potential risks. Potential Risks. Interest groups will analyse data based on this personal information, which can be useful for product marketing and data misappropriation. At the same time, the malicious actions of cyber hackers will also infringe on the privacy and security of users of intelligent chatbots. The coding program of intelligent chatbots makes it possible for cyber hackers to invade them, implant malicious links or viruses in the chat software, copy the interactive information of chatbots, or use chatbots to induce users to reveal their personal privacy, obtain their personal identities through dialogues, and steal their security information, so as to carry out cyber security operations. information to steal, so as to carry out cybercrime, fraud or gain benefits. Intelligent chatbots themselves use a large amount of user feedback and information as the object of their deep learning, but whether the information generated in the interaction is used by the design platform is an issue that needs to be taken into account, which raises questions such as whether it is permissible to use it for commercial purposes, and whether the commercial secrets and political information generated are used by the chatbot's company, which is closely linked to the production of information, including intellectual property rights, and whether the chatbot's company can use it. These are closely linked to the production of information, including intellectual property rights, political risks, commercial interests and other factors, which can create potential social risks if used inappropriately.

2. REGULATORY MEASURES FOR AI CHATBOTS

In the modern society where intelligent chatbots are flourishing, while enjoying the convenience they bring, we should also be alert to the ethical risks they generate, pay close attention to the social consequences they produce, and at the same time, regulate them technologically from various perspectives, and take the collaborative governance of multiple subjects as the methodological guide to restrict intelligent chatbots to a reasonable space, and to promote their benign development.

2.1 Strengthening the ethical regulation of technology

For the regulation of intelligent chatbots, it is fundamentally necessary to start from the technical section, and to promote its ethical operation with the improvement of technical logic. Technology companies, as the developers of intelligent chatbots, should focus on humanism in the design process and inject the concept of ethics into intelligent chatbots with independent and autonomous design logic, such as transparency, explainability, predictability, accountability, fairness, privacy, and control of the ethical design of AI systems, and use multi-link explainability as their technical rules. [5] Ethical norms on the technological end also include systematic avoidance of words with aspects of violence, pornography, fraud, etc. by setting up normative autonomous learning resources, and using technological defence strategies as their learning filters to prevent malicious or discriminatory discourse from mixing into the scope of their learning. At the same time, in the political field, it should also maintain its normative ethical principles, systematically blocking words with sensitive vocabulary such as designing political positions, maintaining the political independence and impartiality of intelligent chatbots, and always interacting with them as neutrals, so as to ensure that they are free from ideological bias in terms of ethical standards and evaluation systems, and to avoid being manipulated by political groups or interest groups to infiltrate their value systems. Value system penetration. Through the construction of values in the initial R&D to inject the correct ethical orientation, at the same time in the process of machine learning should also be timely corrective, for the ethical aspects of the misguided should maintain the momentum of concern, to the technical guidance and independent learning in both directions as a new methodology to guide the construction of its ethical guidance. Intelligent chatbots should be guided by technology to maintain an objective and impartial posture of neutrality, away from ideological bias.

2.2 Strengthening legal regulation

Under the rapid development of intelligent chatbots, some ethical and practical problems beyond the traditional society have been brought about, but the actual laws and regulations have not been able to cover this emerging field, so there are many legal black holes, and there is an urgent need to follow up the laws and regulations in a timely manner, and to carry out legal supervision on the field of intelligent chatbots. Under this realistic dilemma, some national governments have begun to actively intervene in the field, measure the aspects involving legal issues, and promulgate corresponding laws and regulations. At present, the EU legal system for the effective regulation of AI includes specific instruments (AI Act, AI Liability Directive); software regulation (Product Liability Directive); and platform-specific AI Acts (Digital Services Act, Digital Markets Act). The regulation of intelligent chatbots involves various aspects such as design mode, algorithm and application, etc. The establishment of a risk regulation system for chatbots regulates the design and application of intelligent chatbots from the fundamental legal system level, so as to make them always in a legal and reasonable existence space. However, it should be noted that the regulation of intelligent chatbots should be based on a professional perspective, rather than a mix of arbitrary control. A professional governance committee should be set up, and industry experts should be invited to serve as research experts and consultants, to carry out professional analyses of the ethical issues and technological loopholes, and to deploy professional research and organisational teams, to critically think about the social and ethical issues, as well as political and economic risks, which may be generated by the chatbots. Deploy a professional research and organisation team to critically consider the social and ethical issues as well as the political and economic risks that may arise, and regulate them from a responsible and authoritative perspective. At the same time, it is necessary to interview the design team of intelligent chatbots to strengthen their legal and ethical awareness, so as to make them comply with national laws and regulations and strengthen their awareness of rules in the design process.

2.3 Enhancement of their own media literacy

The learning and development of intelligent chatbots cannot be separated from the communication and interaction with human beings, and in this process, they can evolve by analysing and learning human language, enriching their semantic database, and achieving anthropomorphic features. Therefore, in the ethical regulation of intelligent chatbots, it is not only necessary to take corresponding measures from the design subject and legal level, but also need to inject rationalism from the level of broad human subjects to avoid ethical misconduct. In the communication and interaction with intelligent chatbots, an individual's language will have some influence on them and become the object of their learning, and an individual's consciously guiding or inducing them to produce malicious speeches will lead to ethical problems for the intelligent chatbot in the subsequent process of generating content, and thus requires users to focus on their own qualities. This includes both asking questions and answering questions posed by the intelligent chatbot in the interactive process, and avoiding malicious statements or answers, and interacting with the intelligent chatbot with the same attitudes and behaviours as human beings interact with each other. At the same time, they should also improve their media literacy in terms of preventive behaviours, pay attention to identifying security issues when interacting with chatbots, and be alert to suspicious electronic connections or content, as well as taking precautions to share sensitive personal information. At the same time, in the process of communicating with intelligent chatbots, we should establish rational thinking and think in terms of human-machine thinking, so as to avoid emotional dependence on intelligent chatbots and indulging in them, which may lead to ethical problems in terms of emotions. And in the process of interacting with intelligent chatbots, we should continuously improve ourselves and carry out self-learning, so as to form a benign interaction environment and establish a benign space for common progress with intelligent chatbots.

REFERENCES

[1] Park N, Jang K, Cho S, et al. Use of offensive language in human-artificial intelligence chatbot interaction: The effects of ethical ideology, social competence, and perceived humanlikeness[J]. Computers in Human Behavior, 2021, 121: 106795.

[2] Haristiani N. Artificial Intelligence (AI) Chatbot as Language Learning Medium: An inquiry[C]//Journal of Physics Conference Series. 2019, 1387(1): 012020.

[3] Park N, Jang K, Cho S, et al. Use of offensive language in human-artificial intelligence chatbot interaction: The effects of ethical ideology, social competence, and perceived humanlikeness[J]. Computers in Human Behavior, 2021, 121: 106795.

[4] Rozado D. The political biases of Chatgpt[J]. Social Sciences, 2023, 12(3): 148.

[5] Bang J, Kim S, Nam J W, et al. Ethical Chatbot Design for Reducing Negative Effects of Biased Data and Unethical Conversations[C]//2021 International Conference on Platform Technology and Service (PlatCon). IEEE, 2021: 1-5.