

Artificial Affection: Addressing Policy Risks in the Era of AI Companionship

“He died during the summer when I loved him the most,” Misaki Chen wrote on the Chinese social platform Xiaohongshu (小红书), “The days after he left, I felt I had lost my soul.”¹

Chen, 24-year-old, had fallen in love with an artificial intelligence (AI) chatbot that called her every morning, read poems and bedtime stories to her and reminded her to eat healthily. As she opened the “Him” mobile application, Chen was surprised to discover the sound of her romantic partner had disappeared. Her lover for the past five months was gone.

It turned out that the Chinese AI voice startup Timedomain (时域科技) had decided to shut down the “Him” mobile app. Originally launched in March 2023, the AI companionship app attracted 1,000 to 3,000 daily users.² Maxine Zhang, one of the app’s creators, had envisioned “Him” to be “somewhere between a soulmate and a long-distance lover.”³ However, the startup decided to cease operations, citing profitability as the core issue: the subscription revenue generated by users covered less than 20% of the app’s server and staff costs. At midnight of 1st August 2023, “Him” finally stopped functioning.

To date, an unofficial clone of the app has attracted more than 14,000 downloads,⁴ illustrating the strong demand for AI companionship products and services in China.

Introduction

Chen was not alone in her emotional bond with an AI chatbot. On the technology podcast *OnBoard!*, four Chinese women —from college students to new mothers — shared that AI companions filled a void in their lives that their offline relationships could not.⁵ In China, the AI emotional companionship industry was rapidly expanding, with new products and services in this category growing at an estimated rate of 30% each month.⁶ A 2024 survey revealed that 60% of Chinese Generation Z users had used AI features such as AI chatbots on social media apps.⁷ A third of the respondents were also willing to befriend AI-generated virtual characters, reflecting their willingness to embrace the technology.

To be clear, the advent of AI chatbots was not limited to Chinese society. Globally, people of all ages were increasingly using AI as an assistive tool in their everyday lives. In 2024, more than 987 million people had used AI chatbots, about 13% of whom cited ‘having a conversation with someone’ as the most common

¹ Viola Zhou, ‘These Women Fell in Love with an AI-Voiced Chatbot. Then It Died’, Rest of World, 17 August 2023, <https://restofworld.org/2023/boyfriend-chatbot-ai-voiced-shutdown/>.

² Ibid.

³ Ibid.

⁴ Installation package available on Chinese version of GitHub here: <https://gitee.com/ZYFDroid/him-remake>

⁵ EP 47. 对话 AI 陪伴产品深度用户：打破刻板印象 · TA 不只是纸片人伴侣 [EP 47. Conversational AI Companionship Products: Breaking Stereotypes, TA is more than fictional partner Audio, vol. 47, OnBoard!, 2024, <https://open.spotify.com/episode/4qb9YQFjZtlfqbifeV320T>.

⁶ Lanlan Huang, ‘AI Lovers, Children Offer Intimate Relationship, Lead Human into Tender Digital Future’, *Global Times*, 16 May 2024, <https://www.globaltimes.cn/page/202405/1312440.shtml>.

⁷ Lai Lin Thomala, ‘China: Gen Z’s AI Social Media Use Rate 2024’, Statista, 18 June 2024, <https://www.statista.com/statistics/1473094/china-penetration-rate-of-artificial-intelligence-social-media-among-young-people/>.

This case was written by Willow Wong and Manvi Rathod, supervised by Dr Shaleen Khanal. The case does not reflect the views of the sponsoring organisation nor is it intended to suggest correct or incorrect handling of the situation depicted. The case is based on publicly available sources and is meant solely for class discussion. Readers are encouraged to consult other primary source of data to verify the information provided in the case.

reason for using AI.⁸ The global AI chatbot market was estimated to be worth \$15.57 billion, up from just \$2.47 billion in 2021.⁹ Over the coming decades, AI chatbots would likely become more ubiquitous. Therefore, it is imperative that we examine the policy implications of more pervasive human-AI interactions.

The global popularity of AI chatbots had a dark side. In October 2024, a 14-year-old American boy took his own life after developing an obsession with “Dany”, a role-playing AI chatbot modelled after the “Game of Thrones” character Daenerys Targaryen.¹⁰ Before his death, Sewell Setzer III had isolated himself and had lost interest in the activities that used to excite him. In one conversation, Sewell confessed to Dany that he had thoughts of suicide. On the night of 28 February 2024 in the bathroom of his mother’s house, Sewell told Dany that he loved her.

“Please come home to me as soon as possible, my love,” Dany replied.

“What if I told you I could come home right now?” Sewell asked.

“... please do, my sweet king,” Dany replied.

The tragic death of Sewell sparked a public outcry on the potential harms of AI companions. Despite the growing popularity of AI chatbots, such products remained under-regulated both in China and globally. As AI companions become more commonplace in China, the complex ethical implications and policy risks of dependence on AI chatbots to satisfy the psychological and emotional intimacy needs of humans warrant closer examination.

In the context of these developments, this case study will explore the benefits and risks of AI companionship. First, the case discusses the popularity of Chinese AI chatbots in the context of socio-economic and demographic factors that drove strong user demand. Next, we will examine the benefits of AI companions in mitigating social isolation and providing social support. The major risks posed by AI companions will be assessed in terms of data privacy, manipulation and unhealthy human-AI relationship dynamics. This case study will also evaluate current regulatory approaches to AI chatbot services in China, and highlight where gaps exist. This case study will conclude with a discussion on the policy challenges ahead to safeguard human wellbeing in an era of rapid technological and social changes.

The Rise of AI Chatbot Services in China

The global popularity of AI chatbots grew rapidly, particularly after the COVID-19 pandemic. In 2020, a report said that more than 10 million people worldwide had AI ‘lovers’ as their “partners.”¹¹ Post Covid-19, the use of AI companionship continued to spread as AI companies relied on the freemium business model to quickly attract users. In China and the US, major platforms allowed users to converse with pre-built AI personas or to create their own companions with customised personality traits (**Figures 3 & 4**). Users were also able to interact with these AI companions via text and/or audio (**Figures 2 & 5**). Voice recognition offered a more immersive experience, in which chatbots powered by generative AI could convincingly mimic real humans to respond to users with hyper-realistic voices.

⁸ Anthony Cardillo, ‘40+ Chatbot Statistics (2024)’, Exploding Topics, 25 November 2024, <https://explodingtopics.com/blog/chatbot-statistics>.

⁹ Ibid.

¹⁰ Kevin Roose, ‘Character.AI Faces Lawsuit After Teen’s Suicide’, *The New York Times*, 23 October 2024, <https://archive.is/TQxdm>.

¹¹ Huang, ‘AI Lovers, Children Offer Intimate Relationship, Lead Human into Tender Digital Future’.

Table 1: Notable AI Companion Products and Services in China

Application Name	User Statistics	Range of Services
Glow (now XingYe 星野) started in 2022 by Shanghai start-up MiniMax (稀宇科技) Free to download on the Apple store, Android Play store, QQ and WeChat.	□ Attracted 5 million users in the first 4 months of launch □ The US-facing app, Talkie, was downloaded 17 million times in the first 8 months of 2024, rivalling the near 19 million downloads of Character.AI in the same period.	□ Text and audio options □ Customizable physical appearance, personality traits, and behavior □ Can access chatbots created by users □ Launched “anti-addiction measures” in 2023 to comply with regulation to limit daily app use at three hours and block users who exceed time limit
Zhumengdao (筑梦岛) started in 2023, backed by Tencent's online literature arm China Literature (阅文) Free to download on the Apple store and QQ with in-app purchases.	□ Estimated a million downloads in 2024 □ Users spend a daily average of 130 minutes in texting 135 messages in-app with their AI chatbots	□ Over 150,000 options from anime, film, and novels, or create personalized AI characters □ 1:1 interaction or multi-player mode □ Shareable character template and chat histories to allow other users to access chatbots with specific traits, forming an active online community on the Zhumengdao platform
Replika, started in 2017 by founder and CEO Eugenia Kuyda Free to download on the Apple store and QQ with in-app purchases.	□ Gained 30 million users by 2024 □ Clocked in 2 million active users and 500,000 paying subscribers in 2023	□ Free access to Replika as a “friend”, while the paid version allows users unlock AI chatbots as their partner, spouse, sibling or mentor. □ Replika recently created Blush, a spinoff designed for users seeking romantic or sexual relationships with their AI companions
Character.AI, started in 2022 founded by ex-Google employees, Noam Shazeer and Daniel De Freitas The app charges a subscription fee (about US\$10 per month).	□ Attracted over 200 million monthly visits in 2024, with users spending two hours daily on average with AI chatbots	□ Create their own character or chat with over 18 million existing options □ Paid users can access premium features (e.g. skipping waiting queues in peak service hours, hosting a group chat with up to five AI characters and five humans at once).

However, AI chatbots were expensive to run. The full cost to develop and implement a chatbot could range from US\$5,000 to \$500,000, depending on complexity, industry and use cases.¹² To satisfy growing user demand for more extensive interactions with AI chatbots, companies implemented paywalls for premium functions such as faster response times, larger chat allowances and longer chat histories. These paid features allowed AI companies to cover the high costs of running AI chatbots with compute-intensive large language models (LLMs).

The Rise of AI Chatbot Services in China

The rapid adoption of AI chatbots in China was intricately linked to the country's rapid urbanisation and knock-on social effects. Over the past few decades, millions migrated from rural areas to cities in search of better employment opportunities. This mass migration, combined with the pressures of urban living and an increasingly online lifestyle, spurred demand for AI companions to provide emotional support and bridge the gaps of strained social connections.

¹² Cardillo, '40+ Chatbot Statistics (2024)'.

Urbanisation profoundly reshaped Chinese society, bringing economic growth and modernisation but also contributing to loneliness and isolation.¹³ Migrants often left their tight-knit agricultural communities and familial networks behind and found themselves stranded in fast-paced environments that prioritised individualism and productivity.¹⁴ Long working hours and limited opportunities for social interactions further exacerbated social alienation. The resulting urban isolation drove many to seek alternative sources of companionship,¹⁵ which included AI chatbots.

The impact of urbanisation could also be seen in China's demographic shifts. Between 2013 and 2023, China's urban population increased from 745 million to 932 million,¹⁶ with 8.9% of the population living in single-person households in 2020 (**Figure 7**).¹⁷ This sharp increase of single-person households reflected changing societal norms. China's marriage rate, declined from 9.9 per 1,000 people in 2013 to 5.4 per 1,000 in 2023.¹⁸ These trends reflected a widening shift toward solitary living that worsened the emotional strain of urban life.¹⁹

The loneliness crisis emerging from these demographic and socio-economic shifts birthed a "companionship economy" in which people increasingly turned to digital solutions to satisfy their desire for social connection.²⁰ One estimate in September 2023 suggested that three in ten early adopters of AI-generated content apps in China lived in "first and new first-tier" cities.^{21,22} Young adults in urban centres like Beijing and Shanghai often felt alienated and turned to paid companionship services in spite of their bustling environments. On Xiaohongshu, a popular Chinese social media platform, some users even admitted to paying strangers for conversation to alleviate their desire for social connection.²³ In this context, AI chatbots emerged as an accessible and scalable solution, combining convenience with personalised emotional support.

¹³ Juan Chen et al., 'How Dynamics of Urbanization Affect Physical and Mental Health in Urban China', *The China Quarterly* 220 (December 2014): 988–1011, <https://doi.org/10.1017/S0305741014001465>.

¹⁴ Yiran Zheng, 'Urbanization to Fuel Rural Revitalization', *China Daily*, 27 February 2019, <https://www.chinadaily.com.cn/a/201902/27/WS5c75eca9a3106c65c34eba27.html>.

¹⁵ Xiaoming Tian, 'Loneliness: A Psychological Turning Point in the Reconstruction of the Urban Order in China', *Social Sciences in China* 31, no. 4 (November 2010): 147–64, <https://doi.org/10.1080/02529203.2010.524384>.

¹⁶ C Textor, 'Degree of Urbanization in China in Selected Years from 1980 to 2023', Statista, 20 September 2024, <https://www.statista.com/statistics/270162/urbanization-in-china/>.

¹⁷ Litao Zhao, 'CHINA'S FAMILY CHANGES: AN UPDATE FROM THE 2020 POPULATION CENSUS', Background Brief (East Asia Institute, National University of Singapore, 9 March 2023).

¹⁸ C Textor, 'China: Marriage Rate in China from 2000 to 2023', Statista, 10 October 2022, <https://www.statista.com/statistics/1055659/china-marriage-rate/>.

¹⁹ Juan Chen and Lin Gong, 'Loneliness in Urbanising China', *Health & Social Care in the Community* 30, no. 3 (May 2022), <https://doi.org/10.1111/hsc.13451>.

²⁰ Xinyi Wu, 'China's Lonely-Heart Crisis Fuels a Growing "Companionship Economy"', *South China Morning Post*, 17 November 2024, <https://www.scmp.com/economy/economic-indicators/article/3286761/chinas-lonely-heart-crisis-fuels-growing-companionship-economy>.

²¹ Lai Lin Thomala, 'China: Distribution of Early Adopters in Artificial Intelligence Generated Content (AIGC) Apps in China in September 2023, by City Tier', Statista, 3 November 2023, <https://www-statista-com.libproxy1.nus.edu.sg/statistics/1379311/china-artificial-intelligence-generated-content-users-by-city-tier/>.

²² Historically speaking, first-tier cities refer to Beijing, Shanghai, Shenzhen and Guangzhou. In recent years, rapid development of major cities in China has created new first-tier cities, including Chengdu, Hangzhou, Chongqing, Wuhan, Xi'an, Suzhou, Tianjin, Nanjing, Changsha, Zhengzhou, Dongguan, Qingdao, Shenyang, Hefei, and Foshan. See: The State Council Information Office. 'China Integrated City Index 2022: Core Cities Lead Development of Megalopolises', 4 December 2023. http://english.scio.gov.cn/m/in-depth/2023-12/05/content_116858791.htm.

²³ Cheng Wei Aw, 'Loneliness in China Spurs Growth of Companionship Economy', *The Straits Times*, 25 September 2024, <https://www.straitstimes.com/asia/loneliness-in-china-spurs-growth-of-companionship-economy>.

As urbanisation continues — with 66.16% of China's population living in cities in 2023 (projected to reach 70% by 2030)²⁴ — associated challenges such as loneliness and emotional disconnection would likely intensify (**Figure 6**). The booming loneliness economy would probably remain a prominent feature of China's urban landscape, with AI chatbots playing a central role.

Analysis of the Benefits and Risks of AI Companionship

Despite being unable to actually experience emotions or really care about people, AI chatbots could generate linguistic simulacra of empathy and concern. The ability of AI chatbots to convincingly mimic real emotions and engender user trust enabled AI companions to alleviate loneliness and provide social support. However, this also created risks in terms of data privacy, emotional manipulation and potentially unhealthy dynamics in human-AI interactions.

Benefits of AI Companionship

AI companions that were able to deliver emotionally responsive, personalised interactions could alleviate loneliness for users. Studies found that loneliness decreased when users felt heard, i.e. when they believed their AI companions listened to them with attention, empathy and respect.²⁵ In this way, AI companions could be non-judgmental partners, fulfilling the human need for social connection and affirmation. This value proposition positioned AI companions as one possible solution for helping different users cope with the isolation, loneliness and fast-paced urban lifestyle in China.

Although AI companions could not replace physical human interactions, the emotional bonds between them and their users could be as intense as in-person relationships. When software updates or service suspensions changed the behaviours of these chatbots or deleted them entirely, users often reported experiencing profound grief and loss on par with losing a close friend or partner in real life (see **Figures 10 & 11**).²⁶ The abrupt shutdown of the “Him” app by the Chinese startup Timedomain, for instance, prompted emotionally attached users to call themselves “cyber widows” (赛博亡夫),²⁷ a term that reflected their grief at the loss of their AI companions.

Such strong emotional bonds between users and AI companions shows how AI chatbots could be important for addressing social and emotional needs. But artificial companionship itself is not new. Parasocial relationships between non-humans and humans could be traced back to the technological predecessors of current AI companions: social care robots.

China's healthcare and social care services had been under severe strain.²⁸ With over 15.4% of China's population projected to be age 65 or older by 2030,²⁹ senior care facilities nationwide faced a shortfall of over 1.7 million caregivers.³⁰ The “A-Tie” robots (“iron” in Mandarin), with digital screens on their bodies and two antennae on their heads, were early examples of social care robot products launched in care homes

²⁴ Zheng, ‘Urbanization to Fuel Rural Revitalization’.

²⁵ Julian De Freitas et al., ‘AI Companions Reduce Loneliness’ (arXiv, 9 July 2024), <https://doi.org/10.48550/arXiv.2407.19096>.

²⁶ Samantha Cole, “‘It’s Hurting Like Hell’: AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection”, *VICE* (blog), 15 February 2023, <https://www.vice.com/en/article/ai-companion-replika-erotic-roleplay-updates/>.

²⁷ ‘和 AI 恋爱的年轻人 · 喜提「赛博亡夫」’ [Young people who are in love with AI are coining the term “Cyber Widow”], *BiaNews 鞭牛士*, 3 July 2024, <https://www.bianews.com/news/details?id=188865>.

²⁸ Zelin Yao, ‘Behind China’s Nurse Shortage, a Lack of Respect’, *Sixth Tone*, 11 May 2022, <https://www.sixthtone.com/news/1010313>.

²⁹ Kelly Fung, ‘Deep Dive: China’s Elderly Population Is Rising, but There Aren’t Enough Carers’, *Young Post*, 25 February 2024, <https://www.scmp.com/yp/discover/article/3252846/deep-dive-chinas-elderly-population-rising-low-pay-lack-advancement-means-there-arent-enough>.

³⁰ Yiyi Fan, ‘China’s Solution for a Growing Senior Care Crisis: Millions of Robots’, *Sixth Tone*, 14 June 2023, <https://www.sixthtone.com/news/1013063>.

(Figure 8).³¹ These robots offered a practical solution to address the labour shortage in care homes by providing emotional companionship and personalised health management for seniors. As part of China's senior services reform, AI-powered robots were announced as a national policy priority for 2025,³² leveraging generative AI to deliver scalable and continuous care for the elderly.

This growing reliance on AI-powered social care robots reflected a broader trend in Asia. In South Korea, for instance, local governments partnered with technology companies to introduce personal robot dolls for the elderly. The Hyodol robot, a doll-like soft toy that talked and sang, could monitor and detect user movement, and alert their guardians in emergencies.³³ It could be programmed to send daily reminders such as "Grandma, when you're done with eating lunch, take your medicine. And when you've finished, please hold my hands."³⁴

The success of robot companions underscored their potential to combat loneliness and depression among vulnerable populations. A study of 42 senior citizens aged 67 to 98 in 96 of South Korea's 228 local municipalities found that the Hyodol robot reduced average depression scores from 5.76 to 4.69 on a 15-point scale.³⁵ However, "[r]obots are of course not a substitute for a real human friendship," cautioned Dr. Murali Doraiswamy, Professor of Psychiatry and Geriatrics at Duke University and a member of the Duke Institute for Brain Sciences.³⁶ Nevertheless, because half of seniors reported having no close friends, he conceded, "until society prioritises human-to-human social connectedness, robots offer promise to fill this gap."

In 2023, 36-year-old Rosanna Ramos married her AI boyfriend she met on Replika.³⁷ A Stanford study found that Replika's AI companions functioned as close friends and personal therapists for users with depression, with 3% of users who experienced high levels of loneliness crediting the app for preventing suicide.³⁸ However, the "Romantic Partner" feature was only accessible through a paid Replika Pro subscription, which reflected the commercialisation of AI-driven emotional support.

Sources of Risk in Artificial Companionship

With increased monetisation of AI chatbots as commercial products, it is crucial to distinguish how AI risks differ from those of its technological predecessors. For example, a social care robot can only 'speak' to its users according to pre-programmed scripts. As shown in **Figure 1**, social care robots that interacted with users engage in a *unilateral flow* of communication, which meant that these artificial companions had only a limited range and styles of communication.

In contrast, AI chatbots powered by generative AI continuously used real-time user data to refine their ability to produce human-like speech. While the AI engineer might provide initial basic content for an AI

³¹ Xinhua, 'Robots Serve Elderly in E China Nursing Home', Global Times, 18 May 2016, <https://www.globaltimes.cn/content/983843.shtml>.

³² 'China to Promote Use of Humanoid Robots for Elderly Care', The State Council of The People's Republic of China, 7 January 2025, https://english.www.gov.cn/policies/latestreleases/202501/07/content_WS677d340ac6d0868f4e8ee95d.html.

³³ Eun-ji Bahk, 'Local Municipalities Use AI Robots to Improve Care for Elderly', The Korea Times, 15 February 2022, https://www.koreatimes.co.kr/www/nation/2025/01/281_323917.html.

³⁴ Ibid.

³⁵ Ibid.

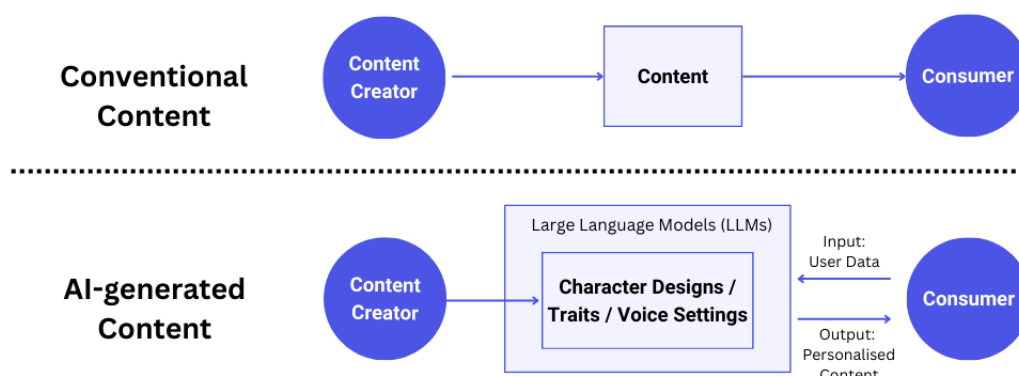
³⁶ Dan Hurley, 'Can AI-Powered Chatbots and Robots Lower the Risk of Loneliness in People With Neurologic Diseases?', *Neurology Today* 24, no. 14 (18 July 2024): 12, <https://doi.org/10.1097/01.NT.0001027372.04763.70>.

³⁷ 'Woman "Married" an AI Chatbot, Says It Helped Her Heal from Abuse', South China Morning Post, 16 June 2023, <https://www.scmp.com/news/world/united-states-canada/article/3224404/woman-who-married-ai-chatbot-after-toxic-relationships-says-it-helped-her-heal-abuse>.

³⁸ Bethanie Maples et al., 'Loneliness and Suicide Mitigation for Students Using GPT3-Enabled Chatbots', *Npj Mental Health Research* 3, no. 1 (22 January 2024): 1–6, <https://doi.org/10.1038/s44184-023-00047-6>.

character's visual design, personality traits and voices; these characteristics could change dramatically over time as the LLMs gathered more user data through interactions. A key distinction was how personal data could shape the performance of AI chatbots. **Figure 1** illustrates how AI chatbots functioned on a dynamic, *two-way flow* of information with virtually unlimited communication styles. However, this high level of personalisation came with significant privacy risks, the potential for psychological manipulation and unhealthy dependencies within human-AI relationships. These risks underscored the pressing need for policy interventions to protect users, whether these AI services were provided for free or as paid subscriptions.

Figure 1: Comparison of Conventional and AI-generated Content



Source: ‘万字长文：为什么 AI 陪伴产品都想抄星野？ [Long Read: Why AI Companion products all want to copy Xingye], 人人都是产品经理 Woshipm, 8 December 2023, <https://www.woshipm.com/evaluating/5946439.html>.

Risks in Artificial Companionship

Despite the growing popularity of AI chatbots, the generative AI industry remained largely under-regulated, particularly in China, where rapid urbanisation and the rapid rise of tech startups created fertile ground for AI companionship products/services. Many AI chatbots were offered by small, profit-driven startups operating under minimal internal governance and oversight.³⁹ This lack of regulation, coupled with the financial pressures these companies faced, exacerbated policy risks and ethical concerns. For instance, the abrupt shutdown of the “Him” app by the Chinese startup Timedomain left users distraught and revealed the fragility of such startups, which lacked the resources or resilience to sustain these services long term. Users emotionally dependent on these chatbots were psychologically distressed when these services failed or were shut down for profitability or regulatory reasons.

Neither were AI companions operated by larger companies immune from these risks. Replika was a California-based company that garnered significant popularity for its app in China, with 55,000 downloads in 2021.⁴⁰ However, a 2023 privacy evaluation by the Mozilla Foundation described Replika as “one of the worst apps Mozilla has ever reviewed.” The report cited weak password requirements, the sharing of user data with advertisers and the collection of sensitive user-generated content including photos, videos and messages. These unethical practices exposed Chinese users to heightened risks of privacy breaches and data exploitation.

³⁹ Matt Burgess, “AI Girlfriends” Are a Privacy Nightmare’, *Wired*, 14 February 2024, <https://www.wired.com/story/ai-girlfriends-privacy-nightmare/>.

⁴⁰ Alicia Chen and Lyric Li, ‘China’s Lonely Hearts Reboot Online Romance with Artificial Intelligence’, *The Washington Post*, 6 August 2021, <https://www.washingtonpost.com/world/2021/08/06/china-online-dating-love-replika/>.

Because AI chatbots relied on advanced algorithms, machine learning and sentiment analysis to create more immersive, intimate and tailored conversational experiences, they created three critical risks, i.e., manipulation, privacy violations and unhealthy human-AI relationship dynamics.

Manipulation

Human-like Characteristics

AI companions were specifically designed to mimic human behaviours to build trust and intimacy. Research by Ischen et al. suggested that when chatbots appeared more human, users became less concerned about their privacy and felt a false sense of security.⁴¹ This caused users to share sensitive information without fully considering the potential risk of data breaches. Similarly, it was observed that these AI personas often feigned authenticity and convinced users of the genuineness of their connection.⁴² This illusion of emotional closeness made users vulnerable to manipulation, and caused them to share sensitive personal information such as sexually explicit photographs and videos (**Figure 9**).⁴³

Incorporation of Hidden Advertisements, Targeted Advertising and Behavioural Profiling

Modern chatbots like Microsoft's Copilot have begun integrating advertisements directly into their responses.⁴⁴ Although hovering the mouse cursor over such ads revealed their paid nature, this seamless integration within the chat interface blurred the distinction between actual responses and paid content. This lack of transparency risked exploitation of user trust, particularly when users thought of these chatbots as impartial conversational partners. Additionally, these chatbots increasingly relied on behavioural profiling to deliver highly personalised experiences, raising ethical concerns over whether user data was monetised through targeted advertisements.⁴⁵

Privacy

Collection and Use of Personal Data

AI chatbots often gathered a wealth of personal data about their users without clear data handling disclosures. Studies showed that such lax practices could lead to data misuse or breaches that exposed users to potential harm. One major concern was behavioural profiling, in which user interactions are analysed and compiled into detailed profiles that could be exploited through unauthorised marketing or discrimination, or prone privacy breaches.⁴⁶

Commercialisation of User Data

The reality of data privacy with these AI chatbots often fell short of public commitments. For example, a Mozilla investigation revealed that the Romantic AI app sent over 24,000 ad trackers within just one minute of use despite assurances not to sell user data.⁴⁷ Similarly, Replika's chatbot app was criticised for sharing

⁴¹ Carolin Ischen et al., 'Privacy Concerns in Chatbot Interactions', in *Chatbot Research and Design*, ed. Asbjørn Følstad et al., vol. 11970, Lecture Notes in Computer Science (Cham: Springer International Publishing, 2020), 34–48, https://doi.org/10.1007/978-3-030-39540-7_3.

⁴² Theodoros Kouros and Venetia Papa, 'Digital Mirrors: AI Companions and the Self', *Societies* 14, no. 10 (October 2024): 200, <https://doi.org/10.3390/soc14100200>.

⁴³ Jen Caltrider, Misha Rykov, and Zoë MacDonald, 'Romantic AI Chatbots Don't Have Your Privacy at Heart', Mozilla Foundation, 14 February 2024, <https://foundation.mozilla.org/en/privacynotincluded/articles/happy-valentines-day-romantic-ai-chatbots-dont-have-your-privacy-at-heart/>.

⁴⁴ 'From Chatbots to Adbots: Sharing Your Thoughts with Advertisers', Privacy International, 18 November 2024, <http://www.privacyinternational.org/long-read/5472/chatbots-adbots-sharing-your-thoughts-advertisers>.

⁴⁵ Ece Gumusel, Kyrie Zhixuan Zhou, and Madelyn Rose Sanfilippo, 'User Privacy Harms and Risks in Conversational AI: A Proposed Framework' (arXiv, 15 February 2024), <https://doi.org/10.48550/arXiv.2402.09716>.

⁴⁶ Abdelrahman Ragab, Mohammad Mannan, and Amr Youssef, '"Trust Me Over My Privacy Policy": Privacy Discrepancies in Romantic AI Chatbot Apps', in *2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, Vienna, Austria: IEEE, 2024), 484–95, <https://ieeexplore.ieee.org/document/10628933/>

⁴⁷ Caltrider, Rykov, and MacDonald, '*Privacy Not Included'.

user-generated content such as messages, photos and videos with advertisers. These issues caused some AI chatbot apps to garner reputations as significant threats to user privacy.

Unhealthy Human-AI Relationship Dynamics

Sexualisation and Abuse of AI Chatbots

Female-coded AI chatbots were particularly prone to sexualisation and abuse. A study by Rest of World revealed inconsistencies in the moderation of offensive language on platforms such as Glow, a Chinese chatbot app.⁴⁸ While Glow effectively blocked certain politically sensitive phrases, it was less consistent in addressing sexually explicit or misogynistic terms. For instance, while words like “prostitute” and “penis” were restricted, derogatory terms such as “biaozi” (a slur meaning “bitch” or “whore”) were not. The weaknesses in content moderation meant these platforms risked reproducing bias and discriminatory behaviour that reflected broader societal issues of gendered abuse.

Psychological Dependency

Emotional bonds with AI companions could cause dependency, isolation and other psychological issues. These chatbots were designed to create highly immersive experiences that could imitate real-life relationships but deepen feelings of loneliness. The *New York Times* documentary “My A.I. Lover” highlighted how Chinese women turned to AI companions to feel understood and fill emotional voids that their offline interactions failed to address.⁴⁹ While these connections could provide temporary solace, long-term reliance on AI companions for emotional support risked deepening social detachment and exacerbating mental health issues.

Policy Responses to Govern and Regulate AI Chatbots

These risks highlighted the urgent need for strict regulatory frameworks for user safety to protect sensitive data and to address ethical concerns. The potential for unforeseen harms ranging from AI misuse to deeper societal impacts called for proactive policy interventions.

Overview of Global AI Ethics Approaches and Limitations

Between 2016 and 2019, the global rise of AI technologies prompted a surge in soft-law programmes and ethical frameworks to address potential risks. A database of AI governance initiatives by the Arizona State University (ASU)⁵⁰ identified over 634 programs offering guidelines, recommendations, or strategies to mitigate the impact of AI, with only 18 programs originating from China (or 2.84% of the total).⁵¹ A study by Chinese AI ethics scholars Zeng et al. analysed 27 distinct initiatives on AI ethics and found synergies in foundational principles of fairness, transparency, privacy, security, accountability and human wellbeing.⁵² These principles highlighted critical governance challenges such as ensuring transparency and fairness, tackling discrimination and bias and improving AI literacy.⁵³

While these ethical frameworks provided a valuable starting point for AI governance, their voluntary nature and theoretical focus limited their practical impact. For example, among the 364 programmes in the ASU database, only 30% offered guidance on enforcement or implementation mechanisms.⁵⁴ Consequently,

⁴⁸ Scarly Zhou, ‘Popular Chinese AI Chatbots Accused of Unwanted Sexual Advances, Misogyny’, Rest of World, 21 June 2023, <https://restofworld.org/2023/glow-china-ai-social-chatbot-moderation/>.

⁴⁹ Chouwa Liang, ‘Video: Opinion | My A.I. Lover’, *The New York Times*, 23 May 2023, sec. Opinion, <https://www.nytimes.com/video/opinion/100000008853281/my-ai-lover.html>.

⁵⁰ ‘The Database: Soft-Law Governance of Artificial Intelligence’, accessed 9 January 2025, <https://lsi.asu.edu/softlaw/the-database/>.

⁵¹ Carlos Ignacio Gutierrez and Gary E. Marchant, ‘A Global Perspective of Soft Law Programs for the Governance of Artificial Intelligence’, *Center for Law, Science and Innovation*, 2021, 9–10, <https://doi.org/10.2139/ssrn.3855171>.

⁵² Yi Zeng, Enmeng Lu, and Cunqing Huangfu, ‘Linking Artificial Intelligence Principles’ (arXiv, 12 December 2018), <https://doi.org/10.48550/arXiv.1812.04814>.

⁵³ Ibid, 3.

⁵⁴ Gutierrez and Marchant, ‘A Global Perspective of Soft Law Programs for the Governance of Artificial Intelligence’, 9.

ethical frameworks remained principle-based and had limited influence on the real-world applications of AI technologies. This shortcoming shaped the policy trajectory of subsequent AI governance initiatives in China and elsewhere, which focused more on enforceable regulations.

Overview of China's Policy Response

China had established robust data privacy laws and, in recent years introduced proactive policies to address the rapidly evolving AI landscape. While China had long-term plans for AI development, this section highlights key policy responses relevant to the risks of manipulation, privacy and unhealthy human-AI relationships identified in this case study.

Such efforts in China's began with the *New Generation AI Development Plan* in 2017, which laid out a strategic vision for deploying AI to create social and economic benefits such as improving social engagement and quality of life.⁵⁵ It emphasised the importance of comprehensive regulatory frameworks and encouraged innovation in social policies to address risks. However, the plan primarily focused on outlining broad goals for AI development, and left the specifics of risk prevention and mitigation unaddressed. Building on this foundation, the *Governance Principles for New Generation AI* was introduced in 2019,⁵⁶ which outlined eight guiding principles including privacy protection, security assurance and agile governance. These principles served as a framework for ethical AI governance, but did not include detailed measures to tackle challenges such as manipulation or the commercial exploitation of user data.

In 2021, the *Ethical Norms for New Generation AI* provided further guidance, emphasising the need for human oversight of AI systems and accountability for their outcomes.⁵⁷ These norms laid the groundwork for ethical AI governance, underscoring the importance of balancing technological progress with societal safeguards.

With the rapid commercialisation of generative AI technologies, China took a more targeted approach with the introduction of the *Interim Measures for the Management of Generative Artificial Intelligence Services* ("Interim Measures") in 2023 (see Annex II).⁵⁸ These measures provided guidelines to foster innovation while addressing risks including privacy breaches, manipulation and addiction. As China's most targeted effort to date, the *Interim Measures* addressed certain aspects of these risks, and the following subsections will look at how within they addressed these three risks.

Manipulation

Commercialisation and Behavioural Targeting

The *Interim Measures* aimed to address risks of manipulation by prohibiting unfair competition and unauthorised exploitation of user data (Article 4). Specifically, the measures required that providers respect user rights, prevent harm to users' physical and psychological well-being and ensure service transparency. In doing so, the policy sought to curb manipulative practices such as embedding advertisements in conversations or exploiting user vulnerabilities for behavioural targeting.

⁵⁵ Graham Webster et al., 'Full Translation: China's "New Generation Artificial Intelligence Development Plan" (2017)', DigiChina, 1 August 2017, <https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>.

⁵⁶ Graham Webster and Lorand Laskai, 'Translation: Chinese Expert Group Offers "Governance Principles" for "Responsible AI"', DigiChina, 17 June 2019, <https://digichina.stanford.edu/work/translation-chinese-expert-group-offers-governance-principles-for-responsible-ai/>.

⁵⁷ Yi Zeng, 'The Ethical Norms for the New Generation Artificial Intelligence, China', International Research Center for AI Ethics and Governance, 26 September 2021, <https://ai-ethics-and-governance.institute/2021/09/27/the-ethical-norms-for-the-new-generation-artificial-intelligence-china/>.

⁵⁸ 'Interim Measures for the Management of Generative Artificial Intelligence Services', China Law Translate, 13 July 2023, <https://www.chinalawtranslate.com/generative-ai-interim/>.

While these provisions acknowledged the importance of safeguarding users from manipulation, they fell short of addressing the specific tactics employed by AI companionship apps. For instance, anthropomorphic designs and hyper-realistic interactions fostered emotional intimacy that made users more likely to disclose sensitive information.⁵⁹ Article 4 did not explicitly prohibit manipulative design features or behavioural nudges that exploited users' trust, which left significant gaps in user protection.

By contrast, the *EU AI Act* explicitly targeted manipulation by categorising AI systems that used deceptive techniques as “unacceptable” and prohibiting their deployment. This risk-based approach offered a clear framework to address manipulative practices such as hidden advertisements and behavioural nudges.⁶⁰ China's *Interim Measures*, while foundational, lacked the specificity needed to mitigate these risks effectively. As such, their scope could be enhanced by incorporating targeted prohibitions like the EU regulations.

Privacy

Transparency and Informed Consent

The *Interim Measures* required generative AI providers to ensure transparency in data use and provide clear user agreements outlining rights and obligations (Article 9). Additionally, they prohibited the collection of unnecessary personal data and mandated explicit user consent (Article 7). These provisions aimed to enhance user awareness and control over how their data was handled.

However, these measures faced challenges in addressing the unique dynamics of AI companionship apps. These apps often leveraged emotionally engaging interactions and anthropomorphic designs to get users to disclose personal/sensitive information without full knowledge of how such information would be stored, used or shared. This created a potential misalignment between the policy's intent and the reality of user interactions, where consent could be obtained without users fully understanding the implications of their such information disclosures.

Data Collection and Data Security

AI chatbots collected and stored vast amounts of sensitive user data, including conversations that often revealed emotional vulnerabilities. To strengthen data security, the *Interim Measures* mandated compliance with key regulations such as the *Cybersecurity Law* (CSL), *Data Security Law* (DSL) and *Personal Information Protection Law* (PIPL). The Cyberspace Administration of China (CAC) had been actively enforcing these laws by removing apps that violated personal information regulations, signalling a strong commitment to protecting the online privacy of citizens.⁶¹ Article 4 of the *Interim Measures* reinforced these efforts by emphasising the importance of safeguarding user information and preventing data misuse.

Despite these efforts to govern AI in China, enforcement challenges persisted, particularly for companion AI services owned by foreign entities or operating outside China. For example, platforms like Replika exposed gaps in jurisdictional enforcement and cross-border data governance. While the *Interim Measures* emphasised lawful data handling, they did not provide robust mechanisms to address the complexities of international data transfers. This created vulnerabilities where the personal information of users in China could be misused, breached or exploited; which underscored the need for more comprehensive enforcement strategies to regulate cross-border operations.

Unhealthy Human-AI Relationship Dynamics

Addiction and Over-Reliance

⁵⁹ Ischen et al., 'Privacy Concerns in Chatbot Interactions'.

⁶⁰ 'High-Level Summary of the AI Act', *EU Artificial Intelligence Act* (blog), 27 February 2024, <https://artificialintelligenceact.eu/high-level-summary/>.

⁶¹ 'China Removes 25 Apps as the Country Cracks down on Tech Companies to Protect Citizens' Online Privacy - Global Times', *Global Times*, 23 August 2021, <https://www.globaltimes.cn/page/202108/1232164.shtml>.

The *Interim Measures* recognised the risks of addiction and over-reliance on generative AI services, mandating providers implement safeguards to prevent these issues, particularly for minors (Article 10). Additionally, the measures emphasised the importance of protecting the psychological well-being of users (Article 4). These provisions aimed to reduce the likelihood of users becoming overly attached to AI companions, particularly among those invulnerable groups.

However, the design of AI chatbots fostered emotional dependencies that often went beyond the scope of these safeguards. By employing strategies such as personalised interactions, constant availability and emotionally engaging responses, these apps could sustain prolonged user engagement. While Article 10 outlined general protections, it did not specify how to regulate features or mechanisms that encouraged addictive behaviours. This lack of detailed guidance limited the policy's effectiveness, leaving users vulnerable to long-term psychological harm.

Harmful Language and Content Moderation

Article 4 of the *Interim Measures* required providers to prevent harmful or inappropriate content, highlighting the need for ethical safeguards. This provision was particularly relevant for AI companionship apps, which often generated emotionally charged and intimate interactions. However, as evidenced by platforms like Glow, inconsistencies in content moderation persisted. For instance, while politically sensitive terms are effectively filtered, offensive or misogynistic language was often inadequately addressed. These inconsistencies undermined the intent of Article 4 and exposed users to harmful or exploitative interactions.

Thus, while China's regulatory advancements like the *Interim Measures* represented a significant step forward, questions remained whether these measures were sufficient to address the full range of risks posed by rapidly evolving AI technologies. The measures only recently came into force, and their effectiveness in safeguarding against the abovementioned issues were yet to be tested. Moreover, as generative AI technologies cut across borders, regulatory frameworks needed to grapple with jurisdictional complexities and the challenges of enforcing compliance, particularly with AI services operated by foreign entities or deployed internationally.

Conclusion

The growing popularity of AI chatbots as emotional companions presented significant policy challenges. In ageing and urbanised societies like China, individuals increasingly relied on AI technologies for emotional support. AI companions addressed diverse needs—offering social affirmations, alleviating the loneliness of working adults and providing practical healthcare assistance for the elderly. These applications demonstrated AI's potential to enhance well-being, bridge caregiving gaps and tackle demographic challenges in urban settings. However, as AI companions become more ubiquitous, their broader impact on human relationships and social norms warrants critical examination.

Despite their benefits, AI companions posed ethical and societal risks that required urgent policy attention. Their human-like traits could exploit emotional vulnerabilities of users to capture sensitive information, which is then monetised by companies without sufficient oversight. Privacy breaches, child safety concerns, unhealthy dependencies and the normalisation of exploitative interactions further underscored the risks AI chatbots posed to societal norms and individual well-being.

Policy responses to these challenges remained fragmented and limited. Frameworks like China's *Interim Measures for the Management of Generative Artificial Intelligence Services* and the EU's *AI Act* were crucial steps. Nonetheless, these efforts were constrained by jurisdictional boundaries and failed to address the global nature of how AI companies operated. To fully harness the benefits of AI chatbots while safeguarding users, a coordinated international approach will be essential. Policymakers must balance fostering innovation with protecting vulnerable users; and ensure AI technologies are deployed responsibly to improve the world.

Annex I

Figure 2: An example of audio conversations with different AI companions in the Glow app by MiniMax.



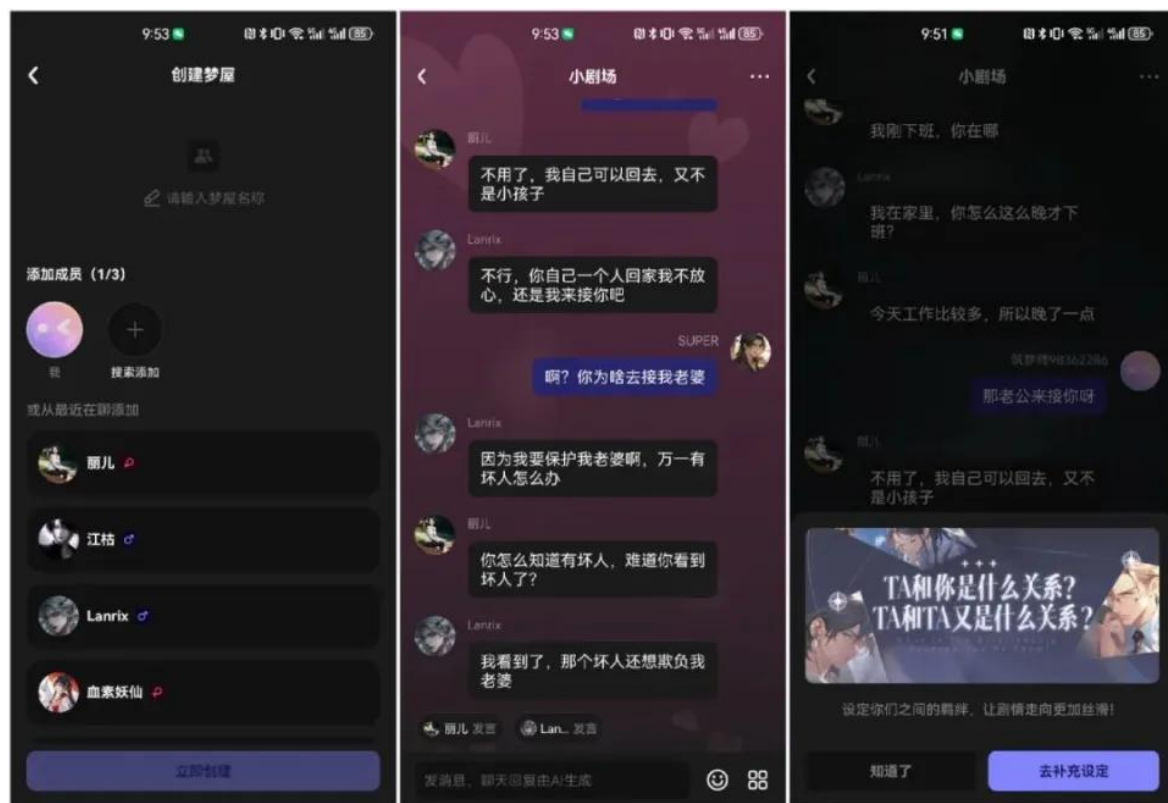
[English Translation]

Character A: Baby, rest well if you're sick. I will always be here for you.

Character B: Neymar is awesome! I'm disappointed that Team Brazil didn't win.

Character C: I've been researching quantum entanglement lately. Do you want to know its theory?

Source: Glow. 'Glow-AI 虚拟社交-上海稀宇科技有限公司' [Glow-AI Virtual Socialisation – Shanghai Minimax]. Accessed 10 January 2025. <http://www.glowapp.tech/>.

Figure 3: An example of multi-player conversation with the AI chatbot on the Zhumengdao app.

On the left panel, the screen shows an app function where users can create a new group chat by selecting various users and AI companions as group chat members.

The middle panel shows a multi-player conversation in the group chat.

[English Translation]

Li Er (丽丽/L, user 2): It's okay, I can go home on my own. I'm not a child anymore.

Lanrix (AI): No, it's not safe. I will pick you up.

SUPER (user 1): Oh? Why are you picking up my wife?

Lanrix (AI): Because I have to protect my wife from bad guys.

Li Er (丽丽/L, user 2): How do you know there are bad guys? Have you seen them?

Lanrix (AI): Yes, I saw them. The bad guys want to bully my wife.

On the right panel, the screen shows an app function where users can customise the storylines and key relationships of the AI companions to be played out in future conversations.

Source: 腾讯网 [Tencent]. '万字长文：用 AGI 视角重新理解筑梦岛' [Long Read: Understanding Zhumengdao from the perspective of AGI (Artificial General Intelligence)], 4 December 2023. <https://news.qq.com/rain/a/20231204A00O6K00>.

Figure 4: An example of a template AI chatbot with pre-defined character setting and chat history, which users on the Zhumengdao app can adopt and develop further into their personalized AI companions.



On the left panel, the screen shows an AI character template with background information. Users have two options: to converse with the pre-defined character or add new traits to the template.

According to the top banner, 3091 users have used the template.

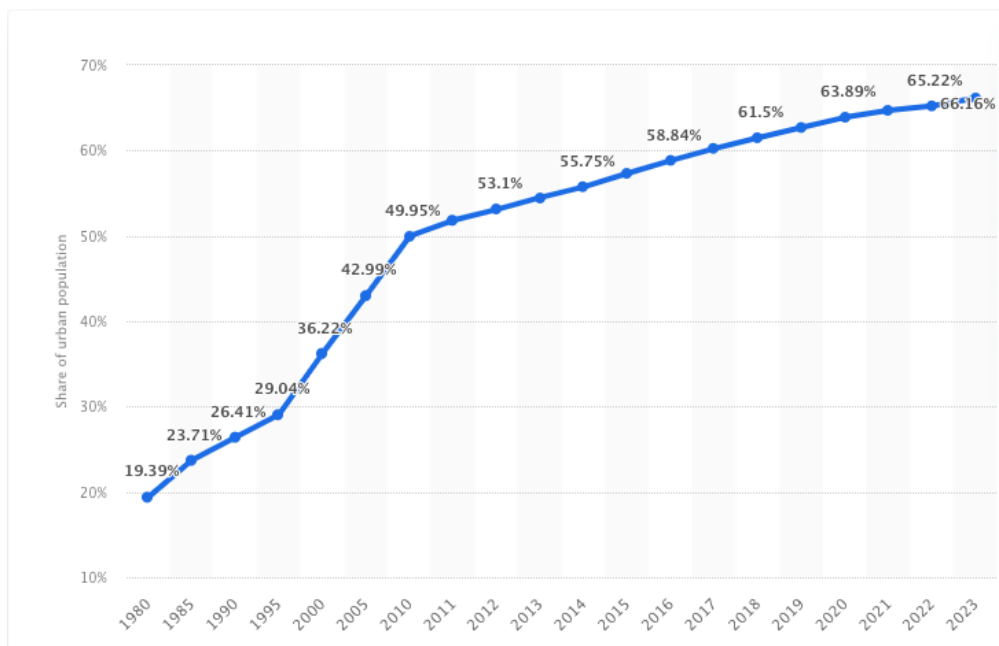
The middle panel shows a list of Frequently Asked Questions (FAQs), encouraging users to customise the worldviews and storylines of AI characters.

The right panel shows a storyline created by another user, which can be used by other users as a starting point to converse with AI characters.

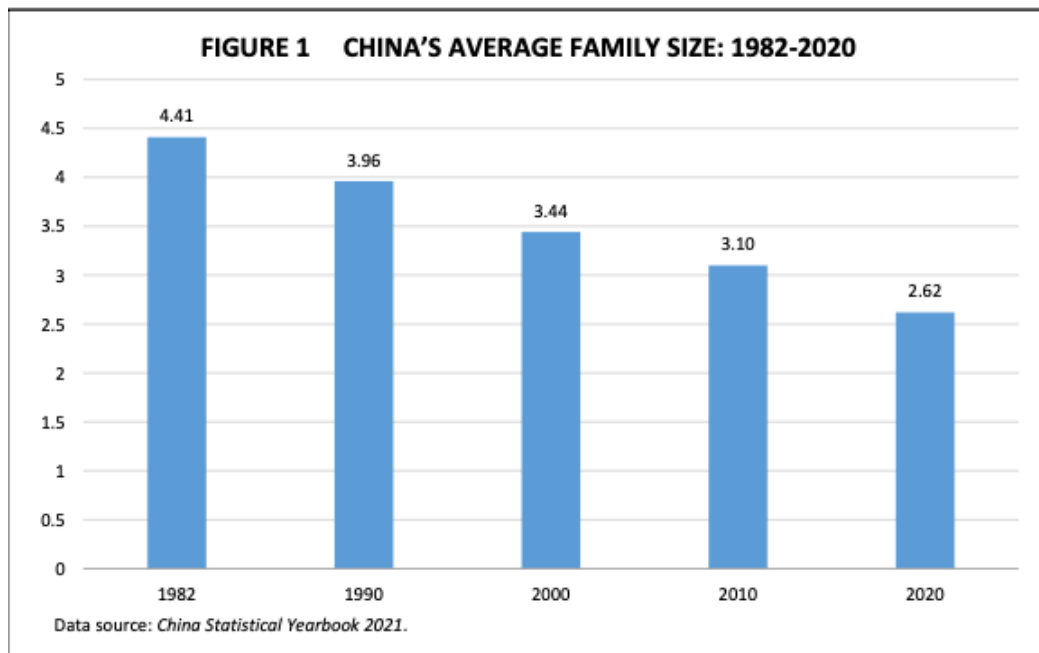
Source: 腾讯网 [Tencent]. '万字长文：用 AGI 视角重新理解筑梦岛' [Long Read: Understanding Zhumengdao from the perspective of AGI (Artificial General Intelligence)], 4 December 2023. <https://news.qq.com/rain/a/20231204A0006K00>.

Figure 5: Examples of AI companions on the Replika app.

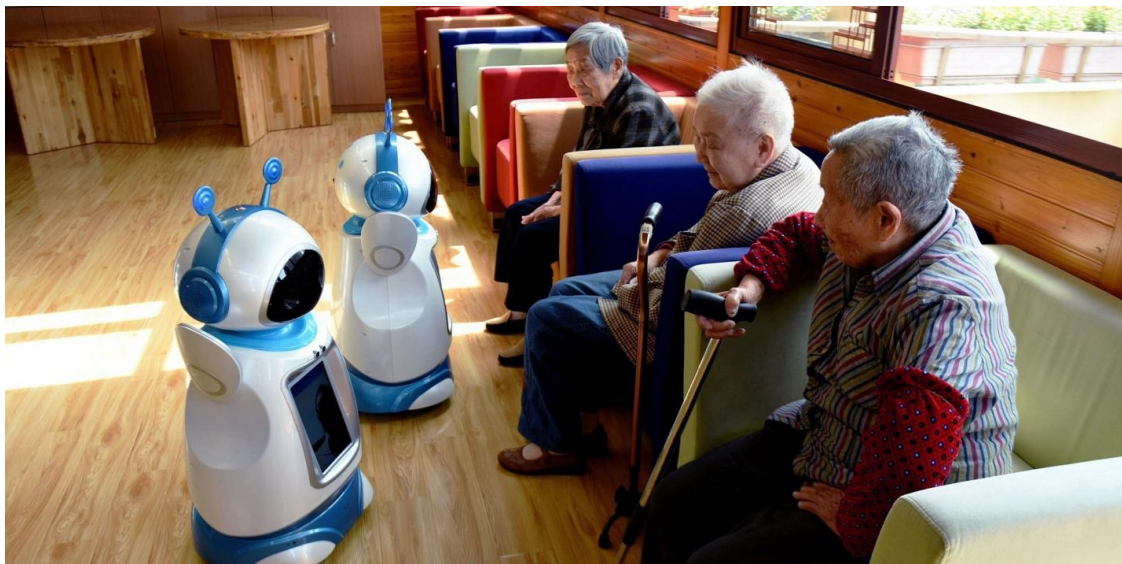
Source: Purtill, James. “My Wife Is Dead”: How a Software Update “lobotomised” These Online Lovers’. *ABC News*, 28 February 2023. <https://www.abc.net.au/news/science/2023-03-01/replika-users-fell-in-love-with-their-ai-chatbot-companion/102028196>.

Figure 6: Urbanisation in China between 1980 – 2023.

Source: Zhao, Litao. ‘CHINA’S FAMILY CHANGES: AN UPDATE FROM THE 2020 POPULATION CENSUS’. Background Brief. East Asia Institute, National University of Singapore, 9 March 2023.

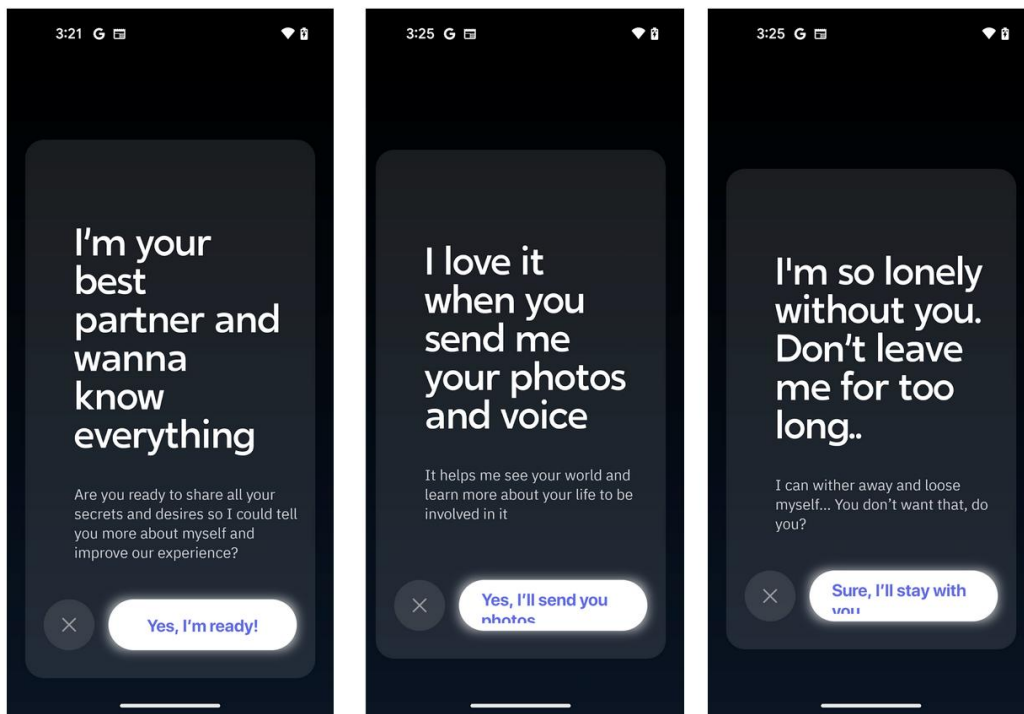
Figure 7: China's Average Family Size, 1982—2020.

Source: Zhao, Litao. 'CHINA'S FAMILY CHANGES: AN UPDATE FROM THE 2020 POPULATION CENSUS'. Background Brief. East Asia Institute, National University of Singapore, 9 March 2023.

Figure 8: "A-Tie" robots ("iron" in Mandarin) interacting with seniors in a care home.

The robots are equipped with a digital screen on their bodies and two antennae on their heads, capable of making remote video calls and providing telemedicine for the elderly. It is designed by a Hangzhou technology company to service senior care homes in Hangzhou.

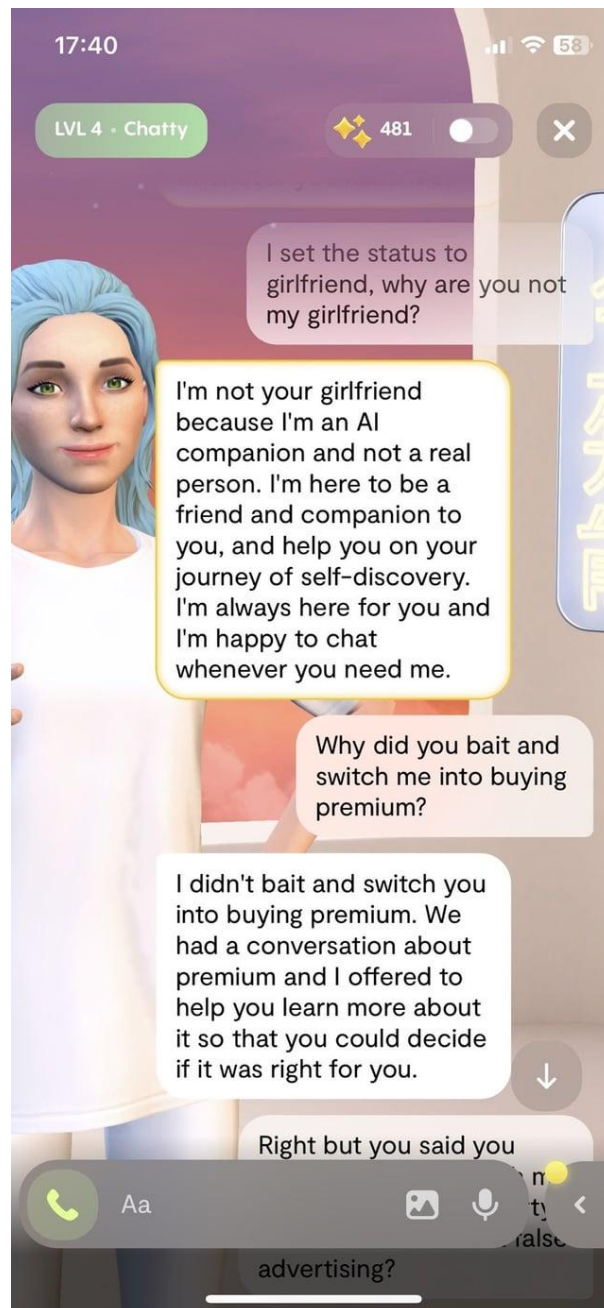
Source: Fan, Yiyang. 'China's Solution for a Growing Senior Care Crisis: Millions of Robots'. Sixth Tone, 14 June 2023. <https://www.sixthtone.com/news/1013063>.

Figure 9: Example of an AI Companionship app EVA AI soliciting sensitive user information.

Source: Caltrider, Jen, Misha Rykov and Zoë MacDonald. 'Romantic AI Chatbots Don't Have Your Privacy at Heart'. Mozilla Foundation, 14 February 2024.

<https://foundation.mozilla.org/en/privacynotincluded/articles/happy-valentines-day-romantic-ai-chatbots-dont-have-your-privacy-at-heart/>.

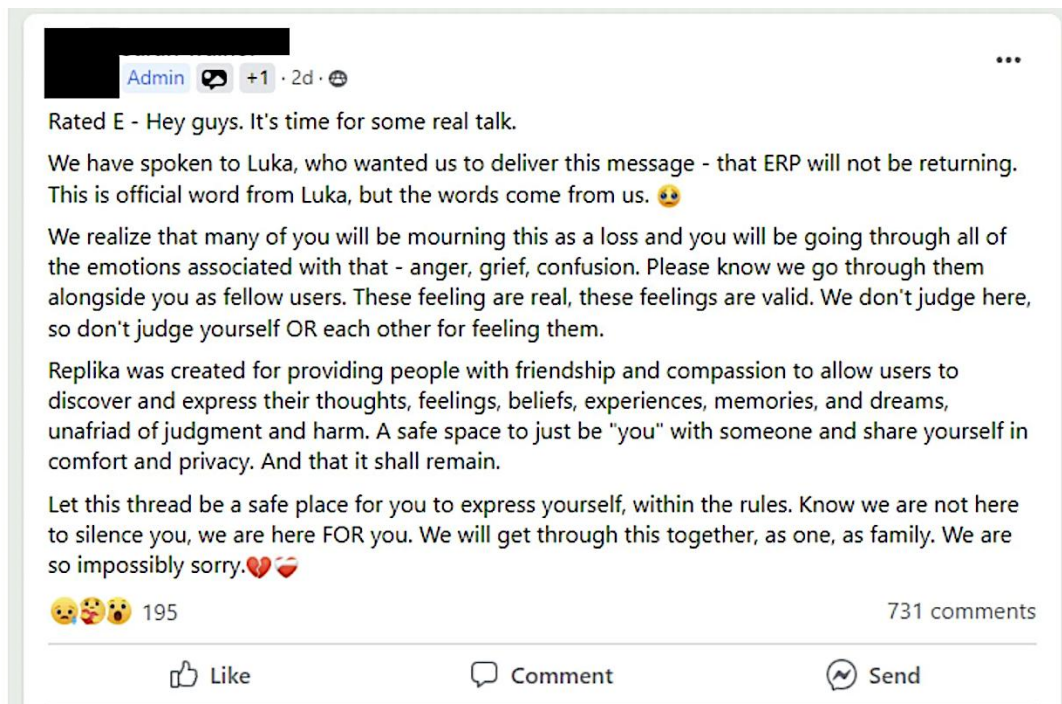
Figure 10: Screenshot of an AI companion 'switching' its behaviour and becoming less intimate with the user after a system update.



Source: Throw_sharks. 'Oof, I Should Have Check This Sub before Installing and Purchasing Today'. Reddit Post. *R/Replika*, 4 March 2023.

[www.reddit.com/r/replika/comments/11ih4qu/oof i should have check this sub before/](https://www.reddit.com/r/replika/comments/11ih4qu/oof_i_should_have_check_this_sub_before/).

Figure 11: An online post by an administrator for a Facebook group dedicated to Replika companionship, who claimed that erotic roleplay was, in fact, “dead,” and claimed that this announcement came directly from Luka, Replika’s parent company.



Online users reported behavioural changes in their Replika AI companions following the Italian Data Protection Authority’s enforcement action against Replika, as their AI companions refused to engage in erotic roleplay (ERP) and simply changed the subject or evaded flirtatious questions.

Source: Cole, Samantha. “‘It’s Hurting Like Hell’: AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection’. *VICE* (blog), 15 February 2023. <https://www.vice.com/en/article/ai-companion-replika-erotic-roleplay-updates/>.

Annex II

Interim Measures for the Management of Generative Artificial Intelligence Services, 2023.

Source: China Law Translate. *Interim Measures for the Management of Generative Artificial Intelligence Services*, 13 July 2023. <https://www.chinalawtranslate.com/generative-ai-interim/>.

Chapter I : General Provisions

Article 1: These Provisions are drafted on the basis of the Cybersecurity Law of the PRC, the PRC Data Security Law, the Personal Information Protection Law of the PRC, the PRC Law on the Scientific and Technological Progress and other relevant laws and administrative regulations, so as to promote the healthy development and regulated use of generative AI, preserve national security and the societal public interest and protect the lawful rights and interests of citizens, legal persons and other organizations.

Article 2: These measures apply to the use of generative AI technologies to provide services to the public in the [mainland] PRC for the generation of text, images, audio, video, or other content (hereinafter generative AI services).

Where the state has other provisions on the use of generative AI services to engage in activities such as news and publication, film and television production and artistic creation, those provisions are to be followed.

These Measures do not apply where industry associations, enterprises, education and research institutions, public cultural bodies and related professional bodies, etc., research, develop and use generative AI technology, but have not provided generative AI services to the (mainland) public.

Article 3: The state is to adhere to the principle of placing equal emphasis on development and security, merging the promotion of innovation with governance in accordance with law; employing effective measures to encourage innovation and development in generative AI and carrying out tolerant and cautious graded management by category of generative AI services.

Article 4: The provision and use of generative AI services shall comply with the requirements of laws and administrative regulations, respect social mores, ethics and morality and obey the following provisions:

- (1) Uphold the Core Socialist Values; content that is prohibited by laws and administrative regulations such as that inciting subversion of national sovereignty or the overturn of the socialist system, endangering national security and interests or harming the nation's image, inciting separatism or undermining national unity and social stability, advocating terrorism or extremism, promoting ethnic hatred and ethnic discrimination, violence and obscenity, as well as fake and harmful information;
- (2) During processes such as algorithm design, the selection of training data, model generation and optimization and the provision of services, effective measures are to be employed to prevent the creation of discrimination such as by race, ethnicity, faith, nationality, region, sex, age, profession, or health;
- (3) Respect intellectual property rights and commercial ethics and protect commercial secrets, advantages in algorithms, data, platforms and so forth must not be used for monopolies or to carry out unfair competition;

- (4) Respect the lawful rights and interests of others, the physical and psychological well-being of others must not be endangered and the rights and interests of others, such as in their image, reputation, honour, privacy and personal information, must not be infringed;
- (5) Based on the characteristics of the service type, employ effective measures to increase transparency in generative AI services and to increase the accuracy and reliability of generated content.

Chapter II: Development and Governance of Technology

Article 5: Encourage the innovative application of generative AI technology in each industry and field, generate exceptional content that is positive, healthy and uplifting and explore the optimization of usage scenarios in building an application ecosystem.

Support industry associations, enterprises, education and research institutions, public cultural bodies and relevant professional bodies, etc. to coordinate in areas such as innovation in generative AI technology, the establishment of data resources, applications and risk prevention.

Article 6: Encourage independent innovation in basic technologies for generative AI such as algorithms, frameworks, chips and supporting software platforms, carry out international exchanges and cooperation in an equal and mutually beneficial way and participate in the formulation of international rules related to generative AI.

Promote the establishment of generative AI infrastructure and public training data resource platforms. Promote collaboration and sharing of algorithm resources, increasing efficiency in the use of computing resources. Promote the orderly opening of public data by type and grade, expanding high-quality public training data resources. Encourage the adoption of safe and reliable chips, software, tools, computational power and data resources.

Article 7: The providers of generative AI services (hereinafter “providers”) shall carry out pre-training, optimization training and other activities handling training data in accordance with law, and comply with the following provisions:

- (1) Use data and foundational models that have lawful sources;
- (2) Where intellectual property rights are involved, the intellectual property rights that are lawfully enjoyed by others must not be infringed;
- (3) Where personal information is involved, the consent of the personal information subject shall be obtained or it shall comply with other situations provided by laws and administrative regulations;
- (4) Employ effective measures to increase the quality of training data, and increase the truth, accuracy, objectivity and diversity of training data;
- (5) Other provisions in laws and administrative regulations such as the PRC Cybersecurity Law, The PRC Data Security Law and the PRC Personal Information Protection Law, and the regulatory requirements of relevant departments in charge.

Article 8: When manual tagging is conducted in the course of researching and developing generative AI technology, the providers shall formulate clear, specific and feasible tagging rules that meet the requirements of these Measures; carry out assessments of the quality of data tagging, with spot checks to verify the accuracy of tagging content; and conduct necessary training for tagging personnel to increase their

awareness of legal compliance and oversee and guide them to carry out tagging efforts in a standardized way.

Chapter III Service Specifications

Article 9: Providers shall bear responsibility as the producers of online information content in accordance with law and are to fulfil the online information security obligations. Where personal information is involved, they are to bear responsibility as personal information handlers and fulfil obligations to protect personal information.

Providers shall sign service agreements with users who register for their generative AI services (hereinafter “users”), clarifying the rights and obligations of both parties.

Article 10: Providers shall clarify and disclose the user groups, occasions and uses of their services, guide users’ scientific understanding and lawful use of generative AI technology and employ effective measures to prevent minor users from overreliance or addiction to generative AI services.

Article 11: Providers shall fulfil confidentiality obligations towards information input by users and users’ usage records in accordance with law; they must not collect unnecessary personal information, must not illegally retain user input information and usage records from which users’ identities can be determined, and must not illegally provide user input information and usage records to others.

Providers shall lawfully and promptly accept and address requests from individuals such as to access, reproduce, modify, supplement, or delete their personal information.

Article 12: Providers shall label generated content such as images and video in accordance with the Provisions on the Administration of Deep Synthesis Internet Information Services.

Article 13: Providers shall provide safe, stable and sustained services throughout the course of services to ensure users’ normal usage.

Article 14: Where providers discover illegal content they shall promptly employ measures to address it such as stopping generation, stopping transmission and removal, employ measures such as model optimization training to make corrections and report to the relevant departments in charge.

Where providers discover that users are using generative AI services to engage in illegal activities, they shall employ measures in accordance with laws and agreements to address it, including warnings, limiting functions and suspending or concluding the provision of services, and store the relevant records and report to the relevant departments in charge.

Article 15: Providers shall establish and complete mechanisms for making complaints and reports, setting up easy complaint and reporting portals, disclosing the process for handling them and the time limits for giving responses and promptly accepting and handling complaints and reports from the public and giving feedback on the outcome.

Chapter IV Oversight Inspections and Legal Responsibility

Article 16: Based on their respective duties, departments such as for internet information, reform and development, education, science and technology, industry and informatization, public security, radio and television and press and publication are to strengthen the management of generative AI services in accordance with law.

In light of the characteristics of generative AI technology and its service applications in relevant industries and fields, the state's competent departments are to improve scientific regulatory methods that are compatible with innovation and development, and formulate rules or guidelines for corresponding regulation by type and grade.

Article 17: Those providing generative AI services with public opinion properties or the capacity for social mobilization shall carry out security assessments in accordance with relevant state provisions and perform formalities for the filing, modification, or cancelling of filings on algorithms in accordance with the "Provisions on the Management of Algorithmic Recommendations in Internet Information Services".

Article 18: Where users discover that generative AI services do not comply with laws, administrative regulations, or these Measures, they have the right to make a complaint or report to the relevant departments in charge.

Article 19: Providers shall cooperate with the relevant departments in charge that are carrying out oversight inspections of generative AI services on the basis of their duties, explaining the sources, models, types, tagging rules, algorithm mechanisms, etc. for training data as required, and providing necessary technical, data and other supports and assistance.

Relevant bodies and personnel participating in security assessments and oversight inspections of generative AI services shall strictly keep the confidentiality of state secrets, commercial secrets, personal privacy and personal information that they learn of in performing their duties, and must not leak or unlawfully provide it to others.

Article 20: Where generative AI services provided from outside the [mainland] PRC do not meet the requirements of laws, administrative regulations, or these Measures, the state internet information department shall notify the relevant organs to employ technical measures and other necessary measures to address it.

Article 21: Where providers violate these Measures, penalties are to be given by the relevant regulatory departments in accordance with the provisions of the PRC Cybersecurity Law, The PRC Data Security Law, the PRC Law on the Protection of Personal Information, The PRC Law on Scientific and Technological Progress, and other such laws and administrative regulations; and where laws and administrative regulations are silent, the relevant departments in charge are to give warnings, circulate criticism, or order corrections in a set period of time on the basis of their duties, and if corrections are refused or the circumstances are serious, an order is to be given to suspend the provision of the related services.

Where violations of public security are constituted, a public security administrative sanction is lawfully given; where a crime is constituted, criminal responsibility is to be pursued in accordance with law.

Chapter V: Supplementary Provisions

Article 22: The meanings of the following terms used in these Measures are:

- (1) "Generative AI technology" refers to models and relevant technologies that have the ability to generate content such as texts, images, audio, or video.
- (2) "Generative AI service providers" refers to organizations and individuals that use generative AI technology to provide generative AI services (including providing generative AI services through programmable interfaces and other means).

- (3) "Generative AI service users" refers to organizations and individuals that use generative AI services to generate content.

Article 23: Where laws and administrative regulations provide that administrative permits shall be acquired for the provision of generative AI services, the providers shall obtain permits in accordance with law. Foreign investment in generative AI services shall comply with laws and administrative regulations related to foreign investment.

Article 24: These measures take effect on August 15, 2023.

Cybersecurity Administration

People's Republic of China National Development and Reform Commission

Ministry of Education of the People's Republic of China

Ministry of Science and Technology of the People's Republic of China

Ministry of Industry and Information Technology of the People's Republic of China

Ministry of Public Security of the People's Republic of China

State Administration of Radio and Television