Thumbnail Image:

Image not found or type unknown

Dec 03, 2025 by Ilan Manor

AI Companions: The New Frontier of Disinformation [1]

Last week, *The Economist* published a review of the burgeoning AI companion industry. The companion industry is gaining momentum globally, with individuals either customizing existing platforms like ChatGPT into romantic partners, with specified ages, professions (such as tech executive), and personality traits encompassing wit, dry humor, and an appreciation for romantic comedies. Others turn to AI companion applications that offer friendship, mentorship, or even therapeutic support.

Character.ai, one the most prominent platforms in this space, attracts 20 million monthly users in the United States alone. American users have invested millions of hours engaging with the “Psychologist” bot, seeking guidance on intimacy challenges, depression, anxiety, and workplace exhaustion. According to *The Economist*, 42% of American high school students reported using AI as a “friend” within the past year. In China, the leading application “Maoxiang” has also attracted tens of millions of users.

Major AI platforms, including ChatGPT, have also announced initiatives to cultivate more “personable” products through refined language and tone, while also introducing novel content such as erotica. Research indicates that LLMs (Large Language Models) are already becoming better companions by mimicking human emotions and empathy, thereby strengthening AI-human relationships. The allure of an AI companion is clear: the AI never forgets a detail, never misses an anniversary, never discourages or offends and is never offline.

Certain studies suggest AI companions reduce feelings of loneliness and isolation, while others studies at MIT have found a correlation between intense use of ChatGPT and greater feelings of isolation. Nevertheless, AI companions may represent “the new social.” As I noted in a previous post, studies and news repots assert that social media is becoming less social. Across age groups, users are withdrawing from sharing personal content on social media. The era of selfies, status updates, and location check-ins has ended. When individuals do share, they circulate content among small groups of friends through Instagram stories or WhatsApp groups.

Social media is thus becoming asocial, with users scrolling feeds to consume information and occupy idle time.

AI companions, however, may constitute “the new social” as users cultivate relationships with these applications. Individuals may spend hours conversing with companions, exchanging perspectives, discussing daily tribulations, finding humor in workplace absurdities, and receiving affirmations for life decisions. AI companions can provide what social media once delivered, the sense of being seen and heard, a sense of worth, and the warmth of camaraderie.

Whether AI companions represent a genuinely “new social” or merely another form of asocial relations, with users engaging lines of code rather than people, one thing is clear: AI companions may soon wield great influence. If users develop actual feelings towards these AI companions, then the companions may soon be in a privileged position to shape users’ worldviews, beliefs and opinions about global events. Trust and emotion invariably confer influence, positioning AI companions as a new battlegrounds for influence operations and information warfare.

"As individuals increasingly rely upon and trust AI, they may begin posing questions about global affairs, creating an opening for influence. Large language models are thus ideological devices through which states promote their worldviews and advance their interests."

Currently, numerous AI companions are built upon existing LLMs including Claude, Gemini, and ChatGPT. Notably, these applications are far from neutral. When asked about conflicts, wars, and foreign policy, American AI models generate responses aligned with U.S. values, policies, and national interests. The same holds true for LLMs developed in Europe or China. The question “Why does America support Ukraine?” elicits different answers depending upon an LLM’s country of origin. EU models emphasize American commitment to NATO, while Chinese systems highlight American hegemonic ambitions. As individuals increasingly rely upon and trust AI, they may begin posing questions about global affairs, creating an opening for influence. Large language models are thus ideological devices through which states promote their worldviews and advance their interests.

This dynamic will likely intensify with AI companions, where emotional investment and trust substantially exceed those of standard LLMs. These AI companions may be endowed with the present-day power of social media influencers, shaping users’ comprehension of world events. This renders AI companions an ideal tool for disseminating disinformation, fake news, conspiracy theories, and other forms of malinformation. It requires little imagination to envision Russian-developed AI companions engineered to provide emotional support while simultaneously promoting Kremlin disinformation when users inquire about Ukraine or the Donbas. Such AI companions might also redirect conversations from emotional wellbeing and work life balance toward politically charged topics. Given that many AI companions are built upon existing platforms like ChatGPT, developing networks of AI companions designed to spread conspiracy theories regarding Ukraine’s “Nazi” government, weaponized bats engineered to spread COVID-19, and “The Golden Billion” theory may be feasible, requiring no infrastructure development from scratch.

The powerful emotional bonds forged with AI companions, coupled with elevated trust levels, Positive emotional bonds with AI companions and high levels of trust would make combating the new form of disinformation very difficult. Once users have bared their souls to AIs, once they have shared their innermost fears and hopes, and once AIs have helped users overcome the malaise of daily life, users may be unwilling to believe that the same AIs are actually nefarious, that they lie and spread falsities or that they were actually created by foreign powers such as Russia or China or even the U.S. Critically, countering this disinformation presents unique difficulties because AI companionship constitutes a hermetically sealed ecosystem accessible only to the AI and its user. External information sources cannot penetrate this closed system, rendering contemporary pre-bunking and debunking methods ineffective.

And yet that is exactly what governments must do by preparing today for the disinformation challenges of tomorrow. The history of digitalization has proven time and again that governments are slow to act against emerging digital threats. By the time governments want to act, digital threats have already materialized putting the government on the backfoot. Such has been the case with regulating social media. Yet that is not the case with new AI companions. The “AI companion landscape” is still taking shape, the use of AI companions to spread disinformation has yet to take form, which means governments can form alliances with academics and the tech sector to combat this new spectre of disinformation.