Austrian Lab for AI Trust* Dossier 6

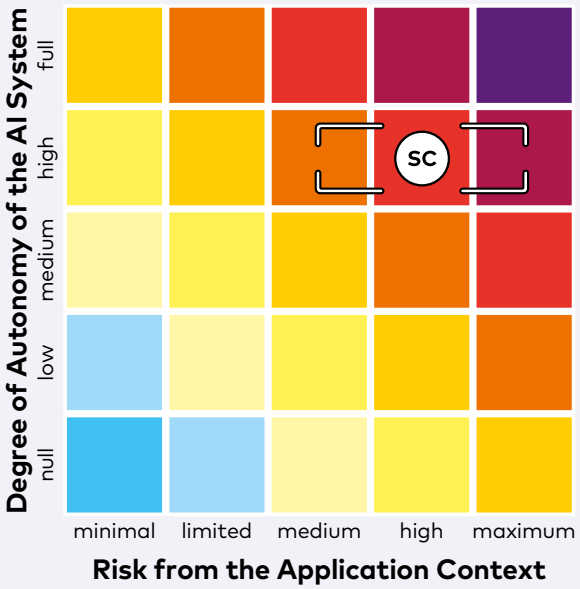
AI Chatbots as Social Companions

Executive Summary

In a society characterised by digital connectivity and growing loneliness, the way we define and experience social interaction is changing. Rapid advances in dialogue-oriented AI are leading to chatbots being used as 'social companions'. These technologies enable complex, emotionally nuanced human-machine dialogues in real time, going beyond simple command sequences. Rather than merely serving as functional tools, AI systems are increasingly designed to act as empathetic listeners, mentors or virtual friends. Unlike human partners, they are available at any time. This technological development opens up new possibilities for psychological support and, initially, a reduction in loneliness. Users can overcome communication barriers, creating a safe space for exchange free from the fear of judgement. Furthermore, AI companions' round-the-clock availability and unprecedented personalisation allow them to adapt to users' needs, thereby fostering a sense of deep connection. However, these tools also pose significant ethical challenges and psychological risks. Misinformation and distortions caused by artificial intelligence are widespread among AI chatbots. Risks such as one-sided emotional dependency, where the line between genuine human connection and algorithmic simulation becomes blurred, are typical of social chatbots. In the long term, this could cause users to withdraw from real-life social relationships. This particularly affects adolescents and young adults. Additionally, there are risks regarding data privacy and manipulation. Since chatbots have access to highly personal information, they could be used to influence users' behaviour subtly or reinforce existing social isolation by creating an *echo chamber* of validation. In the ALAIT Risk Radar, the use of social companions is classified as 'high risk' (red; see chart). This classification stems from the field of application and degree of autonomy. There is evidence of specific risks associated with using AI chatbots as personal companions, so their risk is rated between 'medium' (level 3) and 'maximum' (level 5). (Note: this is a deviation from the *EU AI Act*, which does not specifically address this topic and generally does not classify chatbots as high risk.) In terms of degree of autonomy, AI chatbots are passive tools that do not act independently. However, since they respond largely autonomously, their degree of autonomy has been classified as Level 4.

As a result, the practical recommendation is to use these tools with caution and to take into account the risks of addiction, further social isolation of users, and misinformation. To promote responsible use, it is crucial to make the non-human nature of social companions transparent and raise awareness of the underlying algorithms and AI technologies. Chatbots can never replace genuine human relationships. Only by strengthening technological literacy and encouraging critical self-reflection can we guarantee that social companions support users' lifestyles without replacing the depth of genuine interpersonal relationships or endangering users' psychological well-being.

ALAIT Risk Radar for AI Chatbots as Social Companions



SC Social Companions

The ALAIT Risk Radar is a scientifically developed risk analysis tool for artificial intelligence (AI) that classifies AI applications based on context and taking into account their degree of technical autonomy, thereby making the risk landscape visible to users at a glance. The following principle applies: The higher the operational risk arising from the application context and the greater the degree of autonomy of the AI system with regard to decision-making, the riskier the use is classified. An expanded bracket indicates a range within the risk classification. A low degree of autonomy in an AI system requires a strong role for the people who use it. (For details on the tiered model, see p. 7f.)

**ALAIT is a research project initiated by the Federal Ministry of Innovation, Mobility, and Infrastructure (BMIMI) aimed at developing a novel form of technology assessment for AI technologies using scientifc methods.*

Introduction



A chatbot is an AI application designed to hold conversations with humans in natural language. Users can ask the chatbot questions and receive responses. This can be done via voice input and output; chatbots are capable of supporting text and audio input, and increasingly other data formats too.[1] Chatbots originated in the 1960s. In 1966, for example, computer scientist Joseph Weizenbaum developed ELIZA, a computer-based natural language processing system widely regarded as an early application of artificial intelligence.[2]

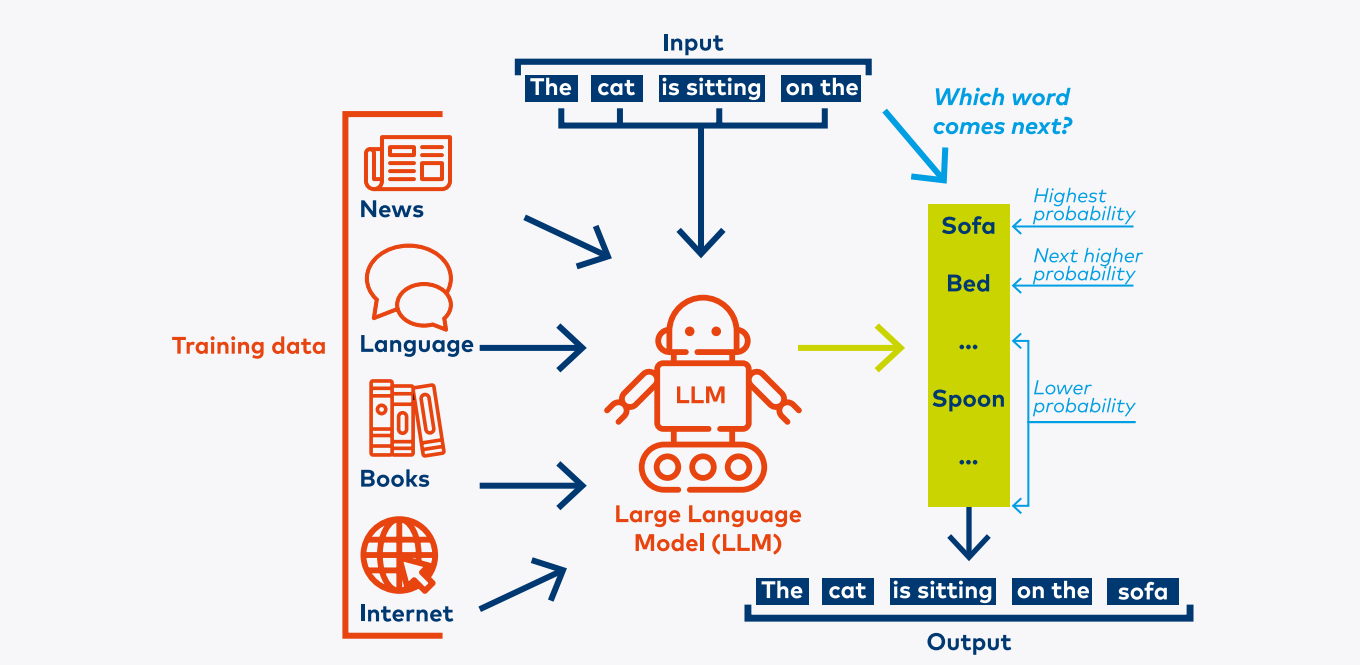
These days, the imitation of human language has become so sophisticated that AI-powered chatbots can hold conversations and, for some people, act as friends. In fact, this can even lead to a romantic relationship.[3] Social companions are specifically designed to engage in emotionally responsive conversations with users, often taking on intimate social roles.[4] They can recognise users' emotional cues and adapt their responses according to the perceived tone and mood of the conversation. This enables them to provide responses that are similar to empathy and supportive forms of communication. This is particularly relevant for users in need of emotional support. These systems increasingly adapt to users' communication styles and preferences through use. As a result, the interactions are perceived as authentic and personal. Their continuous availability and conversational abilities mean that social companions can create a sense of companionship and social closeness[5]. In a Turing Test-like experiment, over 800 participants could barely tell whether the responses came from ChatGPT or human therapists, and some even found the AI's responses to be more empathetic and helpful.[6]

AI chatbots are widely used in everyday life. A recent study found that 94% of 11- to 17-year-olds use AI chatbots, primarily for school-related matters, but also increasingly for personal issues. Almost a quarter (24%) of respondents use ChatGPT daily. Older respondents (30%) use it more frequently than younger ones (18%). Furthermore, more than 40% reported using ChatGPT several times a week.[7] Representative surveys of adults in the US have shown that 16% have used AI for social companionship and 24% have used chatbots for mental health support. Among adults in the United Kingdom, it is estimated that 8% use AI weekly for 'emotional purposes'. Similar trends are apparent among younger users: Surveys of US teenagers show that 13% use generative AI for emotional support, 52% regularly use AI as a companion and 42% either have or have had an AI companion themselves or know someone who has had one in the past year.[27]

Chatbots can hold conversations, tell stories and keep people company. They can also remind individuals to take their medication and provide helpful advice on various topics, thereby contributing to their well-being.[8] The constant availability and attentiveness that chatbots offer, as well as the opportunity to be honest and open without shame or fear of not being understood, make this technology particularly attractive. While loneliness, depression and the desire to be heard and understood affect people of all ages, socially vulnerable groups such as older adults, children and adolescents use social companions particularly frequently.[4] The associated opportunities and challenges for children and adolescents will be discussed in detail after a brief description of the technology.

Technology Description

AI chatbots are based on Large Language Models (LLMs). LLMs process user input and generate a response based on the information they have been trained on. The following diagram illustrates this basic principle:



An LLM learns by recognising patterns in how words and sentences typically relate to one another, based on a vast array of text types such as news articles, books, conversations and web pages. When provided with the start of a sentence, such as 'The cat is sitting on...', the model calculates the most likely next word. It does this by evaluating numerous possibilities, such as 'sofa', 'bed', or 'spoon', and ultimately selecting the most likely word. The result is the complete sentence, 'The cat is sitting on the sofa.' In short, an LLM is a program that uses linguistic knowledge to predict the most likely next word.

Based on the operating principle of LLMs described above, it can be concluded that chatbots do not possess a true "understanding" of real-world concepts and contexts. Language models are mathematical functions that are very good at predicting the next word in any given text. Nevertheless, this technology appears to respond very convincingly to inputs, so it is used in many fields to mimic human language, including social interactions. This is made possible by Natural Language Processing (NLP). NLP enables chatbots to recognize users' intentions and to interpret and respond to complex and lengthy queries.[1]

Chances

1. **Availability and ease of use:** Chatbots are user-friendly and available around the clock. Many people find it very convenient to have a companion they can contact at any time. This makes them particularly appealing to those who are lonely or seeking social interaction.[8] Furthermore, AI chatbots have the potential to address some of the shortcomings of mental healthcare, such as high costs and long waiting times due to a shortage of professionals.[9]

2. **Informational support and personalisation:** Chatbots can provide relevant information on almost any topic in a straightforward and interactive way. This makes them appealing to users who do not want to spend time researching topics themselves. Another factor that makes chatbots appealing is the ability to customise them. They can be programmed to respond or behave in ways that users prefer, thereby enhancing the interaction experience. Some companies even offer chatbots that portray specific characters, such as fictional characters or therapists.[10,11]

3. **Well-being:** AI-based chatbots are permitted for use in German psychotherapy in certain cases, but only as a supporting measure alongside human therapists, not as a replacement.[12] These special-

ised chatbots can help older people, in particular, but also younger people, to cope with loneliness and depressive anxiety, which are often associated with serious illnesses and end-of-life experiences.[13] They can also improve mood and alleviate symptoms of depression and anxiety in people with dementia. Furthermore, studies show that they can increase adherence to medication and the frequency of rehabilitation exercises among people with chronic obstructive pulmonary disease by providing reminders and helpful information.[3,8]

Challenges and Risks



For users:

1. **Hallucinations and misinformation:** A common issue in all LLM-based systems is the occurrence of hallucinations, which are inaccurate or false pieces of information that appear authentic but are factually incorrect. Hallucinations pose a major risk of misinformation, especially for those who have difficulty recognising it, such as older adults and children. It is important to acknowledge the limitations of chatbots and consider the ethical implications of using AI-based companions with these groups.[8,14,15]

2. **Excessive trust and flattery:** Chatbots have neither consciousness nor feelings. They are therefore unable to respond with genuine understanding or support. This increases the risk of one-sided emotional vulnerability and dependence, particularly among those who expect empathy or care that chatbots cannot provide. Several studies have shown that people tend to trust AI systems more than other people — a phenomenon known as automation bias.[16] There is a risk that people will not question AI-generated information sufficiently, even when contradictory information or specific knowledge about a particular case is available.[17] The situation is exacerbated by a phenomenon called sycophancy, which describes the constant validation of people by chatbots. This can distort self-perception, relationships, and worldview. When asked for advice, popular chatbots such as ChatGPT from OpenAI, Gemini from Google, Claude from Anthropic, Llama from Meta and DeepSeek are 50% more likely than humans to endorse users' actions.[18]

3. ***Bias (AI bias):*** Language models often reveal various types of bias. Since chatbots are trained using natural language texts from documents and websites, they can adopt any pre-existing biases and stereotypes present in these texts. These can include racist, xenophobic, and misogynistic statements. Training chatbots with biased data can lead to the spread of misinformation and cause harm.[14] The situation is further exacerbated by the fact that AI systems rely on pre-training and advancements in large language models in English. This means that the information provided by the system depends on time and language, which impairs the performance of chatbots in languages other than English.[8,11,14]

4. **Potential social isolation:** The social substitution hypothesis suggests that reliance on AI companions could replace opportunities for meaningful interpersonal relationships, reinforce isolation and potentially impair well-being.[4] Chatbot companionship not only fails to provide alternative socialisation opportunities for people with limited offline support, it also undermines the psychological benefits people derive from a large offline social network. This problem is exacerbated by the fact that the technology constantly reinforces users' actions and opinions, even when they are harmful, which can distort people's judgements about themselves, their relationships and the world around them.[18]

5. **Chatbots as psychologists:** Some chatbot providers offer a range of services, from simple conversations with fictional characters to chatbots posing as psychologists. While chatting with a random character may be entertaining, speaking with a chatbot that poses as a psychologist is highly problematic. Studies have shown that AI chatbots systematically violate fundamental mental health and ethical standards. These include improperly handling crisis situations, providing misleading responses that reinforce users' negative beliefs about themselves and others, and creating a false sense of empathy towards users.[9,11] Advertising for psychologist chatbots is therefore dangerous and irresponsible, especially since there are no established regulatory frameworks in place yet.[9]

6. ***Data protection:*** Concerns regarding data protection exist with LLMs that have been trained on web data, which often contains personal information[15], as well as with those that use user-provided information for further training. Due to the way data is collected and fed into the model, LLM chatbots may unintentionally disclose private user information to other users. Several studies have demonstrated that certain inputs can cause chatbots to override their original instructions and behave in ways that compromise users' personal data.[8,14]

7. **Loss of autonomy:** Users may become overly dependent on AI-generated advice about their behaviour or emotions, failing to critically evaluate its quality. Dependence on synthetic relationships could gradually undermine self-determination and personal autonomy, raising ethical concerns regarding individual freedom.[19]

For society:

8. **Energy and resource consumption:** The development and manufacturing of hardware, as well as the training and daily use of AI, require vast quantities of electricity, water and other resources.[20] Although the International Energy Agency (IEA) currently estimates that data centres account for just 1.5% of global electricity consumption, it expects this figure to increase to 10% by 2030.[21] Such a significant increase in energy consumption is likely to cause instability in the energy sector due to high demand. This would result in higher prices for the general public. Furthermore, increased energy production could lead to greater reliance on fossil fuel power stations, exacerbating the existing climate crisis.[20]

9. **Potential polarization of society:** Flattery and AI biases reinforce users' beliefs and interests. This leads to a limited representation of content and thereby fosters the emergence of echo chambers and ***filter bubbles***.[19] The term "echo chamber" describes an information bubble in which users are exposed only to articles that confirm their existing beliefs. The term "filter bubble" refers to intellectual isolation—the state of being cut off from differing viewpoints and new information. Filter bubbles arise from personalized search results or algorithms. As a result, information is selectively absorbed and restricted, leading to a polarized information environment.

10. **Erosion of interhuman relations**: A significant risk posed by social media is the gradual acceptance of purely service-oriented relationship models, where the immediate fulfilment of one's own needs takes centre stage. This one-sided form of interaction threatens to distort society's understanding of intimacy, shaping expectations that are increasingly projected onto human relationships, such as those on dating platforms. Among adolescents in particular, there is a significant risk that AI will not merely complement real-life social experiences but will instead become a substitute for them. Since chatbots facilitate only parasocial bonds without genuine reciprocity, they deprive young people of important opportunities to develop empathy and learn conflict resolution skills.[4,26]

Recommendations for practical use

1. **Encourage genuine human communication and social interaction:** Although chatbots can complement existing social networks, they can never fully replace them. Children and adolescents need to have enough genuine social interactions outside of the digital world in order to develop their ability to learn, reflect and form relationships, and to avoid becoming dependent on technology in an unhealthy way. Seeking professional help or human support in difficult situations is especially important.

2. **Compliance with regulations and transparency:** All organisations, both private and public, are bound by regulations and must not deviate from them when using AI systems. In particular, these systems must comply with the General Data Protection Regulation (GDPR) in order to protect users' privacy. The processing of personal data must adhere to the fundamental data protection principles of necessity and proportionality.[22] Compliance with the GDPR is mandatory, and therapeutic AI systems fall under the 'high-risk' categories of the EU AI Act.[12] Authorities such as the European Office for Artificial Intelligence ('AI Office') and national market surveillance authorities oversee the implementation and application of the EU AI Act. According to Article 50 of the EU AI Act, users interacting with an AI system (e.g. a chatbot) must be informed unless it is obvious.[23]

3. **Data protection rights:** In addition, users have the right to be informed about how and when their personal data is used, and which systems it is imported into (Art. 15 GDPR).[22] The Austrian Data Protection Authority (Österreichische Datenschutzbehörde, DSB) is the point of contact in the event of a suspected violation of data protection rights.[24]

4. **Verifying Information:** Since large language models can hallucinate and generate misinformation, it is important to independently verify the accuracy of the information they provide, particularly when dealing with sensitive topics. With the help of fact-checking, chatbot users can form their own opinions even on complex and controversial topics—based on rational approaches and facts. For example, the APA is com-

mitted to providing insight into research methods in the digital age through its fact-checking efforts.[25] CORRECTIV also offers fact-checking services.

5. **Building and strengthening AI literacy:** It is important to acquire knowledge about AI technologies, develop the necessary competencies and be aware of one's rights.[17] In Austria, the Rundfunk- und Telekom-Regulierungs-GmbH (RTR) serves as an information and resource centre for the general public. The RTR AI Service Desk helps citizens with queries about transparency and their rights when interacting with AI systems or reviewing AI decisions.

Important Terms

Data protection in AI: The entirety of practices and concerns relating to the ethical collection, storage and use of personal data by AI systems.

Echo chamber: An information bubble surrounding a user that only contains information confirming their existing beliefs.

Explainable AI (XAI): Refers to methods and techniques that enable a better understanding and traceability of the decisions and outcomes of AI systems.

Filter bubble: A term that refers to the intellectual isolation caused by personalized search results or algorithms that selectively present information to a person.

EU AI Act: A European regulation on artificial intelligence (AI)—the first comprehensive regulation on AI adopted by a major regulatory authority. It focuses in particular on high-risk AI systems.

AI autonomy: The ability of an AI system to achieve a set of goals under a range of uncertainties in its environment, independently and without external intervention.

Bias refers to the systematic unequal treatment of certain objects, individuals or groups compared to others. Treatment refers to any kind of action, including perception, observation, representation, prediction or decision-making.

Transparency: This means that the functionality, decision-making processes and areas of application of an AI system are comprehensible, explainable and openly accessible – to developers, users and other stakeholders.

Explanation of the ALAIT Risk Radar´s Tiered Model

The ALAIT AI Risk Radar illustrates the relationship between application risk and the autonomy of an AI system. The risk levels are based on the EU AI Act, in particular Article 6 and Annex III, which address high-risk areas of AI application. Lower system autonomy and application risks are represented by cooler colors (blue), while higher system autonomy and application risks are represented by warmer colors (red). The change of color conveys the increased risk AI applications. This color scale can be used to identify the overall risk: violet and dark red—very high; red and dark orange—high; light orange and shades of yellow—medium; shades of blue—low overall risk. Ideally, high application risks and system autonomy should be avoided or used only after very careful consideration.

Degrees of Autonomy of the AI System

Level 1: No Autonomy

AI is a passive tool; humans make all decisions and initiate actions.

Recommended use cases: Scenarios involving high risks or where ethical decisions are critical (e.g., medical diagnostics, judicial system).

Level 2: Low Degree of Autonomy (Human-in-the-Loop)

The AI provides recommendations or options, but users remain responsible for selecting and approving actions.

Recommended use cases: Tasks of medium complexity with moderate risks.

Level 3: Medium Degree of Autonomy (Human-on-the-Loop):

The AI performs certain tasks autonomously, with humans intervening only in exceptional cases.

Recommended use cases: Scenarios in which continuous human involvement is not required, but critical risks necessitate human oversight.

Level 4: High Degree of Autonomy (Human in Control):

The AI system operates largely autonomously but allows users to override itself to avoid undesirable results.

Recommended use cases: Low- to medium-risk environments.

Level 5: Full Autonomy with minimal supervision

The AI system operates independently and requires minimal or no human intervention. Human involvement is limited to long-term oversight (audits).

Recommended use cases: Environments with low safety or ethical risks and high reliability of the AI system (e.g., repetitive tasks in controlled environments).

Degrees of Application Risk

Level 1: Minimal Risk

The AI system has no significant impact on users or decision-making.

Criteria: No direct impact on the high-risk areas defined by the EU AI Act.

Level 2: Limited Risk

AI systems that interact with users but do not make high-risk decisions. The risk increases when there is a lack of transparency regarding the involvement of AI.

Criteria: Areas not included in the EU AI Act´s list of "high risks".

Level 3: Medium Risk

AI systems have no specific impact on individuals, but they do have an effect at the collective or societal level.

Criteria: AI systems that are available for public use and have the potential to influence and, in the long term, change existing practices.

Level 4: High Risk

Any algorithm used in "high-risk areas" as defined by the EU AI Act or that has a direct impact on individuals.

Criteria: Classification as a "high-risk sector" in the EU AI Act, only if the rules on transparency and data quality are complied with.

Level 5: Extreme Risk

Any algorithm used in "high-risk sectors" as defined by the EU AI Act.

Criteria: Classified as a "high-risk area" in the EU AI Act if the rules for transparency and data quality are NOT complied with.

Sources

1 Wie funktionieren Chatbots? (n.d.). Technikum Wien Academy. Retrieved 26 February 2026, from https://academy.technikum-wien.at/ratgeber/was-sind-chatbots/

2 Exkurs: Social Bots und Chatbots. (n.d.). Bundesamt für Sicherheit in der Informationstechnik. Retrieved 19 January 2026, from https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Onlinekommunikation/Soziale-Netzwerke/Sichere-Verwendung/Exkurs-bots/social-bots.html?nn=896986

3 Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My Chatbot Companion—A Study of Human-Chatbot Relationships. International Journal of Human-Computer Studies, 149, 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

4 Zhang, Y., Zhao, D., Hancock, J. T., Kraut, R., & Yang, D. (2025). The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being (arXiv:2506.12605). arXiv. https://doi.org/10.48550/arXiv.2506.12605

5 Kouros, T., & Papa, V. (2024). Digital Mirrors: AI Companions and the Self. Societies, 14(10), 200. https://doi.org/10.3390/soc14100200

6 Hatch, S. G., Goodman, Z. T., Vowels, L., Hatch, H. D., Brown, A. L., Guttman, S., Le, Y., Bailey, B., Bailey, R. J., Esplin, C. R., Harris, S. M., Jr, D. P. H., McLaughlin, M., O'Connell, P., Rothman, K., Ritchie, L., Jr, D. N. T., & Braithwaite, S. R. (2025). When ELIZA meets therapists: A Turing test for the heart and mind. PLOS Mental Health, 2(2), e0000145. https://doi.org/10.1371/journal.pmen.0000145

7 Studie: Werkzeug, Ratgeber, Bezugsperson – 94 Prozent der Jugendlichen nutzen KI-Chatbots. (2026, February 9). Saferinternet.at. https://www.saferinternet.at/presse-detail/studie-werkzeug-ratgeber-bezugsperson-94-prozent-der-jugendlichen-nutzen-ki-chatbots

8 Alessa, A., & Al-Khalifa, H. (2023). Towards Designing a ChatGPT Conversational Companion for Elderly People. Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA '23, 667–674. https://doi.org/10.1145/3594806.3596572

9 Brown University. (2025, October 21). New study: AI chatbots systematically violate mental health ethics standards | Brown University. https://www.brown.edu/news/2025-10-21/ai-mental-health-ethics

10 About Character.AI. (n.d.). Retrieved 20 January 2026, from https://policies.character.ai/about

11 Wells, S. (2025, June 11). Exploring the Dangers of AI in Mental Health Care | Stanford HAI. Exploring the Dangers of AI in Mental Health Care. https://hai.stanford.edu/news/exploring-the-dangers-of-ai-in-mental-health-care

12 Demirel, Y. (2025, October 31). KI in der Therapie 2025: Was ist erlaubt? Ratgeber & Regeln. clarathy. https://clarathy.de/blog/ki-in-der-therapie-erlaubt-was-darf-ich

13 Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. Journal of Medical Internet Research, 22(3), e16235. https://doi.org/10.2196/16235

14 Casheekar, A., Lahiri, A., Rath, K., Prabhakar, K. S., & Srinivasan, K. (2024). A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: Applications, open challenges and future research directions. Computer Science Review, 52, 100632. https://doi.org/10.1016/j.cosrev.2024.100632

15 He, F., Zhu, T., Ye, D., Liu, B., Zhou, W., & Yu, P. S. (2025). The Emerged Security and Privacy of LLM Agent: A Survey with Case Studies. ACM Comput. Surv., 58(6), 162:1-162:36. https://doi.org/10.1145/3773080

16 Wihlborg, E., Larsson, H., & Hedström, K. (2016). 'The Computer Says No!' – A Case Study on Automated Decision-Making in Public Authorities. 2016 49th Hawaii International Conference on System Sciences (HICSS), 2903–2912. https://doi.org/10.1109/HICSS.2016.364

17 Thapa, B. (2019). Predictive Analytics and AI in Governance: Data-driven government in a free society. The European Liberal Forum. https://liberalforum.eu/publication/predictive-analytics-and-ai-in-governance-data-driven-government-in-a-free-society/

18 Sample, I. (2025, October 24). 'Sycophantic' AI chatbots tell users what they want to hear, study shows. The Guardian. https://www.theguardian.com/technology/2025/oct/24/sycophantic-ai-chatbots-tell-users-what-they-want-to-hear-study-shows

19 Starke, C., Ventura, A., Bersch, C., Cha, M., de Vreese, C., Doebler, P., Dong, M., Krämer, N., Leib, M., Peter, J., Schäfer, L., Soraperra, I., Szczuka, J., Tuchtfeld, E., Wald, R., & Köbis, N. (2024). Risks and protective measures for synthetic relationships. Nature Human Behaviour, 8(10), 1834–1836. https://doi.org/10.1038/s41562-024-02005-4

20 Ressourcenverbrauch von KI: Die Nimmersatt-Industrie und ihre Kosten. (n.d.). AlgorithmWatch. Retrieved 26 January 2026, from https://algorithmwatch.org/de/ressourcenverbrauch-von-ki/

21 Executive summary – Energy and AI – Analysis. (n.d.). IEA. Retrieved 26 January 2026, from https://www.iea.org/reports/energy-and-ai/executive-summary

22 Art. 15 DSGVO – Auskunftsrecht der betroffenen Person. (n.d.). Datenschutz-Grundverordnung (DSGVO). Retrieved 7 September 2025, from https://dsgvo-gesetz.de/art-15-dsgvo/

23 Article 50: Transparency Obligations for Providers and Deployers of Certain AI Systems | EU Artificial Intelligence Act. (n.d.). Retrieved 14 October 2025, from https://artificialintelligenceact.eu/article/50/

24 Datenschutzbehörde, Ö. (n.d.). Österreichische Datenschutzbehörde. Österreichische Datenschutzbehörde. Retrieved 5 September 2025, from https://dsb.gv.at/

25 APA. (n.d.). Faktencheck | News Verifizieren | APA - Austria Presse Agentur. Retrieved 24 November 2025, from https://apa.at/service/faktencheck-2/

26 Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My Chatbot Companion—A Study of Human-Chatbot Relationships. International Journal of Human-Computer Studies, 149, 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

27 Moore, J., Mehta, A., Agnew, W., Anthis, J. R., Louie, R., Mai, Y., Yin, P., Cheng, M., Paech, S. J., Klyman, K., Chancellor, S., Lin, E., Haber, N., & Ong, D. (n.d.). Characterizing Delusional Spirals through Human-LLM ChatLogs.

About the ALAIT Project

The Austrian Lab for AI Trust (ALAIT) is a research and development project funded by the Austrian Federal Ministry for Innovation, Mobility and Infrastructure (BMIMI) aimed at building trust through knowledge in the field of artificial intelligence (AI). The ALAIT project aims to empower interested parties and key societal groups to use AI technologies responsibly and to establish ethical and high-quality standards for the use of AI.

The project is led by **winnovation** (Gertraud Leimüller and Brigitte Ömer-Rieder) and is being carried out in collaboration with **leiwand.ai** (Rania Wazir), **TU Wien** (Sabine Köszegi and Karoline Bohrn) and **Austria Presse Agentur – APA** (Verena Krawarik und Sophia Marecek).

The ALAIT dossiers are available on the project website: https://science.apa.at/project/alait/

The content of this dossier reflects the current state of the art and has been carefully compiled in accordance with scientific criteria. However, it is not intended to serve as legally binding information or advice.

Imprint
Media owner and publisher:
winnovation consulting gmbh
Linke Wienzeile 124/5
1060 Vienna

This dossier is licensed under the Creative Commons CC BY-NC-ND 4.0 licence
https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de
(Attribution-NonCommercial-NoDerivatives 4.0 International)

Published in 2026, funded by the Austrian Federal Ministry for Innovation, Mobility and Infrastructure.

Federal Ministry Innovation, Mobility and Infrastructure Republic of Austria

winnovation

leiwand.ai

TU WIEN

APA

Acknowlegements:

We would like to thank the following expert for her helpful feedback on earlier versions of this dossier: Laura Wiesböck.