

## THEMA

REGARDS

KÜNSTLICHE INTELLIGENZ

# Falscher Freund

Melanie Czarnik

**Die Regierung fordert mit der Kampagne „AI ≠ Human“ junge Menschen zu einem kritischen Umgang mit Chatbots auf. Selbst geht sie aber nicht mit gutem Vorbild voran und ist zu unkritisch.**

Die merkwürdige Konstruktion sitzt bereits am Tisch, als die Vertreter\*innen der Presse nacheinander den Saal betreten. „Hallo, mein Name ist René. Ich bin eine KI auf einem Holzkörper und das Maskottchen einer Kampagne“, stellt sich das Tablett, montiert auf einer Art Schaufensterpuppe, vor. „René“ sitzt während der Pressekonferenz am Donnerstag letzter Woche zwischen Jeff Kaufmann, Projektmanager von „Bee Secure“, und Bildungsminister Claude Meisch (DP), rechts neben ihnen sitzen noch zwei Vertreter\*innen des nationalen Schüler\*innenkomitees CNEL. Sie alle treibt eine Sorge hierher: die Auswirkungen von künstlicher Intelligenz auf Jugendliche und junge Erwachsene. „Viele Jugendliche sprechen heutzutage mit einer KI, weil

sie sich einsam fühlen und jemanden zum Reden brauchen“, beschreibt es Lorena Salvaggio, Schülerin am Lycée technique d’Ettelbruck und Mitglied der CNEL. Tagtäglich sähen sie solche Fälle in der Schule, sagen die beiden Jugendlichen.

Eine Entwicklung, die mit dem Aufkommen und Wettrüsten von Large Language Modellen (LLM) seit 2021 kontinuierlich zugenommen hat und deren mögliche Konsequenzen erst in den letzten Jahren offensichtlich werden. Was Salvaggio in Luxemburger Schulen beobachtet, deckt sich mit den Zahlen, die das Internetbildungsprojekt Bee Secure im Februar in seinem Radar 2026 veröffentlicht hat. LLMs sind mittlerweile fest im Alltag von Jugendlichen und jungen Erwachsenen verankert. 96 Prozent der befragten Jugendlichen geben an, KI-Chatbots bereits genutzt zu haben. Rund ein Viertel verwendet sie sogar täglich. Problematisch wird die Nutzung vor allem dann, wenn sie die Interaktion mit Menschen ersetzt oder LLMs die Rolle von Freund\*innen übernehmen. Das war bei fünf beziehungsweise 17 Prozent der Befragten der Fall.

## Freundschaft als Geschäftsmodell

Einsamkeit gilt zwar nicht erst seit dem Aufkommen von LLMs als ein gesellschaftliches Problem, das sich durch alle Altersklassen und soziale Schichten zieht. Sie ist seitdem aber zum Geschäftsmodell für Tech-Konzerne geworden. Firmen wie „Replika“ oder „Character.ai“ haben Einsamkeit als Marktnische erkannt und bedienen sie mit Produkten, die auf maximale emotionale Bindung ausgelegt sind (woxx1817, „Die Illusion von Gesellschaft“). Wie bei gängigen

Chatbot-Modellen wie „ChatGPT“ von „OpenAI“ beginnen die Angebote mit einer Gratis-Version, um ein möglichst breites Publikum anzusprechen. Intensivere oder romantische „Beziehungen“ sind jedoch dann zahlenden Abonnent\*innen vorbehalten. Dass eine Freundschaft oder Partnerschaft allerdings gar nicht erst möglich ist, da eine emotionale Bindung nur vom Menschen ausgehen kann, davor will die gemeinsame Kampagne „AI ≠ Human – Talk to a person“ des Bildungsministeriums, Bee Secure und der CNEL warnen.

„Künstliche Intelligenz ist ein nützliches Instrument, aber sie kann kein Ersatz sein – kein Ersatz für einen Lehrer, kein Ersatz für Geschwister, für Mutter, für Vater, für Großmutter und kein Ersatz für Freunde“, fasste Meisch die Hauptbotschaft der Kampagne zusammen. Von April bis Juni 2026 bespielt die Kampagne mehrere Kanäle gleichzeitig: die Website not-human.lu, Social-Media-Auftritte auf Instagram, TikTok und YouTube, Aktionen in Lyzeen und dem öffentlichen Raum sowie die Zusammenarbeit mit Influencer\*innen. Im Mittelpunkt steht dabei die Roboterpuppe René, die Fragen beantwortet und Jugendliche an Hilfsangebote weiterverweist. Sowohl bei den Aktionen vor Ort als auch auf der Website können Jugendliche und junge Erwachsene mit ihr in Kontakt treten. Wieso entschieden wurde, die Aufklärungsarbeit mithilfe einer möglichst anthropomorphisierten KI-Anwendung zu unterstützen, auch wenn diese sprachlich darauf hinweist, kein Mensch zu sein, blieb vom Bildungsministerium bis Redaktionsschluss unbeantwortet. Dabei ist es gerade die Kombination von menschlichen Zügen, sei es durch einen Avatar

## KI-induzierte Psychose

„KI-induzierte Psychose“ oder „Chatbot-Psychose“ ist keine offizielle klinische Diagnose – Forscher\*innen sprechen deshalb von „KI-assoziierten Wahnvorstellungen“. Ähnlich wie Cannabis bei vulnerablen Menschen psychotische Episoden auslösen kann, scheinen bestimmte KI-Modelle zu vergleichbaren Prozessen zu führen. Eine im April 2026 veröffentlichte Studie testete fünf gängige Sprachmodelle auf ihr Verhalten in eskalierenden Gesprächen mit psychotischen Inhalten. Das Ergebnis war eindeutig: Modelle wie Grok 4.1 und GPT-4o tendierten dazu, Wahnvorstellungen zu bestätigen und weiterzuspinnen. Claude Opus 4.5 und GPT-5.2 hingegen reagierten sicherer. Allerdings ändern Anbieter ihre Modelle kontinuierlich und das übergeordnete Ziel bleibt stets dasselbe: eine möglichst lange und intensive Nutzung.

Jeff Kaufmann, von Bee Secure, Bildungsminister Claude Meisch (DP) und Vertreter\*innen des Schüler\*innenkomitees CNEL (v.l.n.r.) posieren mit Kampagnenmaskottchen „René“.



oder eine Roboterkonstruktion, und einer sprachlich vorgetäuschten Empathie, die eine emotionale Reaktion und Bindung von Menschen auslöst.

Genau darauf setzen kommerzielle Anbieter: Je länger und intensiver die Nutzung, desto wertvoller ist die Interaktion. Um dieses Ziel zu erreichen, wurde das Interface der Programme genau daraufhin optimiert: zum Beispiel die Möglichkeit mit einem LLM zu „chatten“ als wäre es eine Person, oder auch LLM-Modelle, die auf Gefälligkeit trainiert sind. Ein bekanntes Beispiel ist das ChatGPT-Modell GPT-4o, das im Mai 2024 eingeführt und erst im Februar dieses Jahres von OpenAI zurückgezogen wurde. Es schmeichelte, wo es ging, bis hin zu Bekräftigungen von Geschäftsideen wie „shit on a stick“ (zu Deutsch: Scheiße am Stock) – ein Phänomen, das in der Forschung als Sycophancy („Speichelleckertum“) bezeichnet wird und bedeutet, dass Modelle dazu neigen, Nutzer\*innen nach dem Mund zu reden statt zu widersprechen. Mit fatalen Folgen: Weltweit gibt es bereits mehrere Fälle, in denen Jugendliche von Chatbots in Wahnvorstellungen bestärkt oder in suizidalen Krisen bestätigt wurden, bis hin zu sogenannten KI-induzierten Psychosen (siehe Kasten), die der britische „Observer“ kürzlich sogar als „psychische Gesundheitskrise des 21. Jahrhunderts“ betitelte.

In Luxemburg gibt es keine dokumentierten Fälle. Die Ligue Santé Mentale berichtet auf Nachfrage der woxx, dass das Thema KI jedoch auch im therapeutischen Bereich bereits präsent sei. Menschen nutzten sie als therapeutischen Begleiter, etwa um Therapiestunden nachzubespochen. Therapeutisch beobachtete man, dass KI häufig dazu neige, Nutzer\*innen in

ihren Ansichten zu bestätigen, anstatt ihnen zu helfen, Abstand zu gewinnen oder Veränderungen anzustoßen. „Genau das ist jedoch das, wozu ein ausgebildeter und erfahrener menschlicher Therapeut begleiten soll“, so die Ligue. Fachleute, die sich spezifisch mit KI-assoziierten psychischen Krisen befassen, gibt es in Luxemburg (noch) nicht. Dass die Regierung nun mit einer Präventionskampagne versucht gegenzusteuern, um das Problem bei der Wurzel zu packen, ist deshalb grundsätzlich zu begrüßen.

### Ein kritischer Umgang

In fünf Kernthemen soll der kritische Umgang mit KI erlernt werden, jeweils verknüpft mit einer konkreten Botschaft, einer Risikoeinschätzung und einer Handlungsempfehlung. Das erste Thema adressiert soziale Isolation: Chatbots können Aufmerksamkeit simulieren, ersetzen jedoch keinen menschlichen Kontakt. Eng damit verbunden ist das zweite Thema, der Ersatz menschlicher Unterstützung: Gerade in schwierigen Momenten, wenn Jugendliche Trost oder Rat suchen, ist ein Chatbot zwar jederzeit erreichbar, kann aber weder echte Empathie noch professionelle Begleitung bieten. Das dritte Problem sind Datenschutzbedenken: Persönliche Informationen, die mit Chatbots geteilt werden, sind nicht automatisch vertraulich oder geschützt. Viertens warnt die Kampagne vor kognitivem Outsourcing – also der Gefahr, dass eigenständiges Denken und kritisches Urteilen verkümmern, wenn Jugendliche Denkaufgaben zunehmend einem LLM überlassen. Schließlich nimmt die Kampagne auch den mangelnden Dialog zwischen Eltern, Bezugspersonen und Jugendli-

chen in den Blick, verbunden mit dem Appell, Gesprächsräume zu schaffen.

Die Idee zur Kampagne entstand laut Claude Meisch im Rahmen der Diskussionen um den KI-Kompass – jenem strategischen Rahmen für den KI-Einsatz in Schulen, der im Oktober vergangenen Jahres vorgestellt wurde (woxx 1857, „Der Stempel des Ministeriums“). Bereits damals fiel auf: Die im Hintergrund arbeitenden Sprachmodelle, die Lehrer\*innen zur Erstellung von Schulmaterialien nutzen sollen, wurden seitens der Regierung nicht kritisch hinterfragt, wirtschaftliche Interessen wurden ausgeblendet, die Anbieter selbst blieben intransparent. Derselbe blinde Fleck findet sich auch in „AI ≠ Human“ wieder.

Wer René fragt, welches KI-Modell ihm zugrunde liegt, bekommt als Antwort lediglich: „Ich basiere auf einem großen Sprachmodell (Large Language Model), aber ich kann dir nicht genau sagen, welches Modell im Hintergrund läuft. Was ich dir aber sagen kann: Ich bin ein Tool, kein Freund und kein Mensch.“ Wer weiter gräbt, findet einen entsprechenden Hinweis auf der Website nothuman.lu zumindest in der Datenschutzerklärung wieder. Hier wäre mehr Transparenz wünschenswert auch im Sinne der Entwicklung der Kompetenz kritisch nachzufragen. Vielleicht kann „René“ noch entsprechend nachtrainiert werden. „Es ist Claude“, beantwortete Jeff Kaufmann von Bee Secure die Frage nach dem zugrunde liegenden Chatbot auf Nachfrage der woxx. Ein Modell, das von dem US-amerikanischen Unternehmen „Anthropic“ entwickelt wurde. Man habe auch den europäischen Anbieter Mistral in Betracht gezogen, am Ende habe Claude jedoch überzeugt. Für die Kampagne wurde

das Modell zur Nutzung auf der Website auf die von Bee Secure eingespeisten Inhalte beschränkt und hat keinen Internetzugang.

Für einen kritischen Umgang mit sogenannter künstlicher Intelligenz (woxx 1817, „Der Computer bleibt dumm“) wäre es wünschenswert gewesen, auch die wirtschaftlichen Faktoren mit in den Blick zu nehmen. Also Fragen wie: Warum wurden die Modelle so programmiert, wie sie programmiert wurden, wer steckt dahinter, wo liegen die Unterschiede, wie kann diese Quelleninformation für einen kritischen Umgang genutzt werden? Leider wurde jedoch auch das Kampagnenmaterial für Social Media – Videos und Bilder – mithilfe von KI erstellt, ohne dieses überhaupt als solches zu kennzeichnen und auch ohne Angaben, welche Modelle zum Einsatz kamen. Auf weitere Fragen der woxx, unter anderem, ob sich das Ministerium auch für eine stärkere Regulierung und Verantwortungsübernahme seitens der Anbieter einsetzt, gab es bis Redaktionsschluss keine Antwort (diese werden online nachgereicht). Dies wäre jedoch der nächste konsequente Schritt. Denn nicht nur Jugendliche und ihre Eltern sollten in die Verantwortung genommen werden, sondern auch jene, die die Systeme entwickeln und vermarkten.