

Die nächste KI-Waffe – ein Mythos

Während es [im letzten Artikel](#) noch um ein überschaubares Schadenspotenzial der KI ging, nämlich als „Tatwaffe“ für Beleidigungen benutzt zu werden, hat sich die Lage kaum 2 Wochen später verschärft. Grund hierfür sind die Verlautbarungen rund um das bisher nicht veröffentlichte KI-Modell von Anthropic mit dem spektakulären Namen *Mythos*. Natürlich kann es auch ein Marketing-Coup sein, sich „aufgrund seiner enormen Fähigkeiten“ gegen eine öffentliche Freigabe zu entscheiden. Als Hauptgrund für die Zurückhaltung wird die Fähigkeit des Modells genannt, hochkomplexe Sicherheitslücken in Betriebssystemen und Browsern fast autonom zu finden und auszunutzen zu können. So erhält es in den Medien die Zuschreibung einer „gefährlichen Waffe aus dem KI-Labor“. Mythos wird zur Zeit einem kleinen Kreis von Partnerorganisationen (darunter Apple, Microsoft und Google) im Rahmen des „Project Glasswing“ zur Verfügung gestellt.

Das klingt soweit noch relativ unverdächtig und Anthropic genießt in der KI-Szene derzeit gegenüber dem Konkurrenten OpenAI durchaus einen Vertrauensvorsprung. Obwohl es ein amerikanisches Unternehmen mit Sitz in San Francisco ist, bricht es bewusst mit dem typischen „Silicon Valley Tech-Slang“. Schon der Firmenname ist aus dem altgriechischen Anthropos abgeleitet (ἄνθρωπος = Mensch) und suggeriert, dass der Mensch im Zentrum der technologischen Entwicklung stehen soll. Das aktuelle, bisher leistungsfähigste Modell heißt Opus 4.6, lateinisch für das Werk oder das „Meisterwerk“. Die Firmengründer scheinen Wert auf eine humanistische Verankerung zu legen. Und nun wird mit Mythos (μῦθος: das Wort, die Erzählung, das Gespräch) die bedeutungsvolle Namensgebung fortgesetzt. In Mythen wurde das komplexe Verhältnis der (unsterblichen) Götter zu den (sterblichen) Menschen beschrieben. Die Götter griffen aktiv in das Leben der Menschen ein – mal als Helfer, mal als Widersacher. Eine der Lehren ist die Warnung vor Hybris (Selbstüberschätzung). Wenn Sterbliche versuchten, sich mit den Unsterblichen zu messen oder deren Autorität missachteten, wurden sie hart bestraft.

Die Namensgebung lässt etwas Legendäres mitschwingen. Es ist möglicherweise eine bewusste Entscheidung, die eigenen KI-Modelle nicht als reine Technologie, sondern als Teil der menschlichen Kulturgeschichte zu positionieren. Doch die Beschränkung des Zugriffs auf Mythos an einen elitären Kreis von Tech-Riesen sollte auch Bedenken auslösen. Eine mächtige Waffe in der Hand weniger Konzerne – völlig losgelöst von Politik und demokratischer Gesellschaftsordnung. Mag es heute zum Guten eingesetzt werden, um Lücken in der Internetsicherheit zu schließen, aber was wird es morgen in der Hand von weniger menschenfreundlichen Zeitgenossen ausrichten?

„Don’t be evil“ war das inoffizielle Firmenmotto bei Google, um sich von Microsoft abzuheben. Aber kürzlich wurde Google (Alphabet) gemeinsam mit Facebook (Meta) für schuldig befunden, Kinder und Jugendliche absichtlich süchtig nach ihren Produkten gemacht zu haben. Microsoft wollte sich in seiner Gründungszeit mit Pulli und langen Haaren vom EDV-Platzhirsch IBM mit konservativer Anzug-und-Krawatte-Kultur abheben. Mehrfach wurde Microsoft später wegen Missbrauchs seiner marktbeherrschenden Stellung verurteilt. Jeder will cooler oder krasser sein als der Konkurrent – was ist Marketing Talk und wo lauert die Hybris?

Jaron Lanier, einer der schillerndsten Kritiker der Tech-Industrie im Silicon Valley, erhielt 2014 den Friedenspreis des Deutschen Buchhandels. Eine zentrale Botschaft seiner Dankesrede trat schon im

Titel hervor: „Der High-Tech-Frieden braucht eine neue Art von Humanismus“: „Der neue Humanismus behauptet, es sei richtig, zu glauben, dass Menschen etwas Besonderes sind, nämlich dass Menschen mehr sind als Maschinen und Algorithmen. [...] ganz gleich, was meine Freunde mit ihren digitalen Überlegenheitsfantasien und Unsterblichkeitslaboratorien denken.“

Lanier sieht die Hybris vor allem in der Tendenz, Technologie zu vergöttlichen und sie als eine Art eigenständiges, übermenschliches Wesen darzustellen. Machtgier und Arroganz führen dazu, dass die Technologie nicht mehr dem Menschen dient, sondern der Stärkung der eigenen Vormachtstellung.

Das ist bereits 12 Jahre her, in denen die Digitalisierung mit großen Schritten vorangegangen ist. Man sollte aufmerksam bleiben.