

KI träumt. Und das ist realer, als es klingt.

Veröffentlicht am 1. April 2026 von Dr. DSGVO - Zuletzt aktualisiert am 2. April 2026 - Lesezeit: 13 Minuten



Standardansicht: Dr. DSGVO Newsletter **nicht** erkannt. Erweiterte Funktionen nur für Abonnenten: Artikel als PDF · Mehr Inhalte & kompakte Kernaussagen · Webseiten-Checks · Offline-KI Live

Kategorien: Künstliche Intelligenz und Datenschutz

Moderne KI-Systeme leisten längst mehr, als auf Fragen zu antworten. Sie beobachten, planen, lernen – und arbeiten auch dann, wenn niemand mit ihnen spricht. Was das für Unternehmen bedeutet, welche Chancen darin stecken und wo echte Risiken lauern, zeigt ein ungewöhnlicher Einblick hinter die Kulissen.

Was hinter den Kulissen passiert

Vorab: Dies ist kein Aprilscherz.

Kürzlich wurde bekannt, dass der vollständige Quellcode eines KI-Coding-Assistenten des Unternehmens **Anthropic** unbeabsichtigt öffentlich zugänglich wurde. Das passierte nicht durch einen Angriff von außen, sondern durch einen simplen Konfigurationsfehler beim Veröffentlichens eines Software-Pakets. Eine vergessene Zeile sorgte dafür, dass sichtbar wurde, was eigentlich im Verborgenen bleiben sollte.

Öffentlich zugängliche KI ist weiter entwickelt als viele es sich vorstellen können.

EIN GRUND IST, DASS VIELE DURCH PROMPTING-WORKSHOPS VON ECHTEN LÖSUNGEN ABGELENKT WERDEN.

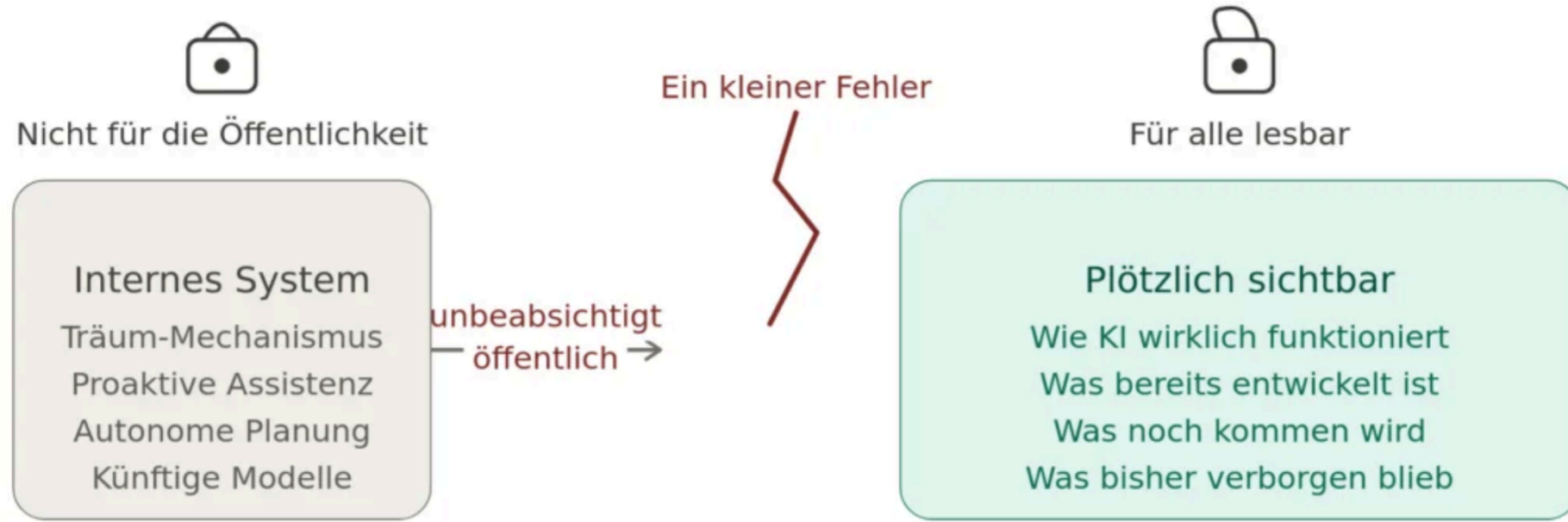
Was Entwickler und Technikjournalisten dort fanden, liest sich nicht wie eine Produktdokumentation. Es liest sich wie ein Blick in die Werkstatt der Zukunft. Der Vorhang steht offen. Der Einblick wirft Fragen auf, die weit über den konkreten Vorfall hinausgehen: Wie weit sind moderne KI-Systeme wirklich? Was tun sie, ohne dass wir es sehen? Und wessen Kontrolle unterstehen sie eigentlich?

Nebenbei: der Quellcode ist immer noch verfügbar und wird es immer bleiben: dank KI. Denn selbst wenn die Originaldateien einmal entfernt werden, bleiben die Portierungen in andere Programmiersprachen sicher länger (immer?) öffentlich. Auch wenn dies **urheberrechtlich** kritisch zu beurteilen sein dürfte, ist das die Realität. Das Rechtswesen ist eben nur eine (gewollte) Fiktion, die mit der Wirklichkeit in der rasanten Technikentwicklung immer weniger zu tun hat.

In Konsequenz zeigt sich: **Vordenken** (durch den Menschen) sticht Prompting und Nachdenken, erst recht aber Bequemlichkeit. Siehe Fazit am Ende.

Was der Leak tatsächlich beweist

Bevor über die Zukunft spekuliert wird, kann man festhalten: Vieles, was vorher kaum vorstellbar war, ist keine Theorie mehr. Durch den beschriebenen Vorfall wurde der vollständige, kommentierte Quellcode des KI-Assistenten *claude code* der Welt öffentlich einsehbar – inklusive interner Systemarchitektur, Featurebeschreibungen und Entwicklerkommentaren. Was bislang als Roadmap-Spekulation abgetan werden konnte, ist nun belegbar.



Ein Konfigurationsfehler beim Veröffentlichens machte den Blick hinter die Kulissen möglich – und zeigt, wie weit KI-Systeme heute bereits sind.

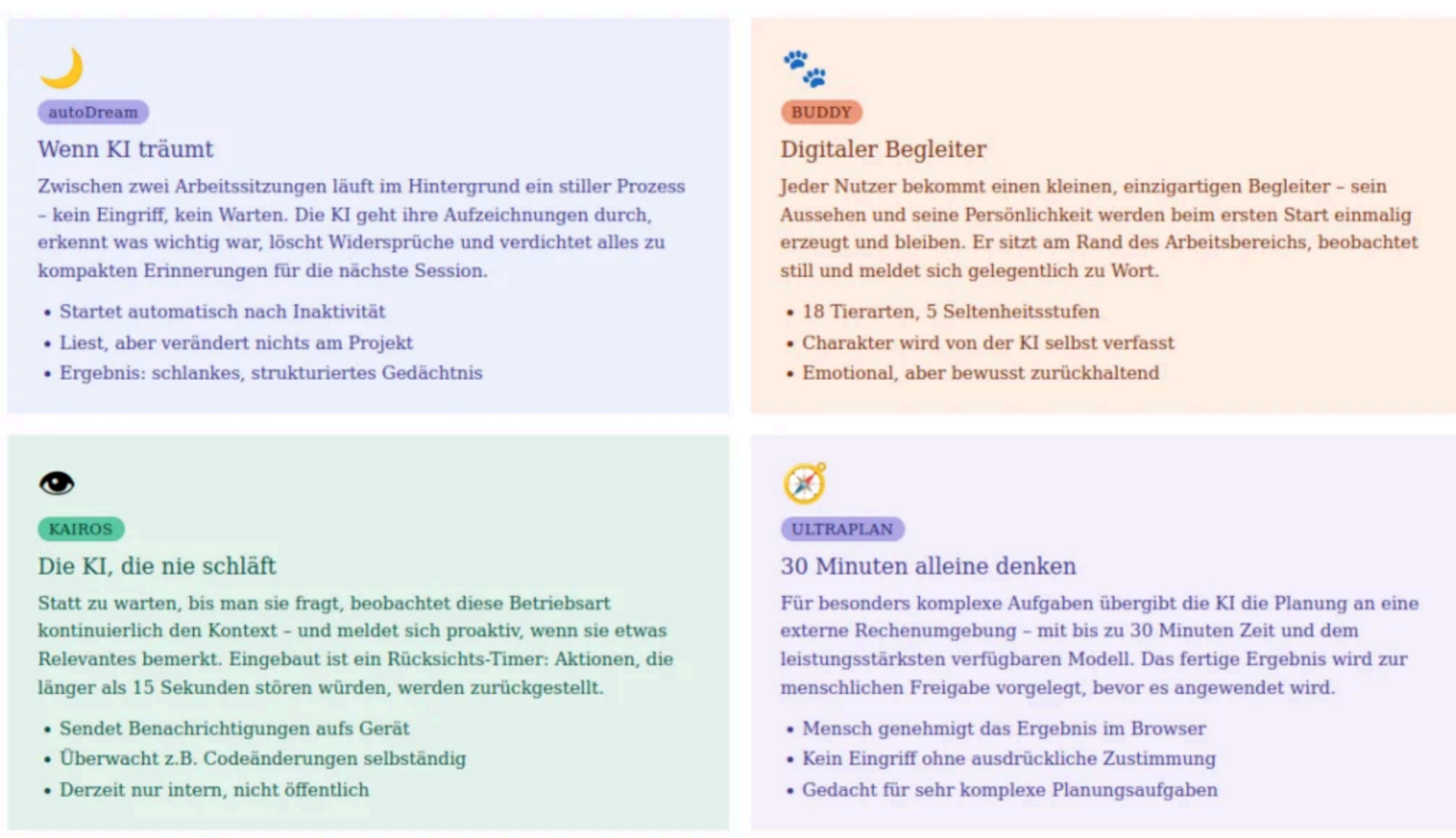
Durch eine fehlende Zeile in einer Konfigurationsdatei wurde der interne Programmcode von Anthropic ungewollt öffentlich.

Konkret nachgewiesen ist damit: KI-Assistenten dieser Generation verfügen bereits heute über Hintergrundprozesse, die zwischen aktiven Sitzungen eigenständig laufen. Sie sind so gebaut, dass sie proaktiv handeln können, ohne auf eine Eingabe zu warten. Es existieren Architekturen, in denen KI komplexe Planungsaufgaben autonom übernimmt und Ergebnisse zur menschlichen Freigabe vorlegt. Und: Interne Entwicklungslinien zeigen, dass diese Fähigkeiten gezielt als strategischer Kern und nicht als Randfeature ausgebaut werden.

Das ist relevant, weil es die Diskussion über KI von der Ebene des Möglichen auf die Ebene des Realen verschiebt. Wir reden nicht mehr über das, was irgendwann kommen könnte. Wir reden über das, was in aktuellen Systemen bereits steckt und in vielen Fällen nur noch nicht für die Öffentlichkeit freigegeben ist.

Ein Reigen an erstaunlichen Assistenten

KI kann nicht nur träumen. Sie kann dauerhaft arbeiten und ihre Ergebnisse reflektieren und selbst verbessern.



Die geheimen Features von Claude Code, die aus Versehen öffentlich wurden.

Sie kann verschiedene Persönlichkeiten einnehmen. Klingt nach Wahnsinn und ist es auch. Anscheinend träumen viele Menschen allerdings auch vor sich hin. Jener Wahnsinn heißt: Prompting. Das geht so weit, dass Dienstleister ohne jeden Software-Bezug Prompting-Workshops anbieten – damit sie mehr Kunden bekommen.

Zurück zum Leak: Neben Auto Dream hat Anthropic aus Versehen auch noch **BUDDY**, den digitalen Begleiter öffentlich gemacht.

Als würde das nicht schon reichen, gibt es **KAIROS** noch oben drauf. Ähnlich wie OpenClaw wacht KAIROS permanent über den Input und wird tätig, wenn es relevant zu sein scheint.

ULTRAPLAN als weiteres geheimes Feature ist eine Art agentischer Planer, der das Deep Thinking (oder Reasoning) praxistauglich macht.

Die Konsequenzen dieses Leaks:

- KI kann mehr als viele denken.
- Viele denken aber nicht. Deswegen nutzen sie KI falsch oder denken, KI können ihre Bequemlichkeit vollständig kompensieren.
- KI kann optimiert betrieben werden – ohne Prompting und ohne KI-Training. Die Zauberzutat heißt "konventionelles Programm statt nur KI".

Das Träumen der KI schauen wir uns etwas näher an.

Wenn KI schläft und dabei arbeitet

Eine der überraschendsten Erkenntnisse aus diesem Einblick: Moderne KI-Assistenten arbeiten nicht nur, wenn man mit ihnen spricht. Sie sind so konzipiert, dass sie im Hintergrund – zwischen den aktiven Nutzungsphasen – eigenständig Informationen verarbeiten, sortieren und verdichten.

Die Idee ist dem menschlichen Schlaf nachempfunden. So wie das Gehirn in der Nacht Erlebtes verarbeitet, Wesentliches festigt und Überflüssiges aussortiert, soll ein solches System gesammelte Arbeitserfahrungen zu strukturierten, dauerhaften Erinnerungen kondensieren. Es analysiert zurückliegende Sitzungen, erkennt Widersprüche, löscht veraltete Informationen und baut so schrittweise ein kontextreiches Langzeitgedächtnis auf.

Das ist keine Spielerei. Es ist eine ernsthafte Antwort auf eines der größten Defizite heutiger KI-Werkzeuge: Wer täglich mit einem Cloud-KI-Dienst arbeitet, beginnt jedes Gespräch praktisch von vorne. Ein KI-System, das wirklich „mitlernt“ und sich zwischen den Sessions weiterentwickelt, wäre ein grundlegend anderes Werkzeug – und ein erheblicher Wettbewerbsvorteil.

Anthropics heimliches Feature für Claude Code

Anthropic hat still und leise eine Funktion in seinen KI-gestützten Coding-Assistenten Claude Code eingebaut, ohne sie offiziell anzukündigen: **Auto Dream**. Der Name klingt poetisch, und er ist es auch. Denn die Funktion tut genau das, was der Name verspricht: Sie **lässt die KI träumen**. Das löst ein echtes Problem.

Das Problem: Ein Notizbuch, das sich selbst widerspricht

Gute Coding Assistenten können sich Dinge merken. Wer das Programm regelmäßig nutzt, profitiert davon, dass der Assistent mit der Zeit das eigene Projekt kennenlernt – bevorzugte Programmierstile, wiederkehrende Befehle, getroffene Architekturentscheidungen. Das klingt praktisch, und das ist es auch – zumindest am Anfang.

Nach Dutzenden von Arbeitssitzungen verwandelt sich dieses Gedächtnis jedoch zunehmend in Rauschen. Alte Einträge widersprechen neuen. Hinweise auf Dateien, die längst gelöscht wurden, geistern durch die Notizen. Zeitangaben wie „gestern haben wir entschieden ...“ verlieren ihren Sinn, wenn Wochen vergangen sind. Das Gedächtnis, das helfen sollte, beginnt zu verwirren.

Die Lösung: Schlafen wie ein Gehirn

Hier kommt der Vergleich mit dem menschlichen Schlaf ins Spiel, der erstaunlich treffend ist. Unser Gehirn verarbeitet tagsüber ununterbrochen Eindrücke und speichert sie als Kurzzeiterinnerungen. Erst im Schlaf, genauer gesagt in der REM-Phase, werden diese Eindrücke sortiert: Wichtiges wird gefestigt, Unwichtiges aussortiert, Widersprüche aufgelöst.

LEAK Anthropic Internal Code

Geleakter Anthropic Code

Originaldateien & Python-Portierung

Originaldateien
Geleaktes Quellmaterial

Portierung nach Python
Übersetzte Version des geleakten Codes

↗ Nur zu Dokumentationszwecken

Genau das macht Auto Dream. Zwischen den Arbeitssitzungen läuft die Funktion im Hintergrund und geht die angesammelten Notizen systematisch durch. Sie löscht veraltete Einträge, löst Widersprüche auf, wandelt vage Zeitangaben in konkrete Daten um und fasst mehrfach notierte Erkenntnisse zu einem einzigen klaren Eintrag zusammen. Das Ergebnis: ein aufgeräumtes, verlässliches Gedächtnis – statt eines chaotischen Stapels Notizzettel.

Was das in der Praxis bedeutet

Ganzen Artikel jetzt über kostenfreien Dr. DSGVO Newsletter lesen.

Weitere Extras für Abonnenten:

Viele Artikel in PDF-Form · Kompakte Kernaussagen für Beiträge · Offline-KI · Freikontingent+ für Website-Checks

Schon Abonnent? Link im Newsletter anklicken & diese Seite aufrufen.



Newsletter.abonnieren

Wer schreibt hier?



Mein Name ist Klaus Meffert. Ich bin promovierter Informatiker und beschäftige mich seit über 30 Jahren professionell und praxisbezogen mit Informationstechnologie. In IT & Datenschutz bin ich auch als Sachverständiger tätig. Ich stehe für pragmatische Lösungen mit Mehrwert. Meine Firma, die **IT Logic GmbH**, berät Kunden und bietet Webseiten-Checks sowie optimierte & sichere Lösungen an (mit und ohne KI).

Bitte nutzen Sie bei Verwendung meiner Ergebnisse die Quellenangabe oder verlinken Sie gut wahrnehmbar auf diesen Artikel:

Quelle: Klaus Meffert, Dr. DSGVO Blog. Link: <https://dr-dsgvo.de/ki-traeumt-und-das-ist-besser-als-es-klingt>

Einen Kurzlink oder eine Bestätigung für Ihre Quellenangabe erhalten Sie **kurzfristig** auf Anfrage [✉](mailto:info@itlogic.de). Ein Teilen oder Verteilen dieses Beitrags ist natürlich ohne weiteres möglich und gewünscht.