Analyse: KI-Anwendung Claude von Anthropic Stand: Juni 2026

I. Zu bewertendes Verfahren / Tool und Zweckbestimmung

Das im Jahr 2021 gegründete Unternehmen Anthropic mit Firmensitz in San Francisco, USA brachte im März 2023 erstmals den Allzweck-KI-Assistenten Claude auf den Markt.

Claude Chat dient neben der Wissensvermittlung (Fragen beantworten, Themen erklären, Texte schreiben, zusammenfassen und übersetzen) auch der strukturierten Datenauswertung und Mustererkennung, hat von sich aus keinen Zugriff auf lokale Dateien oder externe Systeme.

Claude Code ist ein Terminal-basiertes Tool für Entwickler*innen. Es kann Code schreiben, Repositorien verwalten und Tests ausführen, setzt aber technisches Know-how voraus.

Claude Cowork kann eigenständig Aufgaben auf dem Desktop erledigen. Er kann Dateien lesen, bearbeiten und erstellen – alles über eine Chat-Oberfläche, ohne Programmierkenntnisse.

Die Gemeinsamkeit: Alle drei nutzen dasselbe KI-Modell (Claude). Der Unterschied liegt in den Fähigkeiten und der Zielgruppe. Diese Datenschutzfolgeabschätzung beschäftigt sich vorrangig mit dem Einsatz des Allzweck KI – Assistenten Claude Chat. Claude Code und Claude Cowork sind z.T. erst im Mai 2026 auf den allgemein zugänglichen Markt gekommen. Dass die Erlaubnis zum Zugriff der KI auf ausgewählte Dateien des Desktops und selbstständiger Verarbeitung der Inhalte (Cowork) natürlich weitere Datenschutzrisiken birgt sollte klar sein. Hierauf wird zum Ende der Risikoanalyse kurz eingegangen.

Die Hauptplattform von Anthropic, die auch die Standardversion von Claude (claude.ai oder über ein API) nutzt, ist bei Amazon Web Services (AWS) in den USA gehostet. Nur die Authentifizierung der Nutzer*innen erfolgt über AWS, die Datenverarbeitung selbst erfolgt dann auf der US-Infrastruktur von Anthropic, sofern kein Bedrock- oder Vertex-AI-Routing erfolgt, bei dem eine Datenverarbeitung in der EU vereinbart werden kann.

Es stehen auch spezielle Konfigurationen wie z.B. Zero-Data-Retention (ZDR) optional zur Verfügung. Hierbei werden Eingaben und Ausgaben nach der API-Antwort nicht

Gefördert durch das
Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

gespeichert. ZDR ist nur für berechtigte API-Enterprise-Kunden mit gesonderter ZDR-Vereinbarung verfügbar.

1. Betroffenengruppen deren personenbezogene Daten verarbeitet werden

Claude verarbeitet alle Daten, die eingegeben oder hochgeladen werden. Insofern können alle Menschen (Teilnehmende einer Bildungsmaßnahme, Mitarbeitende, Vertragspartner*innen, unbeteiligte Dritte etc.) betroffen sein, sobald jemand personenbezogene Daten eingibt.

2. Art der Daten

a) Personenbezogene Daten, die Nutzer*innen oder die Einrichtung direkt zur Verfügung stellt

Identitäts- und Kontaktdaten: Anthropic sammelt Kennungen, also Name, E-Mail-Adresse und Telefonnummer bei Anmeldung für ein Anthropic-Konto, oder bei Abfrage von Informationen zu den Diensten.

Zahlungsinformationen: Bei Erwerb des Zugangs zu Produkten oder Dienstleistungen von Anthropic

Ein- und Ausgänge: Mit den Diensten kann in verschiedenen Formaten interagiert werden, z. B. über Chat-, Code- oder Agenten-Sitzungen („Inputs“). Darauf basierend generieren die Dienste Antworten und Aktionen („Outputs“). Werden externe Inhalte oder personenbezogene Daten in die Inputs aufgenommen, werden diese erfasst und können in den Outputs wiedergegeben werden. Dies gilt auch für Anwendungen von Drittanbietern, die integriert werden.
Claude Cowork verarbeitet eigenständig sämtliche Daten, die in dem zur Verfügung gestellten Ordner enthalten sind.

Feedback: Bei Feedback z. B. mit dem Daumen nach oben/ Daumen nach unten – speichert Anthropic die gesamte verwandte Konversation.

Kommunikationsinformationen: Bei der Kommunikation mit Anthropic, auch über den Chatbot oder die Hilfe-Website, werden der Name, die Kontaktinformationen und der Inhalt der Nachricht gespeichert.

b) Bei Nutzung von Claude werden zudem automatisch bestimmte technische Daten übermittelt. Hierzu gehören:

Geräte- und Verbindungsinformationen wie der Gerätetyp, Betriebssysteminformationen, Browserinformationen und Webseiten-Referer, Mobilfunknetz, Verbindungsinformationen, Mobilfunkbetreiber oder Internetdienstanbieter (ISP), Zeitzoneneinstellung, IP-Adresse (einschließlich Informationen über den Standort des Geräts, der von Ihrer IP-Adresse abgeleitet wurde), Kennungen (einschließlich Geräte- oder Werbekennungen, probabilistische Kennungen und andere eindeutige persönliche oder Online-Kennungen) und Gerätestandort.

Nutzungsinformationen wie z. B. die Daten und Zeiten des Zugriffs, den Browserverlauf, die Suche, Informationen über aufgerufene Links oder Seiten und andere Informationen darüber, wie die Dienste genutzt werden.
Protokoll- und Fehlerbehebungsinformationen wie Log-Dateien, Informationen über Fehler, die Zeit, in der der Fehler aufgetreten ist, etc. und alle Kommunikationen oder Inhalte, die zum Zeitpunkt des Auftretens des Fehlers bereitgestellt wurden.
Cookies & Ähnliche Technologien

c) Trainingsdaten: Auch bei der kostenlosen Variante von Claude (Claude free, Pro, Max) werden laut Datenschutzerklärung von Anthropic die Inputs der Nutzer*innen nur mit ausdrücklicher Einwilligung des Nutzers zum KI-Training verwendet.

Anthropic bezieht personenbezogene Daten von Fremdquellen, um die Modelle zu trainieren. Zu den Quellen gehören:

- Öffentlich zugängliche Informationen über das Internet
- Datensätze, die Anthropic durch kommerzielle Vereinbarungen mit Drittunternehmen erhält
- Daten, die Benutzer*innen oder Crowd-Mitarbeiter*innen bereitstellen, einschließlich Ein- und Ausgänge aus den verschiedenen Diensten (es sei denn, Benutzer melden sich ab)
- Feedback, das Benutzer*innen explizit über die Dienste bereitstellen

- Materialien, die für die Sicherheit oder die Überprüfung der Richtlinien gekennzeichnet sind
- Daten, die Anthropic intern generiert

3. An der Verarbeitung beteiligten Komponenten (Systeme und Dienste sowie Prozesse)

Angaben gemäß Auflistung des Trustcenters von Anthropic

Google Cloud Platform – Cloud-Infrastruktur Weltweit

Produkte: Alle Produkte

Amazon Web Services – Cloud-Infrastruktur Weltweit

Produkte: Alle Produkte

Microsoft Azure – Cloud-Infrastruktur Weltweit

Produkte: Alle Produkte

Cloudflare - Traffic-Routing (CDN) Weltweit (lokal beim Kunden)

Produkte: Alle Produkte

Stripe – Abrechnung Vereinigte Staaten

Produkte: Claude Pro/Max,

Claude Developer Platform, Claude for Work

WorkOS - Sicherheit, Single Sign-On Vereinigte Staaten

Produkte: Claude for Work, Claude Developer Platform

Intercom – Benutzersupport Vereinigte Staaten

Produkte: Alle Produkte außer Claude for Government

Nutun – Benutzersupport Südafrika

Produkte: Alle Produkte außer Claude for Government

Boldr – Benutzersupport Kanada

Produkte: Alle Produkte außer Claude for Government

Twilio - Analytik, E-Mail-/SMS-Kommunikation Vereinigte Staaten

Iterable – E-Mail-Kommunikation Vereinigte Staaten

Produkte: Alle Produkte außer Claude for Government

Functional Software, firmierend als Sentry -

Fehlerbehandlung, Benutzer-Support Vereinigte Staaten

Produkte: Alle Produkte außer Claude for Government

Sift - Erkennung von Betrug und Missbrauch Vereinigte Staaten

Produkte: Alle Produkte außer Claude for Government

Arkose Labs -

Betrugs- und Missbrauchserkennung Vereinigte Staaten

Produkte: Alle Produkte außer Claude for Government

Brave Search – Websuche Vereinigte Staaten

Produkte: Alle Produkte

ElevenLabs – Text-to-Speech Vereinigte Staaten

Produkt: Claude for Work

TurboPuffer – Websuche Vereinigte Staaten

Produkte: Alle Produkte außer Claude for Government

II. Schutzbedarfsbestimmung

Gewährleistungsziele:

Technische und organisatorische Maßnahmen sollen eine rechtskonforme Verarbeitung gewährleisten. Die Gewährleistung besteht darin, das Risiko einer nicht rechtskonformen Verarbeitung hinreichend zu mindern. Bei der vorliegenden Prüfung ist zu berücksichtigen, dass neben den klassischen Gewährleistungsziele der IT-Sicherheit auch zusätzliche Anforderungen relevant werden, die sich aus dem Einsatz einer KI ergeben.

Das Gewährleistungsziel Vertraulichkeit bezeichnet die Anforderung, dass keine unbefugte Person personenbezogene Daten zur Kenntnis nehmen oder nutzen kann. Hier ist vor allem der Schutz personenbezogener Daten in Prompts, Output und Logs betroffen.

Die Integrität ist dann gewahrt, wenn die zu verarbeitenden Daten vollständig, richtig und aktuell bleiben und nicht z.B. durch Prompt Injection verfälscht werden. Prompt Injection ist eine Manipulationstechnik, bei der Angreifer*innen durch gezielte Prompts eine KI dazu bringt, unerwünschte oder schädliche Aktionen auszuführen – etwa das Umgehen von Sicherheitsvorgaben oder das Preisgeben vertraulicher Informationen. Im Zusammenhang mit dem Einsatz einer KI setzt Integrität aber auch die Vermeidung von Diskriminierung voraus.

Das Ziel der Verfügbarkeit ist dann gewährleistet, wenn ein zuverlässiger Zugriff auf den Dienst jederzeit möglich ist.

Das Gewährleistungsziel der Intervenierbarkeit setzt die Möglichkeit voraus, dass Daten gelöscht, berichtigt, eingeschränkt oder übertragen werden können. Angesichts dessen, dass der Nutzer*innen sich mit seinem Löschbegehren ebenso wie mit dem Entzug seiner Einwilligung an die Bildungseinrichtung wenden kann, setzt das Gewährleistungsziel der Intervenierbarkeit ein Einwilligungsmanagement voraus.

Das Gewährleistungsziel der Transparenz bezeichnet die Anforderung, dass in einem unterschiedlichen Maße sowohl Betroffene, als auch die Betreiber*innen von Systemen sowie zuständige Kontrollinstanzen erkennen können, welche Daten wann und für welchen Zweck bei einer Verarbeitungstätigkeit erhoben und verarbeitet werden, welche Systeme und Prozesse dafür genutzt werden, wohin die Daten zu welchem Zweck fließen und wer die rechtliche Verantwortung für die Daten und Systeme in den verschiedenen Phasen einer Datenverarbeitung trägt.

Die erforderliche Zweckbindung ist nur dann gewährleistet, wenn Daten nur für den vereinbarten Zweck und nur in dem Umfang genutzt werden, der für die Erreichung des Zwecks erforderlich ist.

Die Anforderung der Nichtverkettung setzt voraus, das Daten zu natürlichen Personen nicht durch Verknüpfung mit Metadaten zur Identifizierung der Person führen dürfen.

Schutzbedarf/ Schadenshöhe:

Der Schutzbedarf richtet sich danach, wie intensiv eine missbräuchliche Verarbeitung personenbezogener Daten den Betroffenen in seinem wirtschaftlichen und/ oder in seinem gesellschaftlichen Belangen beeinträchtigen kann.

Gering: Bei der Verarbeitung des Namens und der dienstlichen Kommunikationsdaten der Geschäftsführung und ggfls. des Leiters der IT handelt sich um personenbezogene Daten, deren missbräuchliche Verarbeitung keine besonders schwerwiegende Beeinträchtigung des Betroffenen erwarten lässt.

Hoch/Gravierend: Ein hoher Schaden kann entstehen, wenn personenbezogene Daten verarbeitet werden, deren missbräuchliche Verarbeitung die gesellschaftliche Stellung oder die wirtschaftlichen Verhältnisse des Betroffenen erheblich beeinträchtigen kann. Darunter fallen personenbezogene Daten besonderer Kategorien (z.B. Gesundheitsdaten), personenbezogene Daten, die dem Berufsgeheimnis unterliegen, deren Verarbeitung zu einer Diskriminierung, einem Identitätsdiebstahl oder -betrug, einem finanziellen Verlust, einer Rufschädigung, der unbefugten Aufhebung der Pseudonymisierung oder anderen erheblichen wirtschaftlichen oder gesellschaftlichen Nachteilen führen kann. Ebenso fallen darunter personenbezogene Daten, die für Zwecke des Profiling verwendet werden können, insbesondere zur Analyse oder Prognose von Aspekten bzgl. Arbeitsleistung, wirtschaftlicher Lage, Gesundheit, persönlicher Vorlieben oder Interessen, Zuverlässigkeit oder Verhalten, den Aufenthaltsort oder Ortswechsel, soweit dies rechtliche Wirkung für die betroffene Person entfaltet, oder sie in ähnlicher Weise erheblich beeinträchtigt.

Es gibt eine Vielzahl an denkbaren Situationen, bei denen sich beim Einsatz eines KI-Assistenten wie Claude ein hohes Risiko manifestieren kann. Claude kann u.a. ebenso zur Korrektur und Benotung von Klausuren eingesetzt werden wie zur Feststellung von Förderbedarf. Die Erstellung einer Kostenberechnung für einen Fördermittelantrag kann sowohl interne finanzielle Aspekte beinhalten wie auch personenbezogene Daten eingesetzter Mitarbeitender. Informationen, wie die über den Förderbedarf eines Teilnehmenden, können für den Betroffenen zu gesellschaftlichen Nachteilen führen. Auch die vorurteilsbehaftete Darstellung einer Person in einer KI-Antwort kann zu gesellschaftlichen oder wirtschaftlichen Nachteilen führen. Die Kenntnis der veranschlagten Kalkulationsposten kann im Rahmen des Wettbewerbs um den Zuschlag

bei Ausschreibungen zu erheblichen wirtschaftlichen Nachteilen führen. Solche wirtschaftlichen Nachteile können sich auch aus anderen Claude Anwendungen ergeben, z.B. wenn Claude Cowork sämtliche verstreut gespeicherten Dateien zu einem optimal nutzbarer Datenstruktur organisieren soll und damit – da die KI dazu zunächst alle Daten lesen muss - Geschäftsgeheimnisse offenbart werden.

III. Ermittlung von Gefährdungen/ Bedrohungen für die Verfahrenskomponenten

Bedrohungen werden anhand des Gefährdungskatalog des BSI - IT Grundschutzkompendium (Stand 2024/2025) ermittelt.

Unbefugter Informationszugriff

Vertrauliche Daten und Informationen dürfen nur den zur Kenntnisnahme berechtigten Personen zugänglich sein. Vertraulichkeit gehört zu den Grundwerten der Informationssicherheit. Für vertrauliche Informationen (wie personenbezogene Daten, Firmen-/ Geschäftsgeheimnisse, Interne Planungen) besteht die inhärente Gefahr, dass diese durch Unachtsamkeit offengelegt werden.

Geben Nutzer*innen - z.B. Mitarbeitende - versehentlich personenbezogene Daten insbesondere sensible, wie etwa Gesundheitsdaten oder Prüfungsergebnisse in Chats ein, können diese in den Trainingsdaten oder Logs von Claude gespeichert und später auch als Output verwendet werden. Zudem können die Daten jedenfalls dann, wenn Claude über externe Clouddienste genutzt wird, durch API-Aufrufe auch an diese Clouddienste weitergegeben werden.
Die Fa. Anthropic und deren Mitarbeitende könnten Zugriff auf eingegebene Daten haben und selbst bei der Datenübertragung zu Claude könnte es durch sog. Man-in-the-Middle-Angriffe zum Abhören der Kommunikation mit Claude kommen.

Verstoß gegen Gesetze oder vertragliche Regelungen

Wenn Informationen, Geschäftsprozesse und IT-Systeme einer Institution unzureichend abgesichert sind (beispielsweise durch ein unzureichendes Sicherheitsmanagement oder unvollständigen Datenschutz), kann dies zu Verstößen gegen Datenschutzbestimmungen oder gegen bestehende Verträge mit Geschäftspartner*innen führen.

Der Umgang mit personenbezogenen Daten ist durch eine Vielzahl von Vorschriften geregelt. Dazu gehören neben der DS-GVO, dem Gesetz über den kirchlichen Datenschutz (KDG) und den Bundes- und Landesdatenschutzgesetzen auch eine Vielzahl bereichsspezifischer Regelungen. Beispielhaft zu nennen wäre hier die Nutzung von Claude mit personenbezogenen Daten ohne einen DS-GVO-konformen Auftragsdatenverarbeitungsvertrag gem. Art. 28 DS-GVO. (AVV) Auch ein Verstoß gegen Art. 5 c DS-GVO (Datenminimierung) durch Sammeln nicht notwendiger Daten (wie z.B. des Standortes bei Nutzung der Claude App) kann als ein Verstoß gegen die DS-GVO gewertet werden. Hinsichtlich bereichsspezifische Regelungen könnten beispielsweise die Löschroutinen von Anthropic gegen den Sozialdatenschutz nach den Sozialgesetzbüchern verstoßen.

Zur Verletzung vertraglicher Pflichten kann es bei Verträgen z.B. mit Fördermittelgeber*inen, aber auch mit Institutionen der öffentlichen Hand wie etwa der Arbeitsagentur im Rahmen von Arbeitsmarktprojekten kommen. Diese Institutionen sehen in den Verträgen zum Teil Datenverarbeitungsverbote vor, oder verbieten die Übermittlung von Daten in Drittländer wie die USA selbst bei Vorliegen der Voraussetzungen nach Art. 44 ff. DS-GVO. Verstöße gegen die vertraglich geregelte Sicherheitsanforderungen, können eine Vertragsauflösung nach sich ziehen.

Die Geschäftsführung eines Unternehmens ist dazu verpflichtet, bei allen Geschäftsprozessen durch Beachtung anerkannter Sicherheitsmaßnahmen eine angemessene Sorgfalt anzuwenden. Je nach Organisationsform der Bildungseinrichtung ist hier z.B. an das KonTraG (Gesetz zur Kontrolle und Transparenz im Unternehmensbereich), das GmbHG (Gesetz betreffend die Gesellschaften mit beschränkter Haftung) oder das Gesetz zum Schutz von Geschäftsgeheimnissen (GeschGehG) zu denken. Aus diesen Regelungen lassen sich zu Risikomanagement und Informationssicherheit entsprechende Handlungs- und Haftungsverpflichtungen der Geschäftsführung eines Unternehmens ableiten.

<u>Identitätsdiebstahl</u>

Beim Identitätsdiebstahl täuschen Angreifer*innen eine falsche Identität vor, benutzen also Informationen über eine andere Person, um in deren Namen aufzutreten. Hierfür werden Daten wie beispielsweise Geburtsdatum, Anschrift, Kreditkarten- oder Kontonummern

benutzt, um sich beispielsweise auf fremde Kosten bei einem Internet-Dienstleister anzumelden oder sich auf andere Weise zu bereichern. Identitätsdiebstahl führt häufig auch direkt oder indirekt zur Rufschädigung, verursacht aber auch einen hohen Zeitaufwand, um die Ursachen aufzuklären und negative Folgen für die Betroffenen abzuwenden. Werden Daten von Teilnehmenden oder Mitarbeitenden gepromptet und die Eingaben zum Training der KI genutzt, stehen die Daten danach auch zum Identitätsdiebstahl bereit.

Missbrauch personenbezogener Daten

Ein Missbrauch personenbezogener Daten kann auch dann vorliegen, wenn eine Institution personenbezogene Daten ohne Rechtsgrundlage - insbesondere ohne Einwilligung - weitergibt. Die Daten der Teilnehmenden einer Bildungsveranstaltung werden von denen zur Durchführung der Veranstaltung zur Verfügung gestellt. Die Weitergabe von personenbezogenen Daten durch Eingabe in ein LLM wie Claude und auch durch die Weitergabe an alle Subunternehmende und Drittanbieter*innen, ist (soweit nicht technisch erforderlich) eine Zweckänderung, die nicht von der ursprünglichen Rechtsgrundlage für die Datenverarbeitung umfasst ist. Soweit personenbezogenen Daten durch Claude (oder auch andere LLM`s) zum eigenen Training genutzt werden und Subunternehmern Zugriff gewährt wird ist keine Rechtsgrundlage vorhanden. Eine Einwilligung in die Verarbeitung personenbezogenen Daten durch eine KI kann wirksam sein, aber nur unter sehr engen Voraussetzungen. In der Praxis ist eine Einwilligung dann nicht ausreichend, wenn die Nutzung der KI verpflichtend ist, die Verarbeitung der Daten in der Cloud-KI undurchsichtig ist und/ oder die Daten in Drittländer übermittelt werden, was bei Claude von Anthropic mit Firmensitz in den USA der Fall ist. Das Datenschutzrecht ist als Verbot mit Erlaubnisvorbehalt konzipiert. Fällt die Erlaubnis mangels wirksam erteilter oder zurück genommener Einwilligung weg, ist die Verarbeitung personenbezogener Daten durch die KI verboten.

IV. Bewertung der Eintrittswahrscheinlichkeit

Selten/Überschaubar: Als selten wird die Eintrittswahrscheinlichkeit beurteilt, wenn der Eintritt des Schadens möglich ist, die konkreten Umstände aber gegen eine solche Entwicklung sprechen und nicht zu erwarten ist, dass sich daran etwas ändert.

Hinsichtlich der Gefahr eines Identitätsdiebstahls hängt der Schadenseintritt nicht nur davon ab, dass Mitarbeitende personenbezogene Daten durch Claude verarbeiten lassen, sondern zudem davon, dass ein*e Identitätsdieb*in konkret nach personenbezogenen Daten einer bestimmten Person fragen muss. Hat der möglichen Identitätsdieb keine Information darüber, dass personenbezogene Daten durch Claude verarbeitet wurde, ist die Wahrscheinlichkeit eines Schadenseintritts denkmöglich, aber selten.

Das Gleiche gilt für die Gefahr, dass die Geschäftsleitung die notwendige Sorgfalt außer Acht lässt und Geschäftsgeheimnisse in das KI-System eingibt z.B. indem es Claude CoWork zur sortierten Archivierung der geschäftlichen Mails einsetzt. Hier ist der Eintritt eines Schadens möglich, er hängt aber von dem Verhalten der leitenden Mitarbeitenden ab, bei denen (nach entsprechender Schulung) davon auszugehen ist, dass sie sich der Gefahr und auch der Haftungsrisiken für sich selbst bewusst sind und von sich auch einem Schadenseintritt entgegenwirken.

Auch die reine Kenntnis der Kommunikationsdaten der Geschäftsführer*in bzw. der Leitung der IT führen wahrscheinlich nicht zum Eintritt eines Schadens.

Gelegentlich/ Gravierend: Die Eintrittswahrscheinlichkeit wird als gelegentlich bewertet, wenn der Eintritt des Schadens möglich ist, aber von einem Verhalten Dritter abhängt, das sich nicht vorhersagen lässt. Je mehr Mitarbeitende mit einem LLM wie Claude arbeiten, umso höher ist das Risiko des fehlerhaften Umgangs und der rechtswidrigen Verarbeitung personenbezogener Daten. Auch wenn aufgrund Art. 4 KI-VO eine Mitarbeiterschulung vor dem Umgang mit KI verpflichtend ist, lässt sich unmöglich vorhersagen, ob sich alle Mitarbeitenden zu jeder Zeit der Risiken der Eingabe von personenbezogenen Daten und/oder Geschäftsgeheimnissen bewusst ist. Selbst wenn eine Einrichtung die Mitarbeitenden intensiv für die Problematik sensibilisiert, ist nicht auszuschließen, dass Mitarbeitende versehentlich personenbezogene Daten eingeben. Insofern kann es zumindest gelegentlich zum Eintritt eines Schadens durch Offenlegung personenbezogener Daten kommen.

Anthropic selbst, aber auch viele der Subunternehmer*innen verarbeiten die Daten weltweit und damit auch in Drittländern, bei denen das angemessene Datenschutzniveaus nicht unbedingt von sich aus gewährleistet ist. Auch in diesem Zusammenhang ist ein

Datenschutzverstoß möglich, die Wahrscheinlichkeit des Schadenseintritts lässt sich aber nicht konkret vorhersagen.

Hinsichtlich der USA als Drittland sind verschiedene Datenschutzgarantien zu unterscheiden.

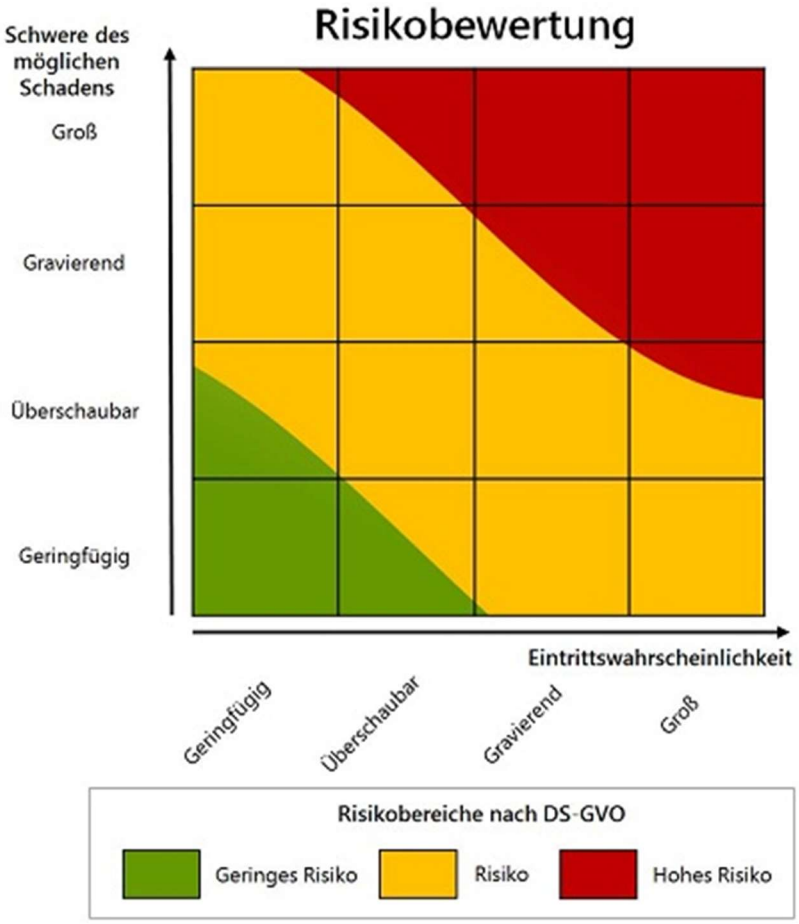
Aufgrund des Angemessenheitsbeschluss der EU-Kommission (2023/1719) gemäß Art. 45 DS-GVO. dient der sog. EU-US Data Privacy Framework zumindest momentan noch als rechtssichere Grundlage für den transatlantischen Datenverkehr. Über dieses Abkommen wird es voraussichtlich im Herbst dieses Jahres seitens des europäischen Gerichtshofs (EuGH) zu einer erneuten Feststellung kommen, ob das Abkommen tatsächlich als Datenschutzgarantie nach Art. 45 DS-GVO ausreicht. Der EuGH hat bereits das vorherige Abkommen - den sog. Privacy Shield – Mitte 2020 vor allem aufgrund der unzureichenden Schutzmechanismen gegen den Zugriff US-amerikanischer Behörden auf personenbezogene Daten von EU-Bürger*innen für ungültig erklärt. US-Präsident Biden hatte daraufhin seinerzeit die Executive Order 14086 erlassen, wonach US-Nachrichtendienste nur noch zielgerichtet und verhältnismäßig auf Daten zugreifen dürfen und keine Massenüberwachung mehr erlaubt ist. Das für die Einhaltung der Order zuständige Kontrollgremium (PCLOB) ist seit Januar 2025 aufgrund der Entlassung mehrerer Gremiumsmitglieder durch den aktuell amtierenden Präsidenten mangels Beschlussfähigkeit handlungsunfähig. Da bereits bekannt ist, dass der EuGH auch die jüngsten Ereignisse bei der Urteilsfindung berücksichtigen wird, gehen nicht wenige Beobachter davon aus, dass der EuGH erneut zur Feststellung der Ungültigkeit des Abkommens kommen wird.

Eine weitere Möglichkeit den transatlantischen Datenverkehr rechtssicher zu gestalten ist die Vereinbarung sog. Standardvertragsklauseln (SCCs) nach Art. 46 Abs. 2 lit. c DS-GVO. Anthropic hat nach eigenen Angaben zumindest für die Enterprise und die API-Variante von Claude bereits Standardvertragsklauseln in den Auftragsdatenverarbeitungsvertrag integriert. Ob diese Vertragsklauseln tatsächlich einen der DS-GVO angemessenen Datenschutz gewährleisten, wird wesentlich davon abhängen, wie die Entwicklung in den USA weiter verläuft. Die Frage ob sich Anthropic dem massenhaften Zugriff von US-Behörden auf personenbezogene Daten von EU-Bürgern entgegenstellen würde, um das

in den Sccc`s vereinbarte Datenschutzniveau durchzusetzen, ist zumindest nicht eindeutig zu bejahen.

Das gleiche gilt für den Fall das Claude über einen Dienst wie AWS Bedrock (Frankfurt) oder Google Cloud (EU) genutzt wird, die eine Datenresidenz innerhalb Europas anbieten. Die Sec. 702 des Foreign Intelligence Surveillance Act (FISA) erlaubt die gezielte Datenerfassung auch über Nicht-US-Bürger*innen. Die Executive Order 12333 und der CLOUD Act erweitern die Datenerfassung auf Internetknotenpunkten und auf außerhalb der vereinigten Staaten belegenden Datenspeichern die unter der Kontrolle US-amerikanischer Telekommunikationsunternehmen stehen. Da es sich bei AWS (Amazon) und Google um amerikanische Unternehmen handelt ist für die Anwendung der oben genannten völlig unerheblich, ob sich der Server physisch in der EU befindet.
Rein formal kann man sich hinsichtlich eines angemessenes Datenschutzniveau der USA noch auf den EU-US Data Privacy Framework und zudem auf die SCC`s im AVV von Anthropic berufen. Die Wahrscheinlichkeit, dass es zu einer ungerechtfertigten Offenlegung von personenbezogenen Daten an die US-Behörden kommt, hängt von der Entwicklung in den USA ab, die sich momentan nicht vorhersehen lässt. Die Tendenz ist aber z.B. einem Schreiben der US-Handelsaufsichtsbehörde im September 2025 an 14 große Tech-Unternehmen zu entnehmen, in dem der Vorsitzende erläutert das die Tech-Unternehmen nicht die EU-Gesetze wie die DS-GVO befolgen müssen, wenn dadurch die Meinungsfreiheit von US-Bürgern eingeschränkt würde.

Bewertung: Geringer bis gravierender Schaden bei seltener bis gravierender Eintrittswahrscheinlichkeit



V. Auswertung:

Vorab bleibt festzustellen, dass hinsichtlich der technischer Sicherheit Claude über ISO 27001 / SOC 2 und Anthropic über SOC 2 Type II-Zertifizierungen verfügen, die als Nachweis für ausreichenden Sicherheitsstandards dienen können.

Das Kernproblem mit Claude Chat von Anthropic ist, dass es sich bei Anthropic um ein US-Unternehmen handelt, personenbezogene Daten wie Nutzeranfragen, Chats, Metadaten auf Servern in den USA und weltweit gespeichert werden und die USA nach der DS-GVO mangels angemessene Datenschutzebene nach Art. 45 ff DS-GVO als unsicherer Drittstaat eingruppiert werden muss.

Die Nutzung von Claude Chat kann formell annähernd datenschutzkonform erfolgen, wenn folgende Gesichtspunkte berücksichtigt werden.

Datenschutzrechtliche Verantwortung:
Die Nutzung der kostenlosen Varianten ist für eine Bildungseinrichtung grundsätzlich ausgeschlossen, weil dann kein AVV abgeschlossen wird und es dementsprechend auch zu keiner Verpflichtung von Anthropic kommt, die Daten nach DS-GVO Standard zu schützen. Der Anthropic AVV ist darauf ausgelegt, die Pflichtanforderungen des Art. 28 DS-GVO grundsätzlich zu erfüllen — einschließlich weisungsgebundener Verarbeitung, Vertraulichkeit, Sicherheit, Subprozessor-Kontrollen, Betroffenen rechten, Löschpflichten und Audioregelungen. Der AVV enthält zudem SCCs für den internationalen Datentransfer.

Rechtsgrundlage:
Die Verarbeitung personenbezogener Daten durch eine KI setzt eine Rechtsgrundlage voraus.
An dieser Stelle muss zwischen der Verarbeitung von personenbezogenen Daten durch Claude Chat zur Generierung eines Outputs und der Nutzung der Daten zur Leistungsoptimierung der KI (Modelltraining) unterschieden werden.
Anthropic erläutert in ihrer Datenschutzerklärung sie nutze die Daten nicht ohne Einwilligung zur Leistungsoptimierung der KI. Eine solche Einwilligung setzt seitens der Bindungseinrichtung ein Einwilligungsmanagement voraus, um zu gewährleisten, das Daten nicht weiterverwandt werden, wenn die Einwilligung nicht erteilt wurde. Die Einwilligung wird vermutlich nicht von allen Mitarbeitenden und/oder Teilnehmenden erteilt werden, da eine wirksame Einwilligung voraussetzt, dass der Zweck der Weitergabe der Daten an Anthropic (also das Training der KI) im Einwilligungstext erläutert wird. Es steht zu vermuten, dass nicht alle Personen dazu bereit sind ihre persönlichen Daten zu Trainingszwecken zur Verfügung zu stellen. Grundsätzlich sollte der Verwendung der Daten zu Trainingszwecken also nicht zugestimmt werden.

Unabhängig von den Anbietenden ist grundsätzlich die Verarbeitung personenbezogener Daten erlaubt, wenn dies zur Erfüllung eines Vertrags oder zur Durchführung vorvertraglicher Maßnahmen, dessen Vertragspartei die betroffene Person ist, erforderlich ist. Gemäß dem EuGH-Urteil vom 04. Juli 2023 zu Meta Plattform müssen Verantwortliche nachweisen können, inwiefern der Hauptgegenstand des Vertrags ohne die betreffende Verarbeitung nicht erfüllt werden könnte. Rechtsgrundlage der Nutzung personenbezogener Daten zur Generierung eines Outputs durch eine KI ist dementsprechend nur selten die Durchführung eines Vertrags oder einer vorvertraglichen

Maßnahme. Vor dem Einsatz einer KI muss allgemein festgestellt werden, ob eine Einwilligung der Betroffenen oder eine andere Rechtsgrundlage wie etwa die des berechtigten Interesses im Sinne des Art. 6 Abs. 1 Buchst. f DS-GVO in Betracht kommt.

Sensible Daten:
Bei der Verarbeitung sensibler, besonders schützenswerter Daten im Sinne des Art. 9 DS-GVO wie z.B. Gesundheitsdaten ist die Verarbeitung solcher Daten z.B. durch Speicherung eines dementsprechenden Inputs auf der Plattform von Anthropic verboten, es sei denn es liegt eine informierte – auf die Risiken hinweisende - Einwilligung des Betroffenen vor. Will man durch Claude Chat sensible personenbezogene Daten verarbeiten, setzt das auch die Information über die Verarbeitung der Daten in den USA und den damit einhergehenden Problemen voraus, denn ansonsten wäre die Einwilligung fehlerhaft und ohne Einwilligung fehlt die Rechtsgrundlage.

Bei der Nutzung der Claude App ist zudem zu beachten, dass die Stimme einer Person dann als biometrische Daten zu werten ist, wenn die Stimme zur Identifizierung einer Person verwendet wird. Sprachaufnahmen ohne Identifizierung fallen unter die allgemeinen DS-GVO Normen des Art. 6 DS-GVO, während die Nutzung der Stimme zur Identifizierung den strengeren n Regeln des Art. 9 DS-GVO unterliegt. Darauf ist bei der Formulierung eines entsprechenden Einwilligungstextes ist zu achten.

Notwendige organisatorische Maßnahmen:
Bei Nutzung der Bezahlvarianten Claude Teams, Claude Enterprise und Claude API ist nach Angaben von Anthropic seit dem 01. Januar 2026 ein AVV mit Standardvertragsklauseln (SCC's) in den Lieferbedingungen enthalten (nicht enthalten bei Claude Pro). Der AVV erfüllt die Mindestanforderungen des Art. 28 DS-GVO.
Die Standardvertragsklauseln, die bei einem möglichen Wegfall des EU-U.S. Data Protection Privacy Framework auch weiterhin ein angemessenes Datenschutzniveau gewährleisten sollen, werden durch die oben erwähnten US-Gesetze zur behördlichen Überwachung auch von Europäern untergraben. Daher ist es notwendig geeignete technische und organisatorische Maßnahmen zu treffen.
Um unter allen Eventualitäten den Datenschutz sicher zu stellen, sollte organisatorisch die Eingabe personenbezogener Daten (aber auch die Eingabe von Geschäftsgeheimnissen)

bei Claude Chat durch eine interne Richtlinie der Bildungseinrichtung ausdrücklich verboten werden.

Technisch kann das bezüglich der personenbezogenen Daten z.T. auch durch ein Pseudonymisierungstool erfolgen, wobei der Einsatz eines solchen Tools allenfalls ein erster Filter ist, allein eingesetzt aber keine Sicherheit gewährleistet. Regelbasierte Systeme scheitern systematisch an der Komplexität, Variabilität und Fehleranfälligkeit realer Daten. Ein mehrschichtiger Ansatz aus Technologie und menschlicher Expertise ist notwendig, sonst ist eine Re-Identifizierung nicht ausgeschlossen und damit ein Datenschutzverstoß vorprogrammiert.

Zur weiteren Risikominimierung sollte - wenn möglich - Zero Data Retention (ZDR) genutzt werden. Dabei werden Inputs nur so lange gespeichert, wie es braucht ein Output zu generieren. Eine weitere Speicherung der Eingabe findet nicht statt. ZDR ist aktuell nur für die Claude Enterprise Variante verfügbar – und auch dort nur für berechtigte Enterprise-API-Kunden und nur auf Basis einer gesonderten Vereinbarung mit Anthropic. In den Standard-Tarifen wie Claude Team, Pro, Free oder der regulären API ist ZDR nicht enthalten. An dieser Stelle trifft sicherlich die Frage der Datensicherheit auf die der finanziellen Machbarkeit. Sollte eine Vereinbarung der ZDR utopisch sein ist es umso erforderlicher, dass eine Richtlinie zum Verbot der Eingabe personenbezogener Daten dann diese Funktion entbehrlich macht. Die Richtlinie sollte allerdings in einer fundierte Mitarbeiterschulung erläutert werden, die auf die Risiken unbedachter Eingaben hinweist und die Richtlinie erläutert.

Anthropic löscht die Daten standardmäßig 30 Tage nach der manuellen Entfernung seitens des Nutzers aus dem Backup. Werden durch die Bildungseinrichtung vertraglich andere Löschfristen vereinbart, sollte die gewählte Bezahlvariante die Möglichkeit einräumen die Löschfristen durch die eigene IT konfigurieren zu können.

Nutzung von Google Vertex AI oder Microsoft Azure (EU):
Bezüglich des Zugriffs von US-Behörden auf Datenbestände der US-Techunternehmen erscheint es der Unterzeichnerin datenschutzrechtlich wenig hilfreich Maßnahmen wie etwa die Nutzung von Claude über Google Vertex AI oder Microsoft Azure (EU) zu vereinbaren. Zum einen macht das das eine Prüfung von Microsoft oder Google bezüglich

deren Datenschutzkonformität erforderlich, zum anderen erlauben – wie oben erläutert - die angeführten US-Gesetze, das US-Behörden auch auf Daten zugreifen können, die sich physisch in der EU befinden, wenn die Verarbeitung durch ein US-Unternehmen erfolgt.

Gemäß Art. 4 KI-VO setzt die Nutzung einer KI eine vorherige Mitarbeiterschulung voraus. Wie oben erläutert sollte bei Einsatz von Claude Chat im Rahmen dieser Schulung ausdrücklich das Problem des Zugriffs amerikanischer Behörden auf alle bei US-Firmen gespeicherten Daten thematisiert werden, um ein Verständnis für das strikte Verbot der Eingabe personenbezogener Daten zu schaffen.

Auf aktuelle Änderung sowohl der Rechtsprechung - mögliche Aufhebung des EU-US Data Privacy Framework (Schrems III) – insbesondere aber auf die Änderung der Datenschutzbestimmungen durch Anthropic muss regelmäßig geachtet werden. Angesichts der schnellen Entwicklungen auf dem KI-Markt ist eine anlasslose Überprüfung der Datenschutzbestimmungen von Anthropic im Abstand von 6 Monaten wohl unentbehrlich.

Zusammenfassend:

- Erforderlich ist der Abschluss eines Auftragsverarbeitungsvertrag (AVV) nach Art. 28 DS-GVO mit SCCs, die den Anforderungen der DS-GVO entsprechen müssen und ggf. durch zusätzliche Maßnahmen (z. B. technische oder organisatorische Schutzvorkehrungen) ergänzt werden, um ein angemessenes Schutzniveau zu gewährleisten.
- Der sicherste Weg zum Datenschutz beim Einsatz von Claude ist der Erlass einer internen Richtlinie mit dem absoluten Verbot der Nutzung personenbezogener Daten.
- Die interne Richtlinie sollte im Rahmen der sowieso notwendigen Mitarbeitendenschulung erläutert werden, um Akzeptanz zu schaffen und das Risikobewusstsein zu fördern.
- Sollten abweichend von der internen Richtlinie in Ausnahmefällen dennoch personenbezogene Daten verarbeitet werden, müssen die betroffene Personen vorab über die Verarbeitung ihrer personenbezogenen Daten durch eine KI und insbesondere über den Drittlandtransfer (USA) informiert werden. Die

Ausnahmefälle sollten organisatorisch geregelt sein, und müssen dokumentiert werden.

- Zudem darf im Ausnahmefall eine Verarbeitung von personenbezogenen Daten durch eine KI nur nach vorheriger schriftlich erteilter Einwilligung erfolgen. Die Einwilligung muss ausdrücklich auf die Risiken bei der Verarbeitung personenbezogener Daten durch US-Unternehmen hinweisen.
- Auf gar keinen Fall sollten sensible Daten durch die KI verarbeitet werden. Gem. Art. 9 DS-GVO sind das personenbezogene Daten, aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person. In Ausnahmefällen ist eine entsprechende informierte Einwilligung als Rechtsgrundlage erforderlich und zu dokumentieren.
- Falls möglich sollte Zero-Data-Retention (ZDR) vereinbart werden, um eine Datenspeicherung personenbezogener Daten von vorneherein zu verhindern. Falls dies nicht möglich ist, kann und sollte seitens der IT die Löschfrist ab 30 Tagen konfiguriert werden um vertraglich vereinbarte Löschfristen (z.B. mit der öffentlichen Hand) auch einhalten zu können.

Wenn diese Maßnahmen getroffen werden, verbleibt dennoch an verschiedenen Stellen ein Restrisiko. Insbesondere hinsichtlich der Frage der Transparenz der Datenverarbeitung durch Claude bleiben Fragen offen. Was der europäische Gerichtshof zur Gültigkeit des EU-U.S. Data-Protection Framwork entscheidet wird und ob es bei Wegfall des Frameworks Anthropic überhaupt möglich sein wird, sich an die vereinbarten SCC’s zu halten, ist ebenfalls nicht eindeutig klar. Ob dieses Restrisiko für eine Bildungseinrichtung akzeptabel ist, richtet sich nach der Art der Daten und ist für jede Einrichtung unterschiedlich. Im Allgemeinen ist eine Nutzung bei gleichzeitigem ausdrücklichem Verbot der Verarbeitung personenbezogener Daten in den allermeisten Fällen ausreichend. Werden hiervon Ausnahmen gemacht, muss klar sein, dass ein Großteil der oben benannten Subunternehmen ebenfalls Kenntnis von den personenbezogenen Daten erhält.

Einrichtungen, die mit einer hohen Anzahl an besonders schützenswerten und sensiblen Daten umgehen, sollten neben der Verbotsrichtlinie als doppelte Sicherung ihr Einwilligungsmanagement überarbeiten und sich zumindest von den Teilnehmenden eine informierte Einwilligung geben lassen. Für den Fall, dass es zu einer versehentlichen Eingabe von personenbezogenen Daten durch einen Mitarbeitenden kommt, ist dafür dann immerhin eine Rechtsgrundlage vorhanden.

Claude Cowork:
Abschließend sei darauf hingewiesen, dass der Einsatz von Claude Cowork ebenfalls voraussetzt, dass keine personenbezogenen Daten oder Betriebsgeheimnisse verarbeitet werden. Dem Nutzer dieser Funktion wird es – jedenfalls zum jetzigen Zeitpunkt – an einer Rechtsgrundlage für die Verarbeitung dieser Daten durch Cowork fehlen

Setzt man beispielhaft Claude Cowork dazu ein, Wissen aus verstreuten Quellen (E-Mails, Chats, Dokumente) zusammenzuführen und durchsuchbar zu machen, fehlt zur Verarbeitung der darin enthaltenen personenbezogenen Daten (Mailabsender, Chatpartner, Namen im Dokument etc.) häufig die Rechtsgrundlage. Das hier in Betracht kommende „berechtigte Interesse“ wird voraussichtlich an der Erforderlichkeit scheitern, denn die Verarbeitungen im Rahmen von KI-Systemen muss bei der Prüfung der Erforderlichkeit möglichen, insbesondere weniger Daten verarbeitenden, Alternativen gegenübergestellt und damit verglichen werden.
Ohne Rechtsgrundlage ist die Verarbeitung rechtswidrig. Sollte der Einsatz einer solchen Funktion angestrebt sein, muss zwingend eine Überarbeitung des Einwilligungsmanagements der Bildungseinrichtung erfolgen, denn durch eine Einwilligung könnte eine entsprechende Rechtsgrundlage entstehen. Dennoch muss sich die Einrichtung darüber klar sein, dass gerade der Einsatz von Cowork im Zusammenhang mit „älteren“ Daten (Mailkontakt vor zwei Jahren) mangels Rechtsgrundlage zu datenschutzrechtlichen Problemen führt. Sollte der Einsatz von Cowork beabsichtigt sein, ist im Hinblick auf die entsprechende Rechtsgrundlage eine ergänzende Risikoabschätzung notwendig.

22.06.2026

Elke E. Thielsch
- externe Datenschutzbeauftragte -

Gefördert durch das
Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen