Leistungsbeschreibung des Gosign privacyRouters

Anlage 1 zum Rahmenwerk über den Zugang · Stand 08.11.2026

Der Gosign privacyRouter (Router) ist eine Programmierschnittstelle, über die Software Anfragen an Sprachmodelle richtet. Jedes Modell erreicht er auf einem Weg, auf dem jede beteiligte Stelle nach § 203 StGB zur Verschwiegenheit verpflichtet ist und der Anbieter Eingaben und Ausgaben nach seinen Zusagen nicht zum Training nutzt und nicht speichert, außer im Zwischenspeicher einer Variante, die der Kunde selbst wählt. Was das je Weg bedeutet, beschreibt Abschnitt 4. Diese Beschreibung gilt für jeden Zugang, gleich ob er einer Kanzlei, einem Unternehmen oder einer selbständig tätigen Person gehört. Kunde ist jeweils die Partei des Rahmenwerks, bei Berufsgeheimnisträgern die Kanzlei.

1. Schnittstelle

- Die Schnittstelle versteht das Format von OpenAI: Chat Completions mit gestreamten Antworten und Werkzeugaufrufen, Einbettungen, Bilderzeugung und die Liste der Modelle eines Schlüssels. Die Schreibweise von Azure OpenAI versteht sie ebenfalls.
- Für die Claude-Modelle versteht sie außerdem das Format der Anthropic Messages API: Nachrichten mit gestreamten Antworten, Werkzeugaufrufen, Bild- und PDF-Eingabe und Zählung der Token. Software auf Grundlage des Claude Agent SDK, Claude Code und Claude Desktop im Modus für Drittanbieter mit einem Gateway stellen dafür die Basis-URL um.
- Bei Claude ist die strukturierte Ausgabe der Schnittstelle nicht freigeschaltet (output_config.format, response_format mit JSON-Schema), weil das Schema dafür nach Angaben von Anthropic bis zu 24 Stunden zwischengespeichert wird. Der Router weist solche Anfragen an Claude ab und entfernt dort die Angabe strict bei Werkzeugen. Die Option outputFormat des Claude Agent SDK arbeitet ohne sie. Bei Gemini reicht der Router response_format mit JSON-Schema durch.
- Für Sprachdialoge in Echtzeit, etwa einen Telefonassistenten, bietet sie eine Echtzeitschnittstelle über eine Verbindung, die für die Dauer des Gesprächs offen bleibt (WebSocket), für Modelle, die Google dafür anbietet. Eingaben sind Text, Ton, Bild und Video, die Antwort ist gesprochen, auf Wunsch der Anwendung mit einer Mitschrift von Eingabe und Antwort. Eine Verbindung endet nach rund zehn Minuten. Die Anwendung baut sie neu auf und gibt den bisherigen Verlauf selbst mit, weil der Router abgebrochene Gespräche nicht wieder aufnimmt (Abschnitt 4).
- Anforderungen eines Zwischenspeichers (cache_control) entfernt der Router, und wiederholte Eingaben kosten jedes Mal den vollen Preis für die Eingabe. Ausgenommen ist die Variante des Wegs über Google mit Zwischenspeicher für eine Stunde, die der Kunde selbst freischaltet und deren Modelle er für einzelne Guardrails selbst wählt (Abschnitt 4): Dort lässt der Router die Anforderung durch, die die Anwendung sendet, und setzt die Dauer des Zwischenspeichers auf eine Stunde, und das Schreiben in den Zwischenspeicher und jeder Treffer daraus kosten die Listenpreise, die Google dafür nennt (§ 5 Abs. 3 des Rahmenwerks). Für das Schreiben gilt Googles Listenpreis für einen Zwischenspeicher von einer Stunde, auch wenn die Anwendung eine kürzere Dauer anfordert. Das Schreiben kostet mehr als die Eingabe ohne Zwischenspeicher, ein Treffer weniger. Fordert die Anwendung einen Zwischenspeicher an, kostet eine Anfrage in der Variante mehr als ohne

Zwischenspeicher, wenn ihr Anfang nicht binnen einer Stunde wiederkehrt. Ein einzelner Treffer gleicht das Schreiben nicht immer aus.

- Die Schnittstellen für Nachrichten, Chat Completions, Einbettungen, Bilderzeugung und Sprachdialoge in Echtzeit sind freigeschaltet. Schnittstellen, über die Inhalte beim Anbieter abgelegt würden, etwa für Dateien oder Stapelverarbeitung, sind gesperrt.
- Software, die eines dieser Formate bereits spricht, ändert Adresse, Schlüssel und Modellkennung. Adresse und Schlüssel zeigt das Konto, wenn der Kunde einen Schlüssel ausstellt.
- Die Übertragung ist verschlüsselt (TLS), zwischen der Software und dem Router ebenso wie zwischen dem Router und den Betreibern der Modelle.

2. Wege und Modelle

Freigegeben werden Wege, nicht Modellnamen oder Modellversionen. Ein Weg ist bestimmt durch die Unternehmen, über die er führt, mit ihrem Ort der Verarbeitung, ihrer Grundlage nach § 203 StGB und ihren Zusagen zu Speicherung und Training, wie sie Abschnitt 4 und die Liste der Unterauftragnehmer nennen. Zu einem Weg gehört jedes Modell, das Gosign über ihn anbietet und für das die Zusagen nach Abschnitt 4 gelten. Zum Weg über Google gehört eine Variante mit Zwischenspeicher für eine Stunde, mit denselben Unternehmen und denselben Zusagen bis auf Zwischenspeicher und Hashwerte (Abschnitt 4).

Weg	Modelle	Ort der Verarbeitung
Google Cloud, EU-Endpunkt	Modelle, die Google dort anbietet, derzeit von Anthropic und von Google	Inferenz in Mitgliedstaaten der EU
Offene Modelle, gehostet in Deutschland	offene Modelle verschiedener Hersteller, ohne Beteiligung der Hersteller	Inferenz in einem Rechenzentrum in Deutschland

Welche Modelle ein Weg umfasst, zeigt der Katalog im Portal des Routers mit ihren Listenpreisen. Welche davon die Guardrails und Schlüssel des Kunden erreichen, zeigen das Konto und die Liste der Modelle eines Schlüssels. Neue Modelle und neue Versionen eines Wegs nimmt Gosign ohne Änderung der Verträge und ohne gesonderte Freigabe auf. Gosign teilt sie mit ihrem Listenpreis mindestens sieben Tage vor ihrer Aufnahme in Textform mit. Ob ein neues Modell zum angekündigten Tag von selbst aufgenommen wird, sofern der Empfang der Mitteilung bis dahin bestätigt ist, oder erst nach der Bestätigung durch eine Person mit der Rolle Inhaber, bestimmt der Kunde im Konto; ein Modell mit einem Preis über dem höchsten der Modelle, die er auf dem Weg schon hat, kommt nur mit dieser Bestätigung. Mit der Aufnahme kommt ein Modell in alle Guardrails des Kunden, die es nicht ausschließen (Abschnitt 3), ein Modell der Variante mit Zwischenspeicher nur in Guardrails, für die der Kunde es wählt. Bestimmte Modellversionen sagt Gosign nicht zu. Stellt ein Anbieter ein Modell ein, bleibt es bis zu seinem Ende nutzbar und entfällt dann aus den Guardrails. Einen Nachfolger teilt Gosign wie jedes neue Modell mit. Einen neuen Weg und jedes neue Unternehmen in der Kette kündigt Gosign nach § 8 Abs. 2 des Auftragsverarbeitungsvertrags mindestens 30 Kalendertage vorher in Textform an, und der Kunde kann widersprechen. Einen neuen Weg nutzt der Kunde erst, wenn er ihn im Konto freischaltet.

Maßgeblich ist der Listenpreis des Anbieters, bei Google die Preisliste der Agent Platform für den jeweiligen Endpunkt in der EU in US-Dollar, bei den offenen Modellen die Preisliste ihres Betreibers in Euro. Befristete Einführungspreise eines Anbieters gibt Gosign nicht weiter, es gilt der reguläre Listenpreis. Wie Gosign abrechnet, regelt § 5 des Rahmenwerks.

3. Konto, Guardrails und Schlüssel

- Der Zugang ist für Anwendungen bestimmt, in denen sensible Daten verarbeitet werden, auch solche, die dem Schutz des § 203 StGB unterliegen. Eine Nutzung außerhalb der Europäischen Union bedarf der vorherigen Zustimmung von Gosign in Textform.
- Jeder Kunde hat ein eigenes Konto im Portal des Routers. Darin legt er Guardrails an, etwa je Team, Abteilung, Anwendung oder Produkt, gibt ihnen ein Budget aus seinem freigeschalteten Guthaben und stellt für sie Schlüssel aus. Schlüssel sperrt und ersetzt er dort selbst. Einen neuen Schlüssel zeigt das Konto beim Ausstellen genau einmal.
- Guardrails enthalten alle Modelle der Wege, die der Kunde freigeschaltet hat, außer denen, die er im Konto ausschließt. Modelle der Variante mit Zwischenspeicher enthalten sie nur, wenn der Kunde sie für diese Guardrails wählt (Abschnitt 4). Ein Ausschluss gilt für das ausgeschlossene Modell und für jede neuere Version desselben Modells auf demselben Weg, nicht für ältere Versionen und nicht für dasselbe Modell auf einem anderen Weg. Neue Modelle kommen mit ihrer Aufnahme in alle Guardrails, die sie nicht ausschließen (Abschnitt 2). Dabei kommen Modelle nur hinzu, und Budgets bleiben unverändert.
- Ein Schlüssel erreicht die Modelle seiner Guardrails, auf Wunsch nur einen Teil davon, und trägt auf Wunsch einen eigenen Höchstbetrag. Begrenzt ist er durch das Budget seiner Guardrails und, wenn gesetzt, seinen Höchstbetrag. Den Durchsatz ihrer Modelle begrenzen die Anbieter; Anfragen über diesen Grenzen weist der Router ab.
- Ist das Guthaben aufgebraucht, das Budget von Guardrails oder der Höchstbetrag eines Schlüssels erschöpft, nimmt der Router über die betroffenen Schlüssel keine Anfragen mehr an, bis neues Budget bereitsteht.
- Guardrails und Schlüssel sind voneinander getrennt: Budget und Höchstbetrag wirken nur auf ihre Guardrails und Schlüssel, und der Router führt Anfragen verschiedener Kunden nicht zusammen. Den Zwischenspeicher der Variante führt Google für das Projekt, das allen Kunden gemeinsam ist (Abschnitt 4).
- Das Guthaben lädt der Kunde im Voraus auf, ab 100 € in Schritten von 100 € netto. Es gilt, sobald Gosign den Eingang bestätigt hat. Guthaben, Testmonat und Rechnungen regeln die §§ 5 und 6 des Rahmenwerks.
- Das Konto zeigt das Guthaben und die Budgets der Guardrails, einen Verbrauch je Person zeigt es nicht. Den Verbrauch je Guardrails und Modell weist Gosign auf Anforderung in Textform aus.
- Im Konto gibt der Kunde Personen Rollen, etwa Inhaber, Manager, Technik, Rechnung, Datenschutz und Betriebsrat. Die Rolle bestimmt, was eine Person im Konto tun darf. Nur Personen mit der Rolle Inhaber nehmen Verträge an, unterschreiben das Rahmenwerk, schalten Wege und die Variante mit Zwischenspeicher frei, erteilen Weisungen zu neuen Modellen und erklären Widersprüche.
- Das Rahmenwerk zeichnet Gosign im Konto. Danach unterschreibt der Kunde es dort durch die Personen, die ihn vertreten, jede mit der Rolle Inhaber, mit Name, Funktion und Ort und nach Anmeldung mit ihrem zweiten Faktor, bei gemeinsamer Vertretung zu zweit. Die Unterschrift gilt für den Text, den das Konto mit seinem Prüfwert zeigt. Statt im Konto zu unterschreiben, kann der Kunde das gezeichnete Rahmenwerk unterschrieben hochladen. Das unterzeichnete Rahmenwerk steht im Konto und geht mit den Anlagen 2 und 3 als Anhang einer Mail an die Personen mit der Rolle Inhaber.
- Mitteilungen in Textform schickt Gosign per Mail an die Personen mit der Rolle Inhaber. Ankündigungen neuer Modelle, Wege und Unterauftragnehmer schickt Gosign per Mail mit

der Ankündigung als PDF an die Personen mit der Rolle Inhaber und Datenschutz, im Konto stehen sie unter Mitteilungen, und den Empfang bestätigen die Empfänger über ihren Link in der Mail.

- Die Prüfspur im Konto hält Änderungen an Guardrails, Schlüsseln, Wegen, Weisungen und Verträgen, die Zeichnung und die Unterschriften des Rahmenwerks mit Person und Zeit fest. Sie lässt sich nicht ändern. Personen mit der Rolle Inhaber, Datenschutz oder Betriebsrat lesen sie und laden einen Auszug herunter.
- Die Telefonleitung, über die Gespräche eine Anwendung erreichen, gehört nicht zum Zugang. Den Vertrag mit ihrem Anbieter schließt der Kunde.

4. Wege, Speicherung und Protokollierung

- Die Modelle bei Google ruft der Router ausschließlich über den Endpunkt von Google für die EU auf, die Multiregion eu. Endpunkte außerhalb der EU sind nicht eingerichtet, Ausweichziele nicht hinterlegt.
- Die offenen Modelle ruft der Router bei ihrem Betreiber in Deutschland auf, der sie auf Servern in einem Rechenzentrum in Deutschland ausführt. Die Hersteller der Modelle sind nicht beteiligt, und kein Unternehmen auf diesem Weg gehört einem US-Konzern an.
- Der Router speichert keine Inhalte, weder Eingaben noch Anweisungen an das Modell noch Antworten. Er protokolliert je Anfrage Zugang, Guardrails und Schlüssel, Modell, Zeitpunkt, Anzahl der Token und Kosten, dazu die Nutzerkennung, sofern die Software sie mitsendet. Diese Verbrauchsdaten bewahrt Gosign bis zum Ablauf von zwölf Monaten nach Ende des Abrechnungsmonats auf.
- Für das Projekt, über das der Router die Modelle bei Google aufruft, gilt Googles Ausnahme von der Protokollierung zur Missbrauchsüberwachung: Eingaben werden zu diesem Zweck weder gespeichert noch von Beschäftigten gesichtet. Die Protokollierung von Anfragen und Antworten ist nicht eingeschaltet. Bei Google wird außerhalb der Variante mit Zwischenspeicher nichts zwischengespeichert: Der Router entfernt jede Anforderung eines Zwischenspeichers, der Zwischenspeicher, den Google bei Gemini ohne Anforderung führt, ist für das Projekt abgeschaltet, und Modelle bei Google, die Eingaben nach Googles Dokumentation ohne Anforderung zwischenspeichern, sind nur freigeschaltet, wenn das für das Projekt abgeschaltet ist.
- In der Variante mit Zwischenspeicher für eine Stunde lässt der Router bei Modellen der Variante die Anforderung eines Zwischenspeichers durch, die die Anwendung sendet, und setzt die Dauer des Zwischenspeichers auf eine Stunde. Google legt die Anfrage bis zu der angeforderten Stelle, etwa mit Anweisungen an das Modell, Werkzeugen und dem bisherigen Verlauf, im Zwischenspeicher ab, auch wenn sie sich nicht wiederholt, und hält sie bis zu einer Stunde nach ihrer letzten Nutzung vor; jede Nutzung verlängert die Frist. Google führt den Zwischenspeicher für das Projekt, über das der Router die Modelle für alle Kunden aufruft; Kennungen des Kunden, etwa Nutzerkennungen und Metadaten, entfernt der Router vor der Weiterleitung an Google, sodass Google vom Router nicht erfährt, welcher Kunde anfragt. Google bildet dafür Hashwerte der Anfragen; diese Hashwerte verarbeitet Google als Verantwortlicher und nicht als Kundendaten und darf sie nach der Google Cloud Privacy Notice auch außerhalb der EU verarbeiten. Einen Ort für Zwischenspeicher und Hashwerte sagt Google nicht zu. Der Zwischenspeicher senkt Wartezeit und, wenn derselbe Anfang mehrmals wiederkehrt, Kosten (Abschnitt 1, § 10 Abs. 6 des Auftragsverarbeitungsvertrags).
- Für die offenen Modelle ist Zero Retention vertraglich zugesichert. Nach den technischen und organisatorischen Maßnahmen des Betreibers werden Eingaben und Antworten nur

flüchtig im Arbeitsspeicher verarbeitet und weder auf Datenträger noch in Protokolle geschrieben. Wiederholte Eingaben hält der Server dabei kurzzeitig im Speicher der Grafikprozessoren vor; die Antwort weist sie als cached_tokens aus, berechnet wird der volle Preis.

- Die Wiederaufnahme abgebrochener Sprachdialoge, für die Google Text, Ton und Video bis zu 24 Stunden speichern würde, lässt der Router nicht zu.
- Suchfunktionen der Modelle, die Inhalte an eine Websuche weitergeben, sind gesperrt.
- Modelle und Funktionen, für die der Anbieter eine Speicherung über den Arbeitsspeicher hinaus voraussetzt, sind nicht freigeschaltet, ausgenommen der Zwischenspeicher der Variante, die der Kunde selbst wählt.
- Weder Gosign noch die Unterauftragnehmer nutzen Inhalte zum Training oder zur Feinabstimmung von Modellen.

5. Verträge und Nachweise

Der Kunde schließt mit Gosign das Rahmenwerk über seinen Zugang. Der Auftragsverarbeitungsvertrag und die Zusatzvereinbarung zur Verschwiegenheit nach § 203 Abs. 4 StGB werden jeweils mit ihrer Annahme im Portal oder mit der Unterzeichnung des Rahmenwerks wirksam, je nachdem, was früher liegt. Bei Berufsgeheimnisträgern ist die Zusatzvereinbarung zugleich der Vertrag nach ihrem Berufsrecht (für Rechtsanwälte § 43e Abs. 3 BRAO, für Patentanwälte § 39c Abs. 3 PAO, für Steuerberater und Steuerbevollmächtigte § 62a Abs. 3 StBerG, für Wirtschaftsprüfer § 50a Abs. 3 WPO, für Notare § 26a Abs. 3 BNotO). Gegenüber Unternehmen und Selbständigen verpflichtet sie Gosign vertraglich zur Verschwiegenheit über alle Inhalte, die sie über den Zugang übermitteln, und ist für Geheimnisse im Sinne des § 203 StGB zugleich die Verpflichtung nach § 203 Abs. 4 Satz 2 StGB. Die Liste der Unterauftragnehmer nennt jedes Unternehmen, das Inhalte verarbeiten kann, mit der Urkunde, mit der es nach § 203 StGB verpflichtet ist, und verweist für die Unterauftragnehmer von Google auf Googles Liste. Die Urkunden und die übrigen Nachweise stellt Gosign im Konto bereit. Einen neuen Weg und einen neuen oder ersetzten Unterauftragnehmer kündigt Gosign mindestens 30 Kalendertage vorher an, und der Kunde kann widersprechen (§ 8 des Auftragsverarbeitungsvertrags). Ankündigungen von Google zu seinen Unterauftragnehmern gibt Gosign unverzüglich weiter.

6. Betreuung

Gosign ist Ansprechpartner für Betrieb, Konto und Abrechnung. Den ersten Schlüssel stellt der Kunde aus, sobald er den Auftragsverarbeitungsvertrag und die Zusatzvereinbarung angenommen hat und ein Guthaben freigeschaltet ist, eine Kanzlei, sobald zudem Gosign ihren Berufsnachweis geprüft hat. Die Software, die einen Zugang nutzt, richtet ihr Betreiber oder der Kunde selbst ein; übernimmt Gosign das auf Wunsch, wird es gesondert vergütet.