SkillPilot Sicherheits- & Rechtsdossier · Band V

Sicherheit, Datenschutz und KI-Transparenz

Pseudonyme Datenhaltung und Löschgrenzen, OpenAI-MCP-Clientbindung

(mTLS, OAuth, 24h-Lernsession) und EU-KI-Verordnung-Compliance

Geltungsbereich dieses Dossiers

Dieser Band vereint die technischen Sicherheits-, Datenschutz- und KI-Transparenznachweise von SkillPilot: das Datenfluss- und Speicherkonzept über Browser, pseudonymes Backend, aktiven Claude-Beta-Pfad und OpenAI-Kandidaten einschließlich 365-Tage-Inaktivitätslöschung, die vierstufige OpenAI-MCP-Clientbindung (HTTPS, Connector-mTLS, OAuth mit PKCE `S256` und 24-Stunden-Lernsession) sowie die strukturierte Compliance-Dokumentation zur europäischen KI-Verordnung (Art. 50 und Anhang III).

Stand: 11. November 2026 · Lizenz: Apache-2.0 (technische Sicherheits- und Compliance-Dokumentation; siehe LICENSING.md).

Technische Architektur- und Bestandsdokumentation — ersetzt keine Rechtsberatung und keine verbindlichen Rechtstexte unter skillpilot.com/privacy bzw. skillpilot.com/legal.

Inhaltsverzeichnis

1 Datenschutz- und Speicherkonzept **3**
1.1 Grundprinzip: Privacy by Design . . . 3
1.2 Datenpartitionierung . . . 4
1.3 Datenfluss-Szenarien . . . 8
1.4 Aufbewahrung und Löschung . . . 8
1.5 Sicherung, Wiederherstellung und verbleibende Betriebsregeln . . . 10

2 OpenAI-MCP: Bindung an den zugelassenen Client **11**
2.1 Schutzziel . . . 11
2.2 Vier unabhängige Nachweise . . . 11
2.3 Verbindliches OAuth-Clientprofil . . . 12
2.4 OAuth-Ablauf und Ressourcenprüfungen . . . 12
2.5 Lernsession als getrennte Anwendungsbindung . . . 12
2.6 mTLS am MCP-Rand . . . 13
2.7 Fail-Closed-Startbedingungen und Secret-Lebenszyklus . . . 13
2.8 Abnahme-, Negativtests, Restannahmen und Referenzen . . . 14

3 KI-Verordnung: Compliance-Stand für SkillPilot **15**
3.1 Rechtskalender . . . 15
3.2 Verifizierter Produktbestand . . . 15
3.3 Art. 50: Einordnung und Umsetzung . . . 16
3.4 Anhang III: Vorläufige Hochrisiko-Abgrenzung . . . 17
3.5 Nachweis-, Release-Status und laufende Pflege . . . 18

1 Datenschutz- und Speicherkonzept

Technisches Datenfluss- und Speicherkonzept

Stand: aktualisiert am 13. September 2026 (mit Abgleich auf den neutralen OpenAI-Endpunkt-Cutover vom Oktober 2026) für die aktive Claude-first-Beta, den nachfolgenden ChatGPT-Integrationskandidaten samt Windows-Desktop-Git-Beta, die Freitextgrenze der Coach-Werkzeuge, freiwilliges Lernziel-Feedback sowie die Lösch- und Aufbewahrungsgrenze der SkillPilot-ID. Die bestehenden Angaben zur Lehreransicht und zum persönlichen Fachzeitplan gelten unverändert fort.

Dieses Dokument beschreibt Datenflüsse und Speicherung aus technischer Sicht. Es ersetzt weder die lernendenseitige SkillPilot-Datenschutzerklärung noch die Datenschutzbedingungen des KI-Anbieters oder eine rechtliche Prüfung vor einer öffentlichen Freigabe. Der öffentliche Datenschutztext wird in `app/src/utils/privacyViewCopy.ts` gepflegt. Die unten genannten Repository-Defaults sind kein Nachweis für die tatsächliche Konfiguration oder Aufbewahrungsdauer jedes produktiven Dienstes, Log-Abzugs oder Backups.

Die verbindliche Arbeitsreihenfolge lautet Claude-Beta → stabiler Kandidat → fokussierte ChatGPT-Integrationsabnahme → offizielle Einreichung. Die providerneutrale Coach-Grenze und das Claude-Runbook regeln die aktive Provider-Spur. Der getrennte OpenAI-MCP-Identitäts- und 24-Stunden-Session-Vertrag sowie 2 beschreiben den nachfolgenden OpenAI-Integrationskandidaten und die explizit autorisierte Desktop-Git-Beta, nicht ein abgenommenes öffentliches Verzeichnis-Release.

1.1 Grundprinzip: Privacy by Design

SkillPilot trennt strikt zwischen **pseudonymer Lernenden-Identität**, **persistentem Lernstand**, **temporärem Provider-Zugriff** und **providerseitigem Gesprächsinhalt**:

- Das SkillPilot-Backend speichert Lernfortschritte unter einer zufälligen, pseudonymen Skill Pilot-ID — ohne Namen, E-Mail-Adresse oder Passwort zu verlangen. Dies ist Pseudonymität, keine Anonymität: Die ID verknüpft eine Lernhistorie, und wer eine Zuordnung zwischen ID und Person besitzt, kann die Person identifizieren.
- Der Browser bildet die Anmeldeoberfläche für SkillPilot. Er hält die dauerhafte SkillPilot-ID lokal vor und kann sie optional mit einem selbst gewählten Passwort verschlüsselt ablegen.
- Die dauerhafte SkillPilot-ID verbleibt ausschließlich in First-Party-Oberflächen von Skill Pilot und gelangt weder in den KI-Chat noch in OAuth-Principals, MCP-Tools oder providergehostete Widgets. Jeder explizite Start im WebGUI erzeugt eine separate, kurzlebige Lernsession-Referenz, die das Backend intern auf die ID auflöst.
- Der KI-Anbieter verarbeitet Chatverlauf, Uploads, projizierten Lernkontext und Tool-Antworten unter dem jeweiligen Provider-Account und dessen Bedingungen.
- SkillPilot erhält kein vollständiges Chat-Transkript des Providers, empfängt aber jeden explizit gesendeten API- oder MCP-Toolaufruf einschließlich der für das Lesen oder Ändern des Lernstands nötigen strukturierten Argumente.

Architekturinvariante für Coach-Werkzeuge: Antworten der Lernenden, Erklärungen, Gesprächszusammenfassungen und vom Coach formulierte Rückmeldungen verbleiben ausschließlich im Provider-Chat. Weder `workFeedback`, `outcomeFeedback`, Recall-Felder wie `feedback/lastFeedback` noch umbenannte Äquivalente gehören in akzeptierte Coach-Eingaben, gespeicherten Lernstand, Replay-Antworten oder Diagnosedaten. Der Coach trifft die pädagogische Einschätzung und formuliert seine Erfolgsmeldung selbst; SkillPilot nimmt ausschließlich die autorisierte

strukturierte Entscheidung entgegen und gibt den bestätigten Zustand zurück (siehe provider-neutrale Coach-Grenze). Das Bereinigungsprotokoll vom 12. September 2026 dokumentiert eine historische Produktionsbereinigung samt damaligen Rollout-Pflichten und ist kein Beleg für den Zustand späterer Deployments.

Der Status der Provider-Spuren bleibt explizit getrennt:

1. **Claude V1 OAuth/MCP:** die aktive Lern-Beta über den persönlichen Claude-Marketplace. Installation und Updates umgehen weder OAuth noch Session-Isolation oder fachliche Autorisierung.
2. **Mehrsprachige OpenAI-V1-OAuth/MCP-App:** der ChatGPT-Integrationskandidat (`skillpilot-coach` auf dem neutralen Zielendpunkt `https://mcp-openai.skillpilot.com/mcp`, flankiert von der explizit autorisierten Windows-Desktop-Git-Beta). Vollständige Host-Abnahmen über Desktop/Voice, Web und Mobile bleiben Voraussetzung einer Verzeichnis-Veröffentlichung; Quellcode, lokale Tests oder bestandene Claude-Tests belegen keine ChatGPT-Host-Abnahme.
3. **Visible Session:** ausschließlich historischer Custom-GPT-Rollback-Vertrag, keine beworbene Beta und kein aktueller MCP-Vertrag. Sein temporäres `sps_...`-Bearer-Token ist bewusst in Startnachricht und Status-Footer sichtbar.

Alle Provider-Spuren nutzen denselben fachlichen Kern, wahren jedoch strikt getrennte OAuth- und Session-Grenzen. Sprachunterstützung und Frontend-Integration sind releasegebunden; bloßes Vorhandensein im Quellcode begründet keine öffentliche Plattformzusage.

1.2 Datenpartitionierung

A. Lokaler Client (Browser / Local Storage) **Vertrauensgrenze:** kontrolliert durch die lernende Person und ihre Browserumgebung. Der Browser kann folgende Daten halten:

- die aktive SkillPilot-ID;
- optionale benannte lokale Profile, in denen die SkillPilot-ID mit einem selbst gewählten Passwort verschlüsselt ist (das Passwort selbst wird nicht gespeichert);
- eine browserlokale OpenAI-Teilnahmebestätigung, gebunden an das aktuelle pseudonyme Skill Pilot-Profil für die Lebensdauer des Browser-Tabs;
- das gewählte Curriculum sowie weitere UI-Einstellungen;
- lokalen SRS-Wiederholungszustand vor oder zwischen Synchronisationen;
- die frisch vorbereitete Provider-Start-URL bzw. Startnachricht mit einer kurzlebigen Lernsession-Referenz, jedoch niemals mit dauerhafter SkillPilot-ID, OAuth-Token oder Client-Secret.

Jeder V1-Coach-Start ruft mit der lokal aktiven dauerhaften SkillPilot-ID einen lernendenspezifischen UI-Endpunkt auf und erzeugt genau in diesem Moment eine frische Lernsession-Referenz. Der Claude-Start öffnet den vorbereiteten Chat direkt; die ChatGPT-Desktop-Beta zeigt die providerspezifische Startnachricht zum Einfügen in einen neuen Desktop-Chat an. Da eine Start-URL oder Startnachricht mit temporärer Referenz selbst schutzbedürftig ist und in der Browserchronik oder kopierten Links auftauchen kann, ist sie entsprechend vertraulich zu behandeln.

Im lehrkraftbegleiteten Einsatz kann der Browser der Lehrkraft bzw. ein institutionell kontrollierter Speicher zusätzlich lokale Klassennamen, Namen oder Aliase, dauerhafte SkillPilot-IDs und zwischengespeicherte Level-2-Personalisierungen halten; SkillPilot speichert diese Klassenzuordnung nicht zentral. Die Lehreransicht liest reguläre Profil- und Mastery-Endpunkte direkt über die bekannte dauerhafte ID. Bei der erstmaligen Planung eines Lernbereichs ermittelt ein zusätzlicher `no-store`-Lesezugriff alle atomaren Ziel-IDs der vollständigen Level-2-Personalisierung des gewählten Fachs samt offener Teilmenge (unabhängig vom veränderlichen Level-3-Fokus) und legt

Ziel-IDs, Summenzähler und Erfassungszeitpunkt als unveränderliche lokale Planungs-Baseline ab — ohne SkillPilot-ID oder numerische Einzel-Masterywerte in der Baseline. Die reguläre Lehreransicht verändert keine lernendenseitig geplanten Ziele und erzeugt serverseitig weder Lehrkraft-, Klassen-, Berechtigungs- noch Mitgliedschaftsdatensätze.

Die Lehreransicht sperrt Lernstandsänderungen im Frontend, während der browserlokale Kursplan editierbare Arbeitsdaten der Lehrkraft bleibt. Eine explizit bestätigte Aktion kann Planbezeichnung, datierte Blöcke und clientseitig materialisierte atomare Ziel-IDs in einen getrennten persönlichen Fachzeitplan unter der bekannten dauerhaften SkillPilot-ID kopieren. Der Server prüft die IDs gegen den personalisierten Fachumfang: Neu hinzugefügte IDs werden nur akzeptiert, solange sie offen sind; bereits enthaltene IDs dürfen zur Wahrung der Plankontinuität erhalten bleiben. Es werden weder Klassen-ID, Alias, Durchführungsjournal, Bestätigungen, Masterywerte, Planungs-Baseline noch Planhistorie übertragen und keine Lehrkraft-Lernenden-Beziehung angelegt. Spätere lokale Planänderungen werden nicht synchronisiert; von der Lehrkraft eingegebene Titel werden unverändert kopiert und können personenbezogene Angaben enthalten.

Die bestätigte First-Party-Aktion **Planung wirksam machen** validiert sämtliche aktuellen Fachpläne vor dem Schreiben vollständig (alles oder nichts), aktiviert das plangesteuerte Lernen unter der bekannten SkillPilot-ID und wählt sofort das erste fällige Ziel mit erfüllten Voraussetzungen (jederzeit pausierbar). Die nachgeordnete Aktion **Nur dieses Fach aktualisieren** ersetzt lediglich die betreffende Fachkopie, ohne den Planmodus zu aktivieren oder ein Ziel zu wählen. Ein expliziter Fachwechsel parkt ein unvollendetes aktives Ziel ohne Mastery-Änderung und wählt ein fälliges zulässiges Ziel im gewählten Fach; nach bestätigtem Abschluss hat ein gültiger Plan mit dem abgeschlossenen Ziel Vorrang, andernfalls greift die deterministische Dringlichkeits- und Fachreihenfolge. Veraltete oder ungültige Pläne bleiben ungenutzt, ohne fälliges zulässiges Ziel wird kein Ersatz erfunden, der generische Autopilot bleibt während aktivem Planmodus unterdrückt, und reiner Kalenderfortschritt löst keinen Schreibvorgang aus. Planexporte lassen die abgeleitete Baseline weg, exportieren Freitexttitel jedoch unverändert.

Die dauerhafte SkillPilot-ID bleibt das Inhabergeheimnis und der Vollzugriffsschlüssel auf den Lernstand (das Löschen einer lokalen Klasse lässt die ID gültig und unverändert). Heruntergeladene lokale Klassenexporte werden daher im Browser zwingend mit einem von der Lehrkraft gewählten Passwort über versioniertes PBKDF2-SHA-256 und authentifiziertes AES-256-GCM verschlüsselt; SkillPilot speichert oder rekonstruiert das Passwort nicht. Ältere unverschlüsselte JSON-Klassenexporte bleiben zur Migration importierbar (im UI als ungeschützt markiert), neue Exporte werden niemals im Klartext geschrieben. Die Dateiverschlüsselung schützt die exportierte Datei im Ruhezustand, verschlüsselt aber weder den lokalen Browserspeicher noch ein entsperrtes Gerät und macht die ID nach dem Entschlüsseln nicht schreibgeschützt.

B. Persistenter Lernstand im SkillPilot-Backend

Vertrauensgrenze: pseudonyme, zentrale Datenhaltung im SkillPilot-Backend. Gespeichert werden insbesondere:

- `Learner`: der dauerhafte pseudonyme Lernendenschlüssel;
- providerspezifische aktive Lernsessions samt absolutem Ablaufzeitpunkt und direkter interner Zuordnung zum Lernenden;
- `Mastery`: Lernziel-Mastery pro Lernendem und Ziel;
- geplante Ziele und Lernkonfiguration (Curriculum, Personalisierung, Scope, aktives Ziel, Filter und synchronisierter Client-Zustand);
- optionale persönliche Fachzeitpläne, deren Revisionen und materialisierte atomare Ziel-IDs sowie die Präferenz für plangesteuertes Lernen (standardmäßig deaktiviert bis zur bestätigten Aktivierung);
- statische Curricula, Kompetenzdefinitionen, SRS-Decks, Ressourcen und Review-Metadaten.

Kein Modellanbieter benötigt die dauerhafte SkillPilot-ID für die aktive Claude-Integration oder den OpenAI-Kandidaten; sie dient ausschließlich serverintern als Verknüpfungsschlüssel von der temporären Session zum Lernstand. Der Betreiber hat bestätigt, dass der SkillPilot-Core in Deutschland gehostet wird; dies bedeutet nicht, dass gewählte KI-Anbieter, Supportdienste oder sämtliche Auftragsverarbeiter ausschließlich in Deutschland arbeiten — die providerseitige Verarbeitung folgt den offengelegten Bedingungen des jeweiligen Anbieters.

C. Historische Visible-Session-Daten — ausschließlich Rollback Dieser beibehaltene Legacy-Pfad ist weder die aktive Claude-Beta noch ein aktueller ChatGPT-MCP-Vertrag. Das Backend speichert eine `ChatSession` mit HMAC-Hash des temporären Tokens, Lernendenzuordnung, Sprache, Start-/Ablauf-/Nutzungszeitstempeln und optionalen Widerrufsdaten; Klartext-`sps_...`-Tokens werden nur bei der Erzeugung ausgegeben und dürfen niemals geloggt werden. ChatGPT sieht das Token in Startnachricht und Pflicht-Footer, den projizierten Lernkontext, Auswahlnummern/-referenzen, öffentliche kanonische Lernziel-IDs sowie Recall-Karten-IDs und Fragen. Das Token ist ein kurzlebiges Berechtigungsmerkmal (absoluter Ablauf nach höchstens 24 Stunden ohne Verlängerung durch Nutzung), keine Anonymisierung des Gesprächs.

D. Daten des mehrsprachigen OpenAI-V1-OAuth/MCP-Kandidaten Dieser Abschnitt beschreibt das implementierte Sicherheits- und Datenprofil der OpenAI-V1-Spur (`skillpilot-coach`, vorbereitet für den neutralen geschützten Endpunkt `https://mcp-openai.skillpilot.com/mcp` als Nachfolger von `https://mcp-coach-v1.skillpilot.com/mcp`; reale Host-Abnahme, Cutover-Rollout und Verzeichnis-Veröffentlichung bleiben getrennte Schritte). Das Backend hält diese Spur strikt getrennt von Visible-Session-Tabellen und Claude-Verbindungen.

Der geschützte `/mcp`-Endpunkt nutzt serverauthentifiziertes HTTPS, OpenAI-Connector-m TLS und ein gültiges OAuth-Access-Token; ein dedizierter Nginx-vHost leitet ausschließlich `/mcp` auf den Loopback-Handler `/internal/openai/v1/mcp` weiter (siehe 2). Für das gehostete ChatGPT-Verzeichnisprofil akzeptiert der Authorization Server genau den vorkonfigurierten vertraulichen OAuth-Client (`client_secret_basic` + PKCE `S256` + exakte Prüfung von Client-ID, Redirect-URI, Resource, Audience und Scopes; Ablage des Client-Secrets nur in Chat GPT-Appkonfiguration und SkillPilot-Secret-Store); für die explizit autorisierte native Windows-Desktop-Route dient ein separates, exakt gepinntes öffentliches CIMD-Profil mit PKCE `S256`, das niemals als Fallback einer fehlgeschlagenen vertraulichen Authentisierung fungiert. Offene DCR, ungepinnte CIMD, `none`, `private_key_jwt` und implizite Profilwechsel sind ausgeschlossen. Der mTLS-Rand prüft OpenAIs veröffentlichte CA-Kette, `clientAuth`-EKU und den exakten Connector-SAN und speichert nur die feste Klassifikation (`VERIFIED`, temporär `OBSERVE_NO_CERT` oder Loopback `LOCAL_OPERATOR`), niemals das Zertifikat selbst (`observe` dient dem Integrationsnachweis; eine Verzeichnis-Veröffentlichung verlangt `enforce`).

Gespeichert werden die OAuth-Autorisierungs-, Consent-, Code-, opaken Access- und rotierenden Refresh-Token-Daten sowie je explizitem **Lernen starten** ein separater Lernsession-Datensatz mit ausschließlich HMAC-Hash der hochentropischen Referenz, Lernendenzuordnung, `started_at`, `expires_at` und Widerrufsmetadaten. OAuth authentisiert die App, die Session adressiert den Lernenden: Jeder fachliche MCP-Aufruf verlangt beides zugleich. Die Start-Intents sind eng begrenzt (aktuelle Lerneinheit, Verified Recall mit Ziel und Batchgröße sowie geprüfter Abi-2026-Prüfungseinstieg mit Ziel und Kursniveau); freie Browser-Prompts werden nicht als Startzustand persistiert.

Anlegen oder Laden der dauerhaften SkillPilot-ID, Bestätigen der Provider-Hinweise sowie die gesamte Level-2-Personalisierung erfolgen ausschließlich im First-Party-WebGUI. ChatGPT hält OAuth-Tokens als reine Transport-Credentials im `Authorization`-Header (unsichtbar für Nutzende und außerhalb von Modellprompts/Argumenten), während die temporäre `learning`

`SessionId` automatisch in den Startprompt für fachliche Toolaufrufe eingesetzt wird, ohne die dauerhafte SkillPilot-ID preiszugeben. Über `CoachStateProjection` projiziert MCP nur den benötigten Workflow-Kontext; Prüfungslösungen und Bewertungsschemata erfordern den autorisierten Evaluationspfad, und Recall-Antworten werden erst nach dem Antwortschritt der Lernenden freigegeben.

E. Aktive Claude-V1-OAuth/MCP-Daten Claude ist der aktive Lern-Beta-Provider: Der persönliche Marketplace liefert Plugin-Paket und Coach-Anweisungen, während fachliche Operationen die isolierte Claude-V1-MCP-Grenze über HTTPS und OAuth aufrufen. Das Claude-Autorisierungs-Principal ist eine generierte, lernendenunabhängige technische Verbindungsreferenz statt der dauerhaften SkillPilot-ID. Jeder explizite Start im First-Party-WebGUI erzeugt eine frische, exakt 24 Stunden gültige Lernsession-Referenz (gespeichert als Hash, Lernendenzuordnung, Erstellungs-/Ablaufzeitpunkt, Locale und Zustandsversion; weder Chataktivität noch OAuth-Refresh verlängern sie). Fachliche Aufrufe erfordern sowohl gültiges OAuth als auch die temporäre Lernsession.

Claudes Client-Authentisierungsprofile sind von OpenAI getrennt (unterstützt werden das eingeschränkte öffentliche Claude-CIMD-Profil sowie ein explizit konfiguriertes vertrauliches Profil mit PKCE und providerspezifischer Prüfung; OpenAI-Connector-mTLS gilt nicht für Claude, und eine öffentliche Client-ID ist kein Geheimnis). Projizierte Kontextdaten enthalten keinen dauerhaften Lernendenschlüssel, bleiben aber pseudonyme Lerndaten. Eingehende Operationen akzeptieren ausschließlich strukturierte Ziel-/Kartenreferenzen, Sprache, Nebenläufigkeits-/Idempotenzschlüssel, autorisierte Capability-Referenzen und eng begrenzte Entscheidungen (optionaler numerischer Prüfungsscore; boolesches `passed` je Recall-Karte); nicht unterstützte Ergebnisfelder werden vor Mutation und Replay-Speicherung abgewiesen, während Antworten, Fotos, Spracheingaben und Erklärungen im Claude-Chat bleiben (`ClaudeV1AppAuthenticationFilter`, `ClaudeV1LearningSessionService`, `ClaudeV1McpContractAdapter`, `ClaudeV1CoachContextProjector`).

F. Bewusst eingereichtes Lernziel-Feedback Cockpit und Lernzielbuch öffnen dasselbe freiwillige öffentliche Feedback-Formular (`credentials: 'omit'`, Übergabe nur des öffentlichen Ziel-/Buchkontexts ohne Lernenden-ID, kein automatischer Export aus Provider-Chats). Unter eigenem Hinweis und expliziten Bestätigungen erfasst es Kategorie, Beobachtung sowie optional Begründung/Beleg, Vorschlag, Quelle und Perspektive ohne Namens-, E-Mail- oder Anhangsfelder (Nutzende werden angewiesen, keine Chats, Zugangsdaten oder unnötigen Personendaten einzufügen, da Längen-/Feldprüfungen freie Personendaten nicht erkennen).

`goal_feedback_submission` speichert den Text mit zufälligen Feedback-/Einreichungs-IDs, Eingangszeitpunkt, öffentlicher Ziel-/Buch-/Seitenbindung samt Fingerprints, Hinweisversion/-sprache, Bestätigungen und technischen Digests **ohne Lernenden-Fremdschlüssel** (das Rate-Limiting hasht die Client-IP kurzlebig im Speicher, was keine Aussage über Infrastruktur-Logs trifft). Offene Einträge werden nach 30 Tagen zur Löschung fällig (tägliche Bereinigung plus Bereinigung bei Intake/Export, daher nennt das Formular höchstens 31 Tage im Online-Posteingang). Betreiber exportieren Batches lokal nach `tmp/goal-feedback/inbox/<export-id>` und quittieren den Digest erst nach verifizierter Speicherung, wodurch der Online-Inhalt gelöscht wird und nur ein inhaltsfreier Batch-Beleg verbleibt. Exportierte Texte dürfen für die offengelegte technisch unterstützte Prüfung genutzt werden; lokale Rohdateien und Codex-Protokolle bilden eigene Aufbewahrungsgrenzen mit manueller Löschpflicht nach Abschluss gemäß Feedback-Intake-Runbook (`GoalFeedbackSubmissionService`, `GoalFeedbackRetentionService`, `GoalFeedbackExportService`, `GoalBookFeedbackPilotView.tsx`). Widerrufs- oder Löschwünsche laufen über `support@skillpilot.com` (vorzugsweise mit Feedback-ID).

1.3 Datenfluss-Szenarien

- **Individuelle Anmeldung:** Beim Öffnen von `https://skillpilot.com` erzeugt der Browser eine SkillPilot-ID, entsperrt ein gespeichertes Profil oder übernimmt eine vorhandene ID, speichert Curriculum und Konfiguration pseudonym im Backend und öffnet das Cockpit oder startet eine verfügbare Coach-Integration.
- **Claude-Beta-Verbindung und Lernstart:** Nach Plugin-Installation und OAuth-Verbindung (getrennt vom Lernendenprofil) wählt die Person im SkillPilot-WebGUI ihren Kontext und startet über **Lernen starten** eine 24-stündige Claude-Lernsession; Claude nutzt OAuth + temporäre Referenz für MCP-Tools, bewertet Antworten im Chat und sendet nur strukturierte Ergebnisse an SkillPilot zurück.
- **Historische ChatGPT Visible Session (nur Rollback):** `POST /api/ui/learners/{skillpilotId}/visible-chat-start` erzeugt ein maximal 24 Stunden gültiges `sps_...`-Token (serverseitig als HMAC-Hash) und öffnet den sprachspezifischen Custom GPT gegen `/api/ai/{lang}/sessions/.../visible`.
- **Kandidaten-Abnahme — OpenAI-V1-App-Verbindung:** Nach browserlokaler Bestätigung der länderspezifischen Alters-/Erziehungsberechtigtenregeln (ohne Geburtsdatum und ohne Altersableitung aus Jahrgangsstufen) durchläuft ChatGPT OAuth Authorization Code + PKCE `S256` gegen das gepinnte Clientprofil (`client_secret_basic` für die gehostete App bzw. separates gepinntes Desktop-CIMD-Profil) und erhält resource-/scope-gebundene Tokens ohne Lernendenauswahl.
- **Öffnen des MCP-Coaches ohne vorbereitete Session:** Fehlt eine `learningSessionId`, ruft das Modell kein Tool auf, fragt weder dauerhafte noch temporäre IDs ab und verweist lokalisiert auf `https://skillpilot.com/` und **Lernen starten**.
- **Kandidaten-Abnahme — OpenAI-Session mit Backend-Locale:** Jeder Klick auf **Lernen starten** erzeugt eine frische 24-Stunden-Session (HMAC-Hash, Lernendenzuordnung, `communicationLocale`, `started_at`, `expires_at`, Widerrufsdaten; wiederholte Starts erzeugen getrennte Sessions ohne Verlängerung älterer) und fügt die Referenz automatisch in die Startnachricht ein; das Backend verlangt für jeden Fachaufruf gültiges OAuth und gültige Session ohne Fallback.
- **Widerruf der App-Autorisierung & Cockpit-Einsicht:** Ein serverseitiger OAuth-Widerruf entwertet Autorisierung und Access-/Refresh-Tokens sofort, lässt getrennte Session-Datensätze ohne gültiges OAuth unbenutzbar und erhält den pseudonymen Lernfortschritt. Die Cockpit-Einsicht (`GET /api/ui/learners/{skillpilotId}/state`) findet ausschließlich zwischen Browser und SkillPilot-Backend statt.

1.4 Aufbewahrung und Löschung

Die Repository-Standardlaufzeiten für Zugangsdaten in Claude V1 und dem OpenAI-V1-Kandidaten betragen:

- **OAuth-Access-Token:** 1 Stunde;
- **Rotierendes Refresh-Token:** 30 Tage;
- **Claude-V1-Lernsession:** feste 24 Stunden ab dem First-Party-Start;
- **OpenAI-V1-Lernsession:** exakt 24 Stunden ab der bestätigten First-Party-Aktion **Lernen starten**.

Die Ablauffrist einer Lernsession wird weder durch MCP-Aufrufe noch durch Access-Token-Refresh, Browser-Reload, neuen Chat oder Kontextkompaktierung verlängert. Refresh-Token-Rotation kann die Token-Gültigkeit fortschreiben und darf nicht als automatische Löschung aller

Verbindungsdaten nach 30 Tagen beschrieben werden; abgelaufene Session-Zeilen werden standardmäßig stündlich bereinigt. OAuth-Laufzeiten sind keine allgemeine Datenbank-Löschfrist für abgelaufene Autorisierungs-, Consent- oder Diagnosedatensätze, und das Löschen eines Lernenden entfernt nicht automatisch jede lernendenunabhängige technische Provider-Autorisierung. Historische Visible-Session-Tokens laufen nach höchstens 24 Stunden ab; Lernfortschritt wird niemals durch Ablauf oder Widerruf einer KI-Verbindung gelöscht.

Der aktive Lernstand unter einer SkillPilot-ID in der SkillPilot-Datenbank einschließlich zugehöriger SkillPilot-Lernsessions und lernendengebundener Verbindungen unterliegt einer gemeinsamen Löschgrenze:

- die lernende Person kann ihn jederzeit manuell über die vorgesehene WebGUI-Funktion löschen; und
- nach 365 aufeinanderfolgenden Tagen ohne erfolgreiche Aktivität wird er zur automatischen Löschung fällig und beim nächsten automatischen Löschlauf entfernt.

Als Aktivität zählen ausschließlich folgende erfolgreich abgeschlossene Vorgänge:

- erfolgreiche Erstellung einer SkillPilot-ID;
- das aktive Laden oder Fortsetzen des Lernstands in der SkillPilot-Weboberfläche;
- ein vom Server abgeschlossener Import oder Export signierter Lerndaten;
- eine serverseitig erfolgreich gespeicherte Änderung des Lernstands;
- eine erfolgreich abgeschlossene SkillPilot-Sitzungs- oder KI-Anbieter-Verbindungsaktion; oder
- ein gültiger Coach-/MCP-Aufruf, den SkillPilot mit einem fachlich erfolgreichen Ergebnis abschließt.

Hintergrund-GET-Anfragen, SSE-Verkehr, OAuth-Token-Aktualisierungen, das bloße Auswählen, Öffnen oder Lesen einer lokalen Datei sowie vom Server nicht abgeschlossene oder fachlich abgewiesene Aktionen zählen nicht und starten die 365-Tage-Frist nicht neu.

Nach der Löschung hat die dauerhafte SkillPilot-ID keinen Zugriff mehr auf diesen Lernstand, zugehörige Lernsessions und lernendengebundene Verbindungen sind unbenutzbar, und eine getrennt bestehende technische OAuth-Autorisierung kann den gelöschten Lernenden weder wiederherstellen noch auswählen. Mangels Klarnamenbindung erfordern Auskunfts- oder Löschwünsche zu einem Lernprofil einen verhältnismäßigen Zugriffsnachweis über einen vereinbarten sicheren Kanal (der Support darf dauerhafte IDs oder temporäre Credentials niemals per gewöhnlicher E-Mail anfordern; Feedback-Anfragen nutzen ihre Feedback-ID). Lokale Dateien und Provider-Chats sind von dieser Löschung unberührt. Bestehende Sicherungskopien gehören nicht zum aktiven Lernstand; weder manuelle Löschung noch der 365-Tage-Ablauf löschen sie unmittelbar einzeln.

Klartext-Sessionreferenzen, OAuth-Credentials/-Secrets, Request-Bodies, Antworten von Lernenden und Startgeheimnisse dürfen nicht in Anwendungs-, Reverse-Proxy-, Trace- oder Support-Logs gelangen. Der allgemeine `RequestLoggingFilter` überspringt geschützte MCP-, OAuth-, Feedback- und Coach-Ergebnisrouten vollständig, während die begrenzte OAuth/MCP-Telemetrie nur Operation, Ergebnis, Latenz, Korrelations-IDs und HMAC-abgeleitete Session-Fingerprints ohne freie Toolargumente erfasst. Andere API-Routen können auf `DEBUG` feldredigierte, größenbegrenzte Bodies protokollieren; optionales AI-Tracing (`tmp/ai-trace.jsonl`, standardmäßig aus) nutzt Feld-/Tokenmuster-Redaktion statt allgemeiner Personendatenerkennung und besitzt weder automatische TTL noch Kopplung an die Lernendenlöschung. Ebenso wird das Legacy-Feld `ChatStartRequest.promptContext` (auf 2.000 Zeichen gekürzt, nicht in der Session gespeichert) vom allgemeinen Debug-Logger nicht als Ganzes redigiert und darf nicht ohne Textprüfung beworben oder getraced werden.

1.5 Sicherung, Wiederherstellung und verbleibende Betriebsregeln

Lernende und Bildungseinrichtungen tragen selbst die Verantwortung für die Aufbewahrung ihrer Zugriffsschlüssel und externen Namenszuordnungen: Signierte Exporte/Importe übertragen Lernstände zwischen Browsern, passwortgeschützte ID-Dateien sichern die SkillPilot-ID verschlüsselt, lokale Klassenexporte verschlüsseln Lehrkraft-Listen (Passwortpflege und Altdatenlöschung liegen bei der Einrichtung), ein Logout leert nur die aktive Browsersitzung, und eine serverseitige Lernendenlöschung belässt lokale Exporte/ID-Dateien auf den Endgeräten.

`scripts/backup_db.sh` setzt restriktive Dateirechte, nutzt standardmäßig `BACKUP_RETENTION_DAYS=30`, erstellt SQL-Dumps ohne Feedback-Inhaltstabellen und entfernt überalterte Dateien nur bei erfolgreichen Sicherungsläufen (ohne eigene Dump-Verschlüsselung im Skript und ohne stundengenaue 30-Tage-Garantie). `application.yml` schreibt standardmäßig `INFO`-Logs nach `logs/skillpilot.log` mit 10-MB-Rotation und `max-history: 7`; die realen produktiven Fristen für Anwendungs-, Proxy-, Trace-, Snapshot-, WAL- und Offsite-Kopien sind gesondert betrieblich zu bestätigen und dürfen nicht stillschweigend mit aktiver Lernenden- oder Online-Feedback-Löschung gleichgesetzt werden.

Wesentliche Betriebsregeln und Restrisiken:

- **Zugangsdaten im Betrieb und Support:** Aktive Lernsession-Referenzen (`Claude`, `OpenAI` oder historische `sps_...`-Tokens) sowie OAuth-Access-/Refresh-Tokens sind Inhaber-Credentials — unredigierte Chats, Screenshots, Browserchroniken, HAR-Dateien oder Logs dürfen niemals geteilt werden, während kanonische Lernziel- und Karten-IDs öffentliche technische Referenzen sind.
- **Reichweite von Dateiverschlüsselung und Provider-Chats:** Die Verschlüsselung von Klassenexporten schützt die heruntergeladene Datei im Ruhezustand und erkennt Manipulationen, verschlüsselt aber weder den aktiven Browserspeicher `skillpilot_classes` noch ein entsperrtes Gerät. Lernende sind darauf hinzuweisen, keine sensiblen Personendaten in den KI-Chat einzugeben oder hochzuladen, da der Provider diese nach eigenen Bedingungen verarbeitet und eine SkillPilot-Löschung keinen Provider-Chat entfernt.
- **Wahrheitsgemäße Release-Gates:** Visible Session bleibt reiner Rollback-Pfad, und weder OpenAI-Kandidatencode noch Claude-Verfügbarkeit belegen eine Host-Abnahme auf einem anderen Client. Vor jedem providerspezifischen Release oder Cutover sind Datenschutzhinweise, Provider-Angaben, reale Aufbewahrungs-/Logging-/OAuth-Profile, echtes Widerrufsverhalten und offene Auftragsverarbeiter-/Backup-Bestätigungen gegen das konkrete Release zu prüfen.

2 OpenAI-MCP: Bindung an den zugelassenen Client

Verbindliche Sicherheitsarchitektur für den OpenAI-MCP-Coach

Ursprungsstand: 31. Juli 2026 · Aktualisiert auf die dauerhafte neutrale Plugin-Entscheidung vom 19. Oktober 2026 (`skillpilot-coach` auf `https://mcp-openai.skillpilot.com/mcp`, die den früheren Endpunkt `https://mcp-coach-v1.skillpilot.com/mcp` vor der Verzeichnis-Einreichung ablöst; OAuth-Issuer bleibt `https://skillpilot.com/api/openai/v1`). Identität, Cutover und Lebenszyklus sind im Versionierungs- und Lebenszyklusplan sowie im Stable-Origin-Runbook festgelegt.

2.1 Schutzziel

Die SkillPilot-ID bleibt das von der lernenden Person verwahrte Geheimnis, an dem ihr Lernstand hängt. Dieses Kapitel behandelt ein davon unabhängiges Schutzziel:

> Ein fremder Client darf die lernendenbezogenen OpenAI-V1-MCP-APIs nicht nutzen, auch wenn er den öffentlichen Serververtrag und eine gültige `learningSessionId` kennt.

Diese Schutzgrenze verändert weder SkillPilot-ID noch Lernziel-, Mastery-, Curriculum- oder Coach-Semantik.

2.2 Vier unabhängige Nachweise

Nachweis	Zweck	Verbindlicher Mechanismus
SkillPilot-Server	Vertraulichkeit und Serverauthentisierung	Reguläres HTTPS/TLS am dedizierten Reverse Proxy.
OpenAI-Connector	Nur gehostete Chat GPT-Verbindungen aus Open AIs Zertifikatskette passieren den produktiven Edge-Pfad	mTLS mit OpenAI Root CA, Connectors-Intermediate, `clientAuth`-EKU und exaktem SAN `mtls.prod.connectors.openai.com`.
Zugelassene MCP-App	Nur der explizit zugelassene Client erhält verwendbare OAuth-Tokens	Vorregistrierter vertraulicher OAuth-Client (`client_secret_basic` + Callback-Allowlist + PKCE `S256`) für den gehosteten Pfad; getrenntes, exakt gepinntes Desktop-CIMD-Profil (`S256`) ohne Fallback für den nativen Desktop-Pfad.
Aktuelle Lernsession	Der konkrete Chat adressiert genau den beim UI-Start gewählten Lernenden	Bei jedem **Lernen starten** frisch erzeugte, exakt 24 Stunden gültige `learningSessionId` als Pflichtargument jedes fachlichen MCP-Tools.

Für jeden fachlichen Aufruf gilt kumulativ:

```
gueltiges, client- und resource-gebundenes OAuth Access Token
AND
gueltige explizite learningSessionId
AND (im produktiven gehosteten enforce-Modus)
gueltiges OpenAI-Connector-Clientzertifikat
```

OAuth authentisiert und autorisiert die App; die Lernsession wählt den Lernenden. Keiner der Nachweise ersetzt den anderen.

2.3 Verbindliches OAuth-Clientprofil

Für den gehosteten ChatGPT-Verzeichnispfad (**SkillPilot Coach**, technischer Name `skillpilot-coach`) existiert genau ein vorkonfigurierter vertraulicher OAuth-Client:

- feste, vom App-Autor vergebene `client_id` und langes kryptografisch zufälliges `client_secret`;
- bytegenau die in der ChatGPT-Appverwaltung hinterlegte Callback-URL;
- Token-Endpunkt-Authentisierung ausschließlich über `client_secret_basic` samt Authorization Code mit PKCE ausschließlich `S256`;
- exakte öffentliche Resource/Audience `https://mcp-openai.skillpilot.com/mcp` (keine Doppel-Audiences zum historischen `mcp-coach-v1`-Ursprung);
- getrennte Read- und Write-Scopes, kurze Access-Token-Laufzeit, Refresh-Token-Rotation und expliziter Widerruf.

Das Client-Secret liegt ausschließlich an zwei geschützten Orten: (1) in der vertraulichen OAuth-Konfiguration der ChatGPT-App und (2) in der geschützten SkillPilot-Serverumgebung — niemals in Repository, Browserbundle, Startprompt, Chat, MCP-Argumenten, Antworten oder Logs.

Ergänzend ist für die native ChatGPT-Desktop-Integration ein **strikt getrenntes, exakt gepinntes öffentliches CIMD-Profil** (`7yTcOAOqrumX`-Client/Callback auf derselben neutralen Resource mit PKCE `S256`; siehe Native-CIMD-Runbook) autorisiert. Dieses Profil ist niemals ein Fallback einer fehlgeschlagenen vertraulichen Authentisierung. Offene Dynamic Client Registration (DCR), ungepinnte CIMD-Dokumente, beliebige `none`-Clients und `private_key_jwt` sind im Produktionsprofil ausgeschlossen.

2.4 OAuth-Ablauf und Ressourcenprüfungen

Der Authorization Server akzeptiert einen Code-Flow nur, wenn (1) `client_id` exakt einem explizit gepinnten Profil entspricht, (2) `redirect_uri` bytegenau in dessen Allowlist steht, (3) PKCE `S256` vorliegt, (4) `resource` exakt die aktive MCP-Resource adressiert und (5) nur erlaubte Scopes angefordert werden. Beim Token-Request muss der vertrauliche Client zusätzlich das Secret per HTTP Basic nachweisen.

Bei jedem MCP-Aufruf prüft der Resource Server Signatur bzw. aktive Introspektion, Aussteller (`https://skillpilot.com/api/openai/v1`), Ablauf und Widerruf, exakte Audience/Resource, exakte Client-ID, den erforderlichen Read-/Write-Scope sowie zusätzlich die explizite gültige `learningSessionId`. Ein gültiges OAuth-Token darf weder einen Lernenden auswählen noch eine Session anlegen; das OAuth-Subject dient niemals als Identitätsquelle oder Fallback.

2.5 Lernsession als getrennte Anwendungsbindung

Gemäß OAuth- und Lernsession-Vertrag entsteht genau beim Klick auf **Lernen starten** für jeden Start eine neue zufällige ID (absolut 24 Stunden gültig, niemals verlängert; serverseitig nur als HMAC-Hash mit Lernendenzuordnung gespeichert). Neue fachliche Operationen benötigen mindestens `PT1H` Restlaufzeit (exakt `PT1H` ist gültig). Ein bereits committeter Write darf sein gespeichertes Resultat nur bei mindestens `PT1H` Restlaufzeit, identischem Toolnamen,

kanonisch gleichen Argumenten, derselben `clientRequestId`, verfügbarer gepinnter Workflow-/Curriculumversion und unveränderter Learner-Revision ohne erneute Mutation replayen.

`SESSION_REQUIRED`, `SESSION_RENEWAL_REQUIRED` und `SESSION_VERSION_UNAVAILABLE` sind keine OAuth-Fehler: Der Coach gibt `instruction` unverändert aus (bzw. den exakten Eintrag aus `instructions` und ergänzt `startUrl` nur, falls noch nicht enthalten) — ohne Fachantwort, ohne Retry mit alter Session und ohne neuen OAuth-Flow. Der Lernende startet über **Lernen starten** im First-Party-WebGUI eine frische Session in einem neuen Chat; die dauerhafte SkillPilot-ID bleibt serverseitig.

2.6 mTLS am MCP-Rand

Gehostete ChatGPT-Connectoren präsentieren beim TLS-Aufbau ein von OpenAI verwaltetes Clientzertifikat. Der dedizierte MCP-vHost vertraut ausschließlich der veröffentlichten OpenAI Root CA und dem Connectors-Intermediate (ohne rotierendes Leaf-Pinning), während ein Loopback-Verifier Kette, `clientAuth`-EKU und den exakten SAN `mtls.prod.connectors.openai.com` prüft:

- **`observe`:** Gültige Zertifikate werden als `VERIFIED` klassifiziert; zertifikatslose Verbindungen passieren vorübergehend bis zur unveränderten OAuth-Schicht und zählen ausschließlich als `OBSERVE_NO_CERT` (für kontrollierte Integrations- und native Desktop-Tests, kein gehosteter Veröffentlichungszustand). Ungültig präsentierte Zertifikate werden stets abgewiesen.
- **`enforce`:** Externe `/mcp`-Aufrufe benötigen zwingend `VERIFIED`. Lediglich unmittelbare Loopback-Socket-Peers dürfen als `LOCAL_OPERATOR` den unveränderten OAuth-Pfad lokal testen; Internet-Header, Queryparameter oder `X-Forwarded-For` erzeugen niemals einen Bypass.

Protected-Resource-Metadaten und Domain-Challenge bleiben ohne Clientzertifikat erreichbar. mTLS authentisiert die OpenAI-Infrastruktur und ersetzt weder OAuth noch die Lernsession.

2.7 Fail-Closed-Startbedingungen und Secret-Lebenszyklus

Bei aktiviertem OpenAI-V1-Provider bricht der Serverstart fail-closed ab, sobald eine Bedingung verletzt ist: eigenständiges hochentropisches `SKILLPILOT_SIGNING_SECRET` mit mindestens 32 Nicht-Leerzeichen (keine Platzhalter/Defaults, niemals geloggt), konsistente gepinnte Client-Konfiguration (feste Client-ID, ausreichend langes Client-Secret ab 32 — vorzugsweise 48+ — Zufallsbytes für `client_secret_basic`, exakte HTTPS-Callback-Allowlist, PKCE `S256`, exakte HTTPS-Resource/OAuth-Endpunkte), exakte Modus-Übereinstimmung zwischen Root-Nginx, Backend und interner Klassifikation sowie geschlossene offene DCR/ungepinnte Profile. Port `8787` darf öffentlich nicht erreichbar sein; der Reverse Proxy terminiert TLS und leitet nur freigegebene Pfade an `/internal/openai/v1/mcp` weiter.

Das OAuth-Client-Secret wird außerhalb des Repositories erzeugt, liegt identisch nur in Chat GPT-Appverwaltung und SkillPilot-Serverumgebung, erscheint nie in Status-/Fehlerausgaben und wird bei Verdacht auf Kompromittierung sofort rotiert (inklusive Token-Widerruf und Neuautorisierung). Eine Rotation von `SKILLPILOT_SIGNING_SECRET` macht laufende Lernsessions gezielt unauflösbar (geplanter Session-Cutover). Ein Clientwechsel reaktiviert niemals alte Sessions als Identitätsersatz.

2.8 Abnahme-, Negativtests, Restannahmen und Referenzen

Vor einer Freigabe müssen objektiv scheitern: Token-Requests ohne/mit falschem Secret oder fremder Client-ID; nicht gepinnte Methoden (`client_secret_post`, offene DCR, ungepinnte `none`/CIMD/`private_key_jwt`); falsche Callback-URL oder fehlendes/falsches PKCE; fehlende/falsche Resource oder fehlender Scope; widerrufenes/abgelaufenes Token; gültiges OAuth ohne `learningSessionId` sowie gültige Session ohne OAuth; unbekannte, manipulierte, widerrufene oder abgelaufene Session; direkter Zugriff auf Port `8787`; im `enforce`-Modus externe `/mcp`-Aufrufe ohne gültiges OpenAI-Clientzertifikat, mit falscher Kette/SAN/EKU oder gefälschten Forwarding-/Klassifikationsheadern. Positiv wird der vollständige Ablauf über die konfigurierte App geprüft (der öffentliche positive mTLS-Nachweis im `enforce`-Modus ist ausschließlich über den echten gehosteten ChatGPT-Connector möglich).

Restannahmen: Nicht abgedeckt sind Kompromittierungen des SkillPilot-Hosts, Reverse Proxys oder der ChatGPT-Appkonfiguration, Verlust von SkillPilot-ID, aktivem Session-Link oder Client-Secret sowie vorsätzliche Weitergabe von Access-Token plus Session — begrenzt durch Secret-Rotation, kurze Tokenlaufzeiten, feste 24h-Sessions, Widerruf, minimale Scopes und redigierte Logs.

Normative OpenAI-Referenzen: Client identification · Client registration & authentication · Resource/Audience parameter · Mutual TLS (mTLS).

3 KI-Verordnung: Compliance-Stand für SkillPilot

Dokumentierter Compliance-Stichtag und fortgeschriebenes Inventar

Stichtag dieser Bestandsaufnahme: **29. Juli 2026** (historischer Compliance-Nachweis im Repository; ersetzt keine Rechtsberatung). Das maschinenlesbare KI-Transparenzinventar wird bei neuen Assets deterministisch fortgeschrieben; für die spätere Provider-Reihenfolge ab September 2026 (aktive Claude-Lern-Beta vor fokussierter ChatGPT-Abnahme) gilt ergänzend die Claude-first-Release-Strategie sowie 1.

3.1 Rechtskalender

Track	Maßgeblicher Termin	Bedeutung für SkillPilot
Art. 50 KI-VO	**2. August 2026**	Transparenz bei direkter KI-Interaktion sowie bei bestimmten synthetischen Inhalten. Die Übergangsfrist bis **2. Dezember 2026** betrifft ausschließlich Art. 50 Abs. 2 für bereits vor dem 2. August 2026 in Verkehr gebrachte generative Systeme.
Anhang III / Art. 6 Abs. 2	**2. Dezember 2027**	Kapitel III Abschnitte 1–3 gelten ab diesem Datum für Anhang-III-Hochrisikosysteme. Dieser Track ist rechtlich von Art. 50 getrennt.
Anhang I / Art. 6 Abs. 1	**2. August 2028**	Späterer, für den aktuellen SkillPilot-Sachverhalt nicht vorrangiger Produktregulierungs-Track.

Rechtsgrundlagen sind die KI-Verordnung (EU) 2024/1689, die seit 27. Juli 2026 geltende Änderungsverordnung (EU) 2026/1744 sowie die am 20. Juli 2026 angenommenen Leitlinien der Europäischen Kommission zu Art. 50 (Auslegungshilfe, nicht Verordnungstext).

3.2 Verifizierter Produktbestand

Das fortlaufend geprüfte maschinenlesbare Transparenzinventar bildet den technischen Nachweis; zum Stichtag 29. Juli 2026 wurden im Repository folgende Fakten und Aussagegrenzen festgestellt:

Klasse	Repository-Fakt am 29. Juli 2026	Aussagegrenze
Lernziel-Visualisierungen	`app/public/assets/goal-visualizations/` enthält 1.494 Bilder (1.484 JPG, 10 PNG), identisch zum kanonischen Source-Bestand.	Dateizahl und Format belegen weder Erstbereitstellung vor einem Stichtag noch eine Deepfake-Einstufung.
Bild-Provenienz	Ein Byte-Scan findet in 1.491 Bildern einen `c2pa`-Marker und in 1.484 Bildern `Created by Google Generative AI`; Prompt-Sidecars nennen häufig Gemini / Nano Banana Pro.	Keine kryptografische Validierung von C2PA-Signatur, Zertifikatskette oder Manifestintegrität — Provenienzindizien statt validiertem Art.-50-Nachweis.
Narrative Illustrationen & Comics	Fünf semantische Sammlungen mit 24 Runtime-Dateien (12 bilinguale Quickstart-Szenen plus 12 Sprach-/Formatvarianten aus vier Comic-Sammlungen inkl. `docs/comic4/`).	Dateizahl entspricht nicht der Zahl eigenständiger Werke; Erzeugungsdienst und Erstveröffentlichung sind nicht lückenlos belegt.

Klasse	Repository-Fakt am 29. Juli 2026	Aussagegrenze
Whitepaper-Rasterfiguren	`docs/whitepaper/` und `app/public/whitepaper/` enthalten 24 gebundene PNG-Dateien (davon 2 Kopien aus `docs/comic4/`), zusammen mit den narrativen Sammlungen also 46 eindeutige Runtime-Bilder außerhalb der Lernziel-Visualisierungen.	Gemischte bzw. unvollständig belegte Provenienz; aus der Inventarisierung folgt ausdrücklich nicht, dass alle 24 Figuren KI-generiert sind.
Lerncoach (Stichtag 29.07.2026)	Zum Stichtag zwei ChatGPT-Varianten; das Dialogmodell läuft beim externen KI-Anbieter, SkillPilot liefert Lernkontext und Werkzeuge.	Die am 29.07.2026 pausierte Claude-OAuth/MCP-Spur wurde erst im September 2026 durch Product-Owner-Beschluss zur aktiven Lern-Beta erhoben (siehe 1).
Texte und Lernkarten	Fachliche und redaktionelle QA-Lanes; Entstehungsgeschichte nicht für jeden Einzeltext einheitlich providerbezogen nachgewiesen.	Keine pauschale Behauptung reiner KI-Erzeugung; die Ausnahme für Texte von öffentlichem Interesse setzt im Einzelfall belegte menschliche Prüfung und redaktionelle Verantwortung voraus.

Audio-Bestand

- **`app/public/audio/intro-de.m4a`** — Dauer `1095,540680 s`, Encoder-Tag `Google`, SHA-256: `11af439ea19741567bd797d7b05314c2a1f5c416413beed5f781e2fd7d1cf196`
- **`app/public/audio/intro-en.m4a`** — Dauer `892,482177 s`, Encoder-Tag `Google`, SHA-256: `e4d53b9f15932b7b1ce863cf06a48b36323a8a1efa09f0238fe96407a471a923`
- **`docs/notebooklm/SkillPilot_macht_Bildung_zum_Navigationssystem.m4a`** — Dauer `801,413515 s`, Encoder-Tag `Google`, SHA-256: `267aafede53d9849a980cccd0304abc47c7069edee5ecac5ab8263c06c594876`
- **`docs/notebooklm/Skill_Graphs_Tame_the_Smart_Liar.m4a`** — Dauer `1134,410884 s`, Encoder-Tag `Google`, SHA-256: `1542430f14f1beee4738dfa5b655b7fc1d819503a95e31dc1f02dbd580da43f8`

Die veröffentlichten Audioeinführungen sind nicht byteidentisch mit den beiden archivierten NotebookLM-Dateien. Ein Encoder-Tag oder UI-Symbol beweist weder den konkreten Erzeugungsdienst noch die verwendete Stimme; die vorsorgliche, providerneutrale Player-Kennzeichnung *„Diese Audioeinführung enthält KI-erzeugte Stimmen“* ersetzt daher keinen vollständigen Segment- und Stimmennachweis.

3.3 Art. 50: Einordnung und Umsetzung

- **Rollenabgrenzung:** Art. 50 Abs. 2 adressiert den Anbieter des generativen KI-Systems (maßgeblich ist das in Verkehr gebrachte System, nicht allein das Modell-Hosting). SkillPilot wird für seine Integration und Veröffentlichung vorsorglich als **Betreiber** behandelt; eine zusätzliche **Anbieterrolle** ist neu zu prüfen, falls SkillPilot ein KI-System unter eigenem Namen in Verkehr bringt, wesentlich verändert oder dessen Zweck ändert. Rollenprüfungen sind je System und Inhalt zu dokumentieren.
- **Direkte Interaktion:** Der Lerncoach läuft im Chat des gewählten Providers; zusätzlich nennt der SkillPilot-Startbereich den KI-Assistenten explizit und warnt vor möglichen Fehlern, statt sich bloß auf „Offensichtlichkeit“ zu stützen (insbesondere mit Blick auf Minderjährige).

- **Vierstufige kumulative Deepfake-Prüfung (Art. 3 Nr. 60):** Ein KI-Asset ist nur dann ein Deepfake, wenn kumulativ (1) KI-Erzeugung/-Manipulation, (2) Bild-/Audio-/Videoformat, (3) Ähnlichkeit mit realen Personen, Gegenständen, Orten, Einrichtungen oder Ereignissen (auch ohne namentlich identifizierbares Einzelvorbild) und (4) ein Kontext vorliegen, in dem es fälschlich authentisch oder wahrheitsgemäß wirkt (unter strengerem Verständlichkeitsmaßstab bei Minderjährigen). Bei schematischen Lernillustrationen spricht der Darstellungscharakter regelmäßig gegen Merkmal (4) — als Klassenannahme ohne pauschale Freistellung und ohne globales Bild-Badge. Die beiden Audioeinführungen tragen vorsorglich direkt am Player den klaren KI-Stimmen-Hinweis, da eine bloße Legal-Unterseite für Art. 50 Abs. 4 allein nicht genügen würde.
- **Altbestand und Fünf-Schritte-Gate für neue Inhalte:** Mangels lückenlosem Erstbereitstellungsbeleg vor dem 2. August 2026 beansprucht SkillPilot keinen pauschalen Altbestandsschutz, sondern stützt sich auf inhaltliche Art.-50-Prüfung und kontextnahe Player-/Coach-Hinweise. Für nach dem 2. August 2026 KI-erzeugte/-manipulierte oder neu eingeordnete Inhalte gilt vor Veröffentlichung verbindlich: (1) Akteur und Rolle feststellen, (2) Inhaltstyp und Deepfake-Merkmale prüfen, (3) bei Texten von öffentlichem Interesse menschliche Prüfung und redaktionelle Verantwortung belegen oder am Inhalt offenlegen, (4) erforderlichenfalls klar, barrierefrei und spätestens bei erster Exposition kennzeichnen sowie (5) Entscheidung, Version, Datum und Prüferrolle im Inventar festhalten. Vorhandene C2PA-Strukturen werden beim Deployment nicht entfernt, ersetzen aber keine erforderliche menschenlesbare Offenlegung.

3.4 Anhang III: Vorläufige Hochrisiko-Abgrenzung

Systemgrenze und KI-Systemeigenschaft (Art. 3 Nr. 1) sind für Lerncoach, deterministische Zustandswerkzeuge und etwaige Bewertungsmodule getrennt nach Zweck, Unterlagen, Vermarktung und Einsatzkontext zu beurteilen (gestützt u. a. auf die nicht bindenden Beispiele des AI Act Service Desk für Bildung):

Einsatzkontext	Vorläufige Einstufung
Selbstgewähltes Lernen außerhalb einer Bildungs- oder Berufsbildungseinrichtung ohne formale Noten-, Zugangs- oder Zertifikatswirkung	Derzeit **nicht** als Anhang III Nr. 3 Buchst. b behandelt; die KI-Einschätzung bleibt ein weiches Hilfsmittel ohne formale Entscheidungswirkung.
Einsatz durch eine Bildungs- oder Berufsbildungseinrichtung, bei dem KI-Ausgaben Lernleistungen bewerten oder darauf gestützt den Lernprozess steuern	**Vorsorglich als Hochrisiko nach Anhang III Nr. 3 Buchst. b behandeln**, bis eine dokumentierte Gegenprüfung anderes trägt; vor Aktivierung ist ein eigenes Release-Gate zwingend.

Sechs Trigger für die Hochrisiko-Spur: (1) Vertrag, Pilot oder Marketing für Schulen, Hochschulen oder Berufsbildungseinrichtungen, (2) Nutzung von KI-Ausgaben oder Mastery-Werten für Noten, Zertifikate, Zulassung, Versetzung oder Abschluss, (3) institutionell vorgegebene Lernsteuerung aufgrund einer KI-Bewertung, (4) neue automatische Bewertungs-, Ranking- oder Prognosefunktionen, (5) Profiling nach Art. 4 Nr. 4 DSGVO innerhalb eines bereits von Anhang III erfassten Systems (bloße Personalisierung ist nicht automatisch Profiling) sowie (6) eigenes Inverkehrbringen, wesentliche Änderung oder Zweckänderung eines KI-Systems.

Eine Ausnahme nach **Art. 6 Abs. 3** greift niemals bei Profiling und verlangt kumulativ fehlendes erhebliches Risiko, fehlende wesentliche Entscheidungsbeeinflussung sowie mindestens eine der vier **alternativen** Fallgruppen: (1) enge Verfahrensaufgabe, (2) Verbesserung des Ergebnisses

einer zuvor abgeschlossenen menschlichen Tätigkeit, (3) Muster-/Abweichungserkennung ohne Ersatz menschlicher Prüfung oder (4) vorbereitende Aufgabe für eine Anhang-III-Bewertung — einschließlich Vorab-Dokumentation und EU-Datenbankregistrierung nach Art. 49 Abs. 2 durch den Anbieter.

3.5 Nachweis-, Release-Status und laufende Pflege

Maßnahme	Dokumentierter Stand (Baseline 29. Juli 2026)
KI-Transparenzabschnitt in `/legal` (DE/EN)	Im Repository umgesetzt, ohne neues Zustimmungs- oder Haftungsverzichts-Gate und ohne erneuten Consent-Zwang.
Allgemeines Haftungshinweis-Gate	Technisch und rechtlich von Art. 50 getrennt und unverändert beibehalten (zwingende gesetzliche Rechte bleiben unberührt).
KI-Hinweise am Audio-Player & Coach-Start (DE/EN)	Im Repository samt lokalem Render-, Copy- und Accessibility-Smoke umgesetzt.
Maschinenlesbares Inventar & CI-/Deploy-Gates	`check:ai-transparency-inventory`, `check:ai-transparency-artifact` und Deploy-Readiness-Checks verifizieren UI, Inventar und gebaute Artefakte.
Altbestandsschutz & C2PA-Kryptovalidierung	Werden beide nicht beansprucht; vorhandene C2PA-Marker sind lediglich als Byte-Struktur dokumentiert.
Institutionelles Hochrisiko-Gate	Zukunftstrigger vor einem institutionellen Pilot oder Release; kein Defizit des freiwilligen Selbstlernangebots.

Pflegeanlässe: Dossier und Inventar werden zwingend neu geprüft bei Provider- oder Modellwechsel, neuer Assetklasse, realistischer Stimme oder Personendarstellung, Änderung von Coach- oder Bewertungsfunktionen, institutionellem Vertrieb sowie wesentlicher Zweckänderung; rollenbezogene KI-Kompetenzmaßnahmen nach Art. 4 KI-VO werden begleitend unterstützt.