

Leistungsbeschreibung

Leistungsbeschreibung

Gegenstand ist die einmalige Erzeugung eines Sprachmodells aus den vom Kunden bereitgestellten Dokumenten sowie dessen digitale Bereitstellung über einen Aktivierungscode und ein Einrichtungsprogramm zum Download. Grundlage ist ein offenes Basismodell der Qwen3.5-Reihe unter Apache-2.0-Lizenz, das per LoRA-Feinabstimmung angepasst wird: Die vortrainierten Gewichte bleiben unverändert, angelernt werden zusätzliche Matrizen. Der Kunde betreibt das Ergebnis auf eigener, selbst beschaffter Hardware in seinem Netz.

STAND 12. August 2026 HERAUSGEBER IonKon GmbH

1. Gegenstand der Leistung

Nach der Auslieferung findet bei IonKon keine Verarbeitung von Kundeninhalten mehr statt. Das ausgelieferte System hat keinen Rückkanal zu IonKon und läuft nach der Einrichtung ohne Internetverbindung. Ein Auftrag erzeugt genau ein Modell aus genau dem Bestand, der zu diesem Auftrag hochgeladen wurde.

Ausdrücklich nicht Gegenstand der Leistung sind:

- Hardware. IonKon liefert kein Gerät und vermittelt keines.
- Eine laufende Dienstleistung, ein Wartungsvertrag oder eine Cloud-Anbindung.
- Betrieb, Betreuung oder Überwachung des Modells im Netz des Kunden.
- Eine spätere Feinabstimmung mit neuem Material; dafür ist ein neuer Auftrag nötig.
- Garantien im Rechtssinne. Beschreibungen auf der Website und in diesem Dokument sind keine Garantien.

2. Kenndaten

Alle Preise sind Nettopreise und verstehen sich zuzüglich der gesetzlichen Umsatzsteuer. Der Preis gilt einmalig; laufende Lizenz- oder Nutzungsgebühren fallen nicht an, und die Zahl der Nutzer im Kundennetz ist nicht begrenzt.

Die Textmenge ist der maßgebliche Deckel, nicht die Dateigröße: gezählt werden Zeichen extrahierten Textes, umgerechnet mit 1.000 Zeichen je Seite Geschäftstext. Die ZIP-Grenze wird beim Upload technisch erzwungen. Der für die Feinabstimmung erzeugte Datensatz hat unabhängig von der hochgeladenen Menge immer dieselbe Größe; der Umfang Ihrer Unterlagen ist also die Grenze, nicht der Grund für die Tarifwahl.

Die Angabe zum Grafikspeicher deckt nur die Modellgewichte. Für jedes gleichzeitig laufende Gespräch kommt ein Zwischenspeicher hinzu, der mit der Länge des Verlaufs wächst. Wie Sie den Bedarf für Ihre Nutzerzahl selbst ausrechnen, steht in [Welche Hardware ein KI-Modell im Haus braucht](#).

Angabe	Wert
Vertragsschluss	mit Zugang der Auftragsbestätigung per E-Mail
Auftragsverarbeitungsvertrag	geschlossen spätestens vor dem Daten-Upload
Anzahlung	keine
Rechnung	mit Beginn der Verarbeitung, voller Modellpreis, netto, zahlbar in 14 Tagen
Bearbeitungsdauer	Tarif M meist wenige Tage, Tarif L rund eine Woche; Termine sind Zirka-Angaben, sofern nicht ausdrücklich verbindlich zugesagt
Bereitstellung	Aktivierungscode per E-Mail an die im Konto hinterlegte Adresse
Download-Fenster	30 Tage ab Bereitstellung
Löschung nach Bereitstellung	nach Ihrer Empfangsbestätigung, spätestens 30 Tage nach Bereitstellung

Angabe	Wert
Löschung ohne Bereitstellung	spätestens 180 Tage ab Auftragsbeginn
Löschnachweis	automatisch, sobald die Löschung erfolgt ist
Meldung eines Datenschutzvorfalls	unverzüglich, spätestens 48 Stunden nach Kenntniserlangung
Gewährleistungsfrist	12 Monate ab Bereitstellung
Prüfrechte	einmal jährlich und bei konkretem Anlass; Vor-Ort-Prüfung mit 14 Tagen Vorlauf
Wechsel eines Subauftragsverarbeiters	Ankündigung vorab in Textform, Widerspruch aus wichtigem Grund binnen 14 Tagen
Lieferumfang Modell	Gewichte des feingetunten Modells, Eigentumszertifikat mit SHA-256-Prüfwert, Apache-2.0-Lizenztext und NOTICE-Dateien
Lieferumfang Software	Einrichtungsprogramm für Windows, Linux und macOS, Bedienoberfläche AnythingLLM für den Browserzugang im Firmennetz
Lieferumfang Unterlagen	Prüfwerkzeug für die Kennzeichnung der Ausgaben, Kurzanleitung zur Prüfung, diese Leistungsbeschreibung

3. Ablauf mit Zeitachse

Der Ablauf ist für beide Tarife gleich. Die Zeitachse beginnt mit der Bestellung und endet mit dem Löschnachweis; der einzige Schritt, der die Dauer nennenswert beeinflusst, ist Ihr Upload, weil die Verarbeitung erst danach startet.

Internetzugang brauchen Sie nur für Anmeldung, Aktivierung und den einmaligen Download. Der spätere Betrieb läuft vollständig im Firmennetz. Den Bearbeitungsstand sehen Sie währenddessen in Ihrem Konto, aufgeschlüsselt nach Prüfung der Unterlagen, Erstellung der Trainingsdaten, Feinabstimmung und Bereitstellung.

Schritt	Was Sie tun	Was IonKon tut
Bestellung	Tarif wählen, Unternehmen und Anwendungsfall beschreiben, AGB zustimmen, Unternehmereigenschaft bestätigen	Bestellung entgegennehmen
Vertragsschluss	Auftragsbestätigung entgegennehmen	Auftragsbestätigung per E-Mail senden, Bestellportal freischalten
Auftragsverarbeitung	Auftragsverarbeitungsvertrag schließen	Vertrag mit den technischen und organisatorischen Maßnahmen als Anlage bereitstellen
Upload	Dokumente als eine ZIP-Datei im Portal hochladen	kurzlebiges, vorsigniertes Upload-Formular ausstellen, Archiv serverseitig prüfen
Verarbeitungsbeginn	Rechnung entgegennehmen, netto, 14 Tage	Rechnung stellen, für diesen Auftrag eine eigene GPU-Instanz erzeugen
Aufbereitung	Fortschritt im Konto verfolgen	Text extrahieren, personenbezogene Angaben entfernen, Trainingsbeispiele erzeugen
Feinabstimmung	keine Mitwirkung nötig	Modell anlernen, Qualität prüfen, Paket verschlüsselt ablegen, Instanz und Datenträger zerstören
Bereitstellung	Aktivierungscode aus der E-Mail entnehmen	Aktivierungscode versenden, Paket 30 Tage zum Download bereithalten

Schritt	Was Sie tun	Was IonKon tut
Einrichtung	Einrichtungsprogramm laden, mit Konto und Code anmelden, Modell auf eigener Hardware einrichten	Anmeldung, Aktivierung und Download bereitstellen
Empfangsbestätigung	Bereitstellung über den zugesandten Link bestätigen	bei ausbleibender Bestätigung zweimal erinnern
Löschung	Löschnachweis entgegennehmen	hochgeladene Dokumente, Trainingsdaten und eigene Modellkopien löschen, Nachweis erteilen
Betrieb	Modell im eigenen Netz betreiben, offline	keine weitere Verarbeitung

4. Was Sie selbst stellen

Die folgenden Punkte liegen beim Kunden. Sie sind Mitwirkungspflichten, nicht Empfehlungen: ohne sie kann die Leistung nicht erbracht oder das Ergebnis nicht in Betrieb genommen werden.

- **Hardware:** ein Server oder Rechner mit unterstützter NVIDIA-Grafikkarte und aktuellem Treiber. Karten der Blackwell-Generation nutzen den schnelleren FP4-Betrieb, alle anderen unterstützten Karten laufen automatisch im BF16-Betrieb, was den Speicherbedarf aus Abschnitt 2 vervierfacht.
- **Software auf dem Zielsystem:** Docker mit Compose v2 und dem NVIDIA Container Toolkit. Das Einrichtungsprogramm prüft beides und bietet bei fehlendem Docker eine Installation auf Knopfdruck an.
- **Netz:** das Firmennetz für den Browserzugang Ihres Teams sowie Internetzugang für Anmeldung, Aktivierung und den einmaligen Download. Auf macOS zusätzlich ein SSH-Zugang zu einem separaten Windows- oder Linux-Server mit passender Grafikkarte, weil Mac-Rechner keine NVIDIA-Karte haben.

- **Personen:** eine Person mit Administrationsrechten auf dem Zielsystem, ein IonKon-Konto mit E-Mail und Passwort sowie Zugriff auf die hinterlegte E-Mail-Adresse, an die der Aktivierungscode geht. Für die Auswahl der Dokumente eine fachlich zuständige Person.
- **Dokumente:** eine ZIP-Datei mit Formaten, die eine Textebene haben, also Word, PowerPoint, Excel mit Textinhalten, textbasierte PDF, Markdown, HTML und TXT. Gescannte PDF ohne Textebene, Fotos, CAD-Zeichnungen, Video und Audio liefern keinen Text und verbrauchen nur Platz im Archiv. Ordnerstruktur und Dateinamen spielen keine Rolle.
- **Rechte an den Inhalten:** die Zusicherung, zur Nutzung der hochgeladenen Inhalte für die Feinabstimmung berechtigt zu sein, und dass sie keine rechtswidrigen Materialien enthalten. Besondere Kategorien personenbezogener Daten nach Art. 9 DSGVO sind vertraglich ausgeschlossen.
- **Verantwortung nach der Auslieferung:** Aufbewahrung und Sicherung der heruntergeladenen Kopie. Sie bleibt von jeder Löschung bei IonKon unberührt und liegt allein bei Ihnen.

5. Was Sie entscheiden

Vier Entscheidungen trägt der Kunde, zwei davon vor der Bestellung. Sie bestimmen Preis, Zeitpunkt der Rechnung, Inhalt des Modells und den zulässigen Einsatzrahmen.

- **Tarif:** M oder L. Entscheidend ist nicht allein der Umfang der Unterlagen, sondern die Aufgabe. M deckt ein klar umrissenes Gebiet mit einheitlicher Sprache ab, L mehrere Gebiete, uneinheitliche oder widersprüchliche Quellen, mehrere Sprachen und Aufgaben über mehrere Schritte. Diese Fähigkeit entsteht mit der Modellgröße und lässt sich nicht durch mehr Trainingsdaten ersetzen.
- **Zeitpunkt:** Die Rechnung geht mit Beginn der Verarbeitung heraus, und die Verarbeitung beginnt nach Ihrem Upload. Sie steuern damit sowohl den Rechnungszeitpunkt als auch den Beginn des 30-Tage-Fensters für den Download.
- **Umfang und Auswahl der Dokumente:** Maßstab ist, was das Modell wissen soll, nicht was gerade greifbar ist. Alte Versionsstände, Dubletten und reine Zahlenkolonnen ohne erklärenden Text verbrauchen Ihr Kontingent ohne Nutzen.
- **Zweckbestimmung:** internes Assistenzsystem auf Basis eigener Unternehmensdokumente. Ein Einsatz für nach der KI-Verordnung verbotene Praktiken oder als Hochrisiko-System nach Anhang III bedarf der vorherigen Abstimmung mit IonKon.

- Oberfläche: die mitgelieferte Oberfläche, in der der Hinweis auf das KI-System bereits dauerhaft sichtbar eingerichtet ist, oder eine eigene Oberfläche, in der Sie diesen Hinweis selbst umsetzen.
- Optionale Angaben bei der Bestellung: ein Name, unter dem das Modell auftritt, und eine Tonalität, die den Stil aller Antworten prägt. Ohne Angabe tritt es neutral als Assistent Ihres Unternehmens auf.

6. Rechte am Modell, Vertrag, Zahlung

Mit vollständiger Zahlung gehen sämtliche übertragbaren Rechte am feingetunten Modell einschließlich der Gewichte unwiderruflich auf den Kunden über. IonKon behält keine Nutzungs-, Verwertungs- oder Rückrufrechte. Bis zur vollständigen Zahlung besteht ein einfaches, widerrufliches Nutzungsrecht; ein Zahlungsverzug berührt die Bereitstellung eines bereits fertiggestellten Modells nicht.

Der Kunde darf das Modell zeitlich, räumlich und inhaltlich unbeschränkt nutzen, verändern, vervielfältigen und wirtschaftlich verwerten, auch weiterverkaufen. Zu beachten sind allein die Apache-2.0-Bedingungen des Basismodells, im Wesentlichen der Lizenzhinweis; Lizenztext und NOTICE-Dateien liegen der Bereitstellung bei. Als Nachweis dient das Eigentumszertifikat mit dem SHA-256-Prüfwert der Gewichte.

Das Angebot richtet sich ausschließlich an Unternehmer im Sinne des § 14 BGB, an juristische Personen des öffentlichen Rechts und an öffentlich-rechtliche Sondervermögen, nicht an Verbraucher. Es gilt deutsches Recht unter Ausschluss des UN-Kaufrechts, Gerichtsstand ist der Sitz der IonKon GmbH. Änderungen und Ergänzungen bedürfen der Textform. Modell und zugehörige Software können dem europäischen und US-amerikanischen Exportkontrollrecht unterliegen; für die Einhaltung bei Weitergabe oder Wiederausfuhr ist der Kunde verantwortlich.

Eine Anzahlung wird nicht erhoben. Mit Beginn der Verarbeitung stellt IonKon eine Rechnung über den vollen Modellpreis, netto zuzüglich gesetzlicher Umsatzsteuer, zahlbar innerhalb von 14 Tagen. Bei Verzug gelten die gesetzlichen Verzugszinsen nach § 288 Abs. 2 BGB. Aufrechnen kann der Kunde nur mit unbestrittenen oder rechtskräftig festgestellten Forderungen.

Das Modell ist als entgeltlich erworbenes immaterielles Wirtschaftsgut grundsätzlich aktivierungs- und abschreibungsfähig; anzusetzen sind die Anschaffungskosten, also der

gezahlte Preis. Die Hardware bleibt ein davon getrenntes Wirtschaftsgut. Die konkrete bilanzielle Behandlung stimmen Sie mit Ihrer Steuerberatung ab; die Gesamtrechnung über drei Jahre steht in [Was ein eigenes KI-Modell wirklich kostet](#).

7. Datenschutz und Löschfristen

Für die hochgeladenen Dokumente ist der Kunde Verantwortlicher und IonKon Auftragsverarbeiter nach Art. 28 DSGVO; für die Stammdaten des Kundenkontos ist IonKon eigener Verantwortlicher. Der Upload geht über ein kurzlebiges, vorsigniertes Formular direkt in den verschlüsselten Objektspeicher der STRATO GmbH in Deutschland und berührt die Webserver nicht. Die inhaltliche Verarbeitung läuft auf einer eigens für den Auftrag erzeugten und danach gelöschten GPU-Instanz der DataCrunch Oy, Cloud-Marke Verda, in Finnland oder der Akenes SA, Marke Exoscale, in Frankfurt, Zagreb oder Zürich; welche davon, richtet sich nach der verfügbaren Kapazität. Die BuchhaltungsButler GmbH in Deutschland erhält für Rechnungsstellung und Zahlungsabgleich ausschließlich Stammdaten und keine Trainingsdokumente. Alle vier verarbeiten in der Europäischen Union oder der Schweiz; ein externer KI-Anbieter erhält zu keinem Zeitpunkt Zugriff auf Kundendokumente.

Es gelten zwei Löschfristen. Kommt es zur Bereitstellung, werden hochgeladene Dokumente, daraus erzeugte Trainingsdaten und die bei IonKon verbliebenen Modellkopien nach Ihrer Empfangsbestätigung gelöscht, spätestens 30 Tage nach der Bereitstellung. Kommt es nicht zur Bereitstellung, etwa weil ein Auftrag technisch scheitert, erfolgt die Löschung spätestens 180 Tage ab Auftragsbeginn. Ist gelöscht, wird automatisch ein Löschnachweis erteilt. Eine bereits heruntergeladene lokale Kopie bleibt unberührt. Bestell- und Rechnungsdaten werden im Rahmen der handels- und steuerrechtlichen Aufbewahrungsfristen nach § 257 HGB und § 147 AO aufbewahrt.

Vor der Erzeugung der Trainingsbeispiele werden personenbezogene Angaben automatisch aus dem extrahierten Text entfernt. Das verringert das Risiko, dass solche Angaben im fertigen Modell verbleiben, beseitigt es aber nicht: Die eigene Folgenabschätzung von IonKon bewertet genau dieses Risiko als mittel bis hoch. Die Pflicht zur Datenschutz-Folgenabschätzung trifft den Verantwortlichen, also den Kunden; IonKon unterstützt nach Art. 28 Abs. 3 lit. f DSGVO und stellt die eigene Ausarbeitung als Grundlage bereit. Stationen, Auftragsverarbeiter und das verbleibende Risiko im Einzelnen: [Wohin Ihre Dokumente gehen, wenn ein Modell entsteht](#).

8. Rollen nach der KI-Verordnung

IonKon ist Anbieter des ausgelieferten Systems im Sinne der Verordnung (EU) 2024/1689 und erfüllt in dieser Rolle die Kennzeichnungspflicht aus Art. 50 Abs. 2: Jede Ausgabe über die Schnittstelle trägt maschinenlesbare, signierte Metadaten, und ein mitgeliefertes Werkzeug prüft die Signatur unabhängig von IonKon. Die LoRA-Feinabstimmung erreicht die Rechenzeitschwelle nicht, ab der eine Anpassung selbst zum Basismodell mit eigener Anbieterpflichtenkaskade würde. Die Zuordnung der Rollen hängt von den Umständen des Einzelfalls ab.

Der Kunde ist Betreiber. Auf ihn entfallen der Hinweis an interagierende Personen nach Art. 50 Abs. 1, der in der mitgelieferten Oberfläche bereits eingerichtet ist, die Einhaltung der Zweckbestimmung, die Förderung der KI-Kompetenz der Beschäftigten nach Art. 4 und die vorherige Abstimmung, falls ein verbotener oder ein Hochrisiko-Einsatz nach Anhang III geplant ist. Welche Pflichten für ein internes Assistenzsystem gerade nicht gelten und welche Fristen laufen: [KI-Verordnung: was für normale Firmen gilt und was nicht](#).

9. Grenzen der Leistung

Die folgenden Punkte sind nicht zugesagt und werden auch nicht zugesichert. Sie stehen hier, damit die Erwartung an das Ergebnis vor der Bestellung feststeht und nicht nach der Auslieferung verhandelt wird.

- Richtigkeit einzelner Ausgaben. KI-Modelle erzeugen Ausgaben maschinell auf statistischer Grundlage; eine bestimmte Richtigkeit, Vollständigkeit oder Eignung ist nicht geschuldet. Vor geschäftskritischer Verwendung sind Ausgaben zu prüfen.
- Freiheit von falschen Aussagen und von wörtlicher Wiedergabe von Trainingsinhalten. Beides lässt der Stand der Technik nicht zusagen; was organisatorisch dagegen wirkt, steht in [Was ein Modell aus Ihren Unterlagen lernt, und was nicht](#).
- Eine bestimmte Ergebnisgüte. Sie hängt maßgeblich von Umfang und Qualität der bereitgestellten Dokumente ab. Sind die Daten unbrauchbar, etwa bei beschädigten Dateien oder fehlender Textebene, erhalten Sie Gelegenheit zur Nachlieferung innerhalb angemessener Frist.
- Tagesaktuelles Wissen. Der Wissensstand ist der Stand der hochgeladenen Unterlagen. Für Einzelfakten, die sich häufig ändern, ist der Dokumentenabruf der mitgelieferten Oberfläche

vorgesehen, den Sie ohne neue Feinabstimmung selbst pflegen: [Dokumente durchsuchen oder Modell anlernen](#).

- Kennzeichnung außerhalb der Schnittstelle. Die Markierung hängt an der Antwort des Systems, nicht am Text. Wird der reine Text in eine E-Mail oder ein anderes System kopiert, geht sie mit dieser Kopie nicht mit.
- Betrieb auf zu kleiner Hardware. Passen die Gewichte nicht in den Grafikspeicher, weicht das System auf den Arbeitsspeicher aus und wird um Größenordnungen langsamer. Ein knapp zu kleines Gerät gibt es nicht.
- Ansprüche aus Verzögerung oder Güte. Nur wenn die Bereitstellung aus Gründen unterbleibt, die IonKon zu vertreten hat, entfällt der Vergütungsanspruch und bereits gezahlte Beträge werden erstattet.
- Rechtsberatung. Dieses Dokument beschreibt den Leistungsstand und ersetzt keine rechtliche Prüfung Ihres Einsatzfalls.

10. Stand und Gültigkeit

Stand: 12. August 2026. Die Kenndaten in Abschnitt 2 geben die Tarife und die Rechtstexte zu diesem Stand wieder. Preise sind freibleibend und verstehen sich netto zuzüglich gesetzlicher Umsatzsteuer.

Verbindlich sind die Allgemeinen Geschäftsbedingungen, der Auftragsverarbeitungsvertrag mit der Anlage zu den technischen und organisatorischen Maßnahmen sowie die Datenschutzerklärung. Bei Abweichungen gehen diese Dokumente vor, im Bereich des Datenschutzes der Auftragsverarbeitungsvertrag. Die Zustimmung zu den Geschäftsbedingungen wird mit Zeitpunkt und Textfassung gespeichert; eine erteilte Zustimmung bleibt mit der Fassung gültig, der zugestimmt wurde.

Dieses Dokument wird nachgezogen, sobald sich Tarife, Rechtstexte oder die technische Umsetzung ändern. Maßgeblich ist stets die auf ionkon.de veröffentlichte Fassung.

Quellen

[IonKon AGB](#)

[IonKon Auftragsverarbeitungsvertrag](#)

[IonKon Technische und organisatorische Maßnahmen](#)

[IonKon Datenschutzerklärung](#)

[IonKon Preise](#)

[IonKon KI-Verordnung](#)

[IonKon Einrichtungsprogramm](#)

[IonKon Modelle und Hardware](#)

[Welche Hardware ein KI-Modell im Haus braucht](#)

[Wohin Ihre Dokumente gehen, wenn ein Modell entsteht](#)

[Was ein Modell aus Ihren Unterlagen lernt, und was nicht](#)

[KI-Verordnung: was für normale Firmen gilt und was nicht](#)

[Was ein eigenes KI-Modell wirklich kostet](#)

[Dokumente durchsuchen oder Modell anlernen](#)

[Verordnung \(EU\) 2016/679 \(DSGVO\), EUR-Lex](#)

[Verordnung \(EU\) 2024/1689 \(KI-Verordnung\), EUR-Lex](#)

[Hu et al., LoRA: Low-Rank Adaptation of Large Language Models, 2021](#)