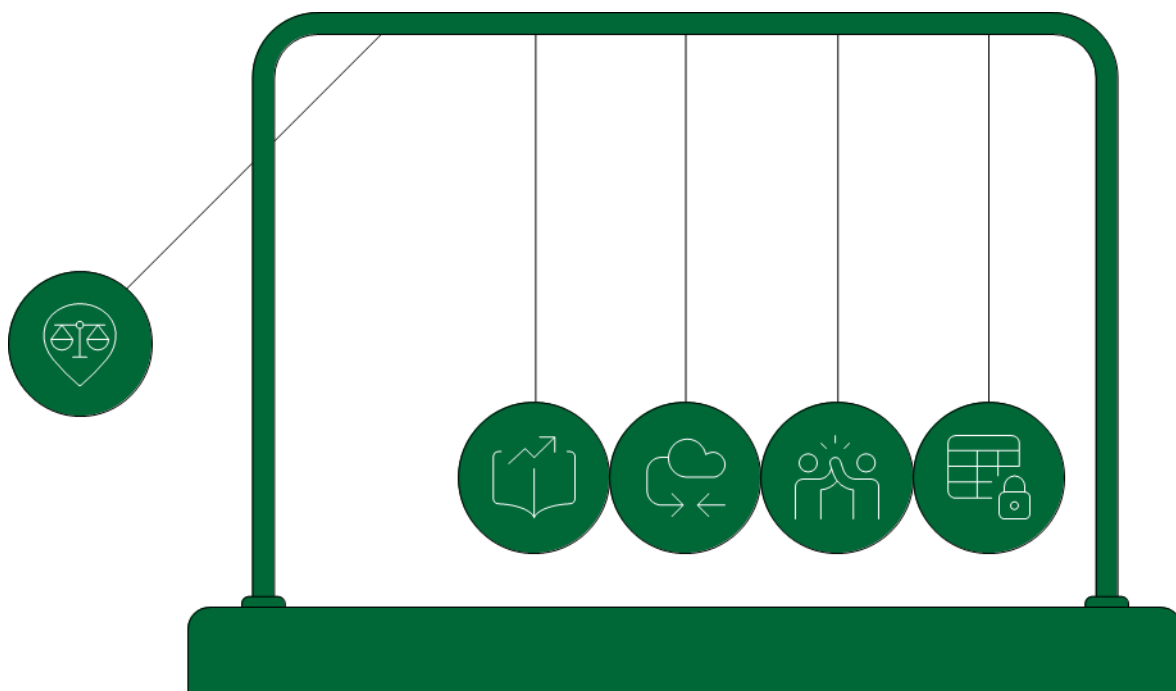


„AI-SourceKit“ -Einkauf von KI

Herausforderungen bei der KI-Beschaffung / Was bedeutet KI-Kompetenz für den Einkauf im Rahmen der Softwarebeschaffung?

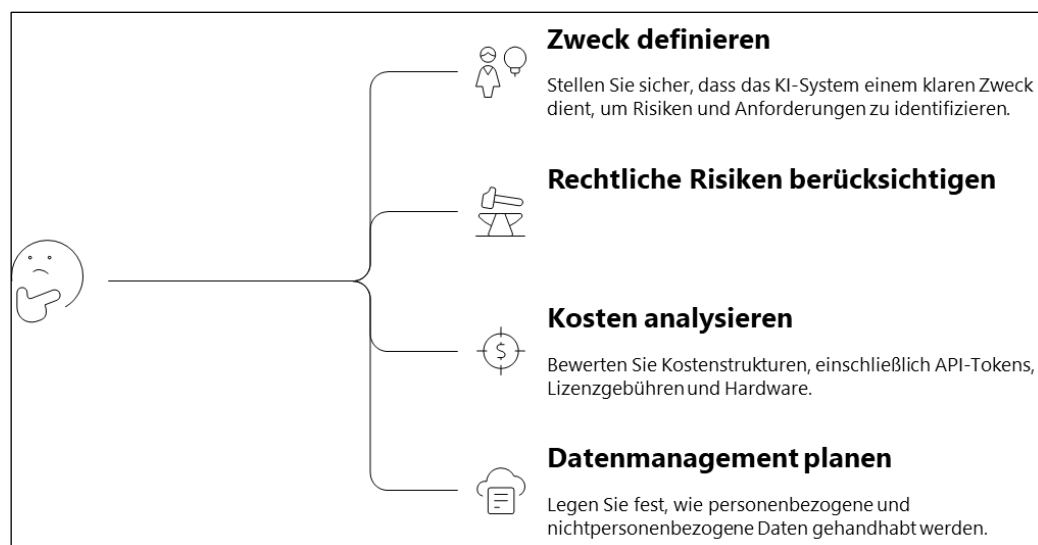


1 Einleitung

Die voranschreitende Entwicklung von KI-Systemen bringt auch für deren Beschaffung neue Herausforderungen mit. Die reine IT-Einkaufsaufgabe wird zur multidisziplinären Governance-Herausforderung. Wer KI beschafft, ohne die gesetzlichen Leitplanken und ökonomischen Hebel zu kennen, riskiert nicht nur horrenden Kosten, sondern auch die rechtliche Handlungsfähigkeit des Unternehmens.

Nach dem Motto: „**Wir führen einfach mal ein KI-System ein.**“, ist wenig zuführend. KI-Systeme umfassen heute ein breites Spektrum, was von Sprachmodellen (LLMs) für Texterstellung und Analyse über Bildgenerierung bis hin zu komplexen Multi-Agenten-Systemen, die mehrere KI-Instanzen orchestrieren reichen. Ausgangspunkt sollte immer der spezifische Zweck sein, für den das KI-System eingesetzt werden soll. Daran sind die jeweils spezifischen Risiken, aber auch technischen Anforderungen zu berücksichtigen.

Neben dem AI Act¹ können auch sektorspezifische Regelungen, wie etwa im Automotive-Bereich die Typengenehmigungsverordnung oder die General Safety Regulation, im Auswahlprozess relevant sein. Aber auch an DSGVO und allgemeines Vertragsrecht ist aus der Sicht der Beschaffung zu beachten.



Im Folgenden werden wir uns anschauen, auf welche Aspekte Beschaffungsteams bei der Auswahl eines KI-Systems achten sollten. Es geht um vertragliche Regelungen, die KI betreffen, aber auch um Daten und Informationsschutz. Hinsichtlich der Kosten sind bei API-basierten KI-Systemen sogenannte Tokens von Relevanz. Bei anderen Modellen sind Lizenzkosten, Implementierungsaufwand sowie Hardware-Anforderungen zu berücksichtigen.

2 AI Literacy: KI-Kompetenz als Fundament der Beschaffung - mehr als „nice to have“

Der AI Act sieht in Art. 4 die Verpflichtung für Unternehmen und Organisationen vor, ein angemessenes Maß an KI-Kompetenz (AI Literacy: siehe auch: <https://varyfy.de/ki-kompetenz/>) sicherzustellen. Diese Pflicht zu KI-Kompetenzvermittlung betrifft Beschäftigte und sonstige Personen, die mit dem Betrieb oder der Nutzung von KI-Systemen befasst sind – insbesondere

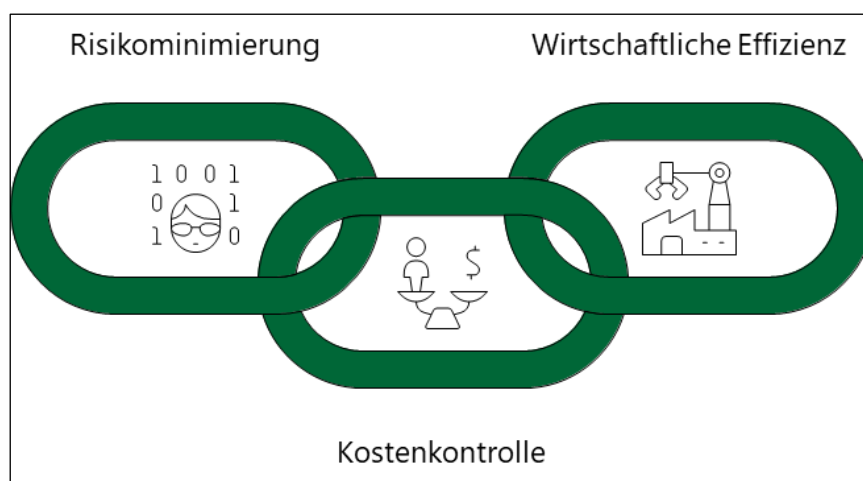
Mitarbeitende, die für die Auswahl, Beschaffung und Implementierung von KI-Lösungen verantwortlich sind.

Beschaffungsteams stehen dabei vor einer besonderen Herausforderung. Sie müssen die spezifischen Chancen, aber auch die Risiken von KI-Systemen vorab erkennen und bewerten. Neben dem AI Act sind ggf. weitere einschlägige Rechtsvorschriften zu beachten. Je nach Einsatzzweck kann hier die Datenschutz-Grundverordnung (DSGVO), das NIS-2-Umsetzungsgesetz bei kritischen Infrastrukturen sowie sektorspezifische Regelungen relevant werden. Ein technisches Verständnis ermöglicht es, die richtigen Compliance-Anforderungen (CE-Kennzeichnung, Konformitätsbewertung, etc.) zu identifizieren und auch eine entsprechende Qualitätsbewertung vornehmen zu können.

- Unterscheidung von KI-Architekturen: regelbasierte Systeme, Machine Learning, Large Language Models (LLMs) und Agentensysteme
- Verständnis von Deployment-Optionen: SaaS, PaaS, On-Premise, Hybrid
- Grundwissen Token-Economics: Wie API-basierte Abrechnungsmodelle funktionieren
- Datenflusskompetenz: Wo fließen Daten hin? Werden sie für Training verwendet?
- Tool-Landschaft: Kenntnis gängiger KI-Plattformen, Automatisierungslösungen und Enterprise-Dienste

Es geht darum, Risiken frühzeitig zu erkennen und zu minimieren, am besten bereits vor der Einführung eines KI-Systems. Gleichzeitig sollen Kostenkontrolle und wirtschaftliche Effizienz gewährleistet sein.

Systematische KI-Kompetenzschulungen können dabei die Grundlage für erfolgreiche und AI Act-konforme Procurement-Prozesse bilden.



3 Anforderungsanalyse und Auswahlprozess

Was wollen wir eigentlich? Vorab die interne Bedarfsermittlung

Die Auswahl eines KI-Systems erfordert besondere Sorgfalt. Vorab sollte immer erst das zu lösende Problem bzw. der zugrundeliegende Zweck klar definiert werden. Bevor eine neue Technologie gesucht wird, ist es hilfreich, zunächst den spezifische Anwendungsfall zu beschreiben. Dies sollte durch den Fachbereich geschehen, der das KI-System einsetzen will.

Maßgeblich ist der konkrete Anwendungsfall, für den das KI-System eingesetzt werden soll.

3.1 Zweck und Anwendungsfall definieren

Zunächst sollte gefragt werden, ob das Problem, was mit dem anzuschaffenden KI-System gelöst werden soll, auch nur mit KI lösbar ist. Möglicherweise kann eine einfache Datenbank, Suchfunktion oder regelbasierte Automatisierung schon ausreichend sein.

Typische KI-Anwendungsfälle in der Beschaffungspraxis umfassen:

- Textbasierte Assistenten für Kundenservice, interne Wissensabfrage oder Dokumentenanalyse
- Automatisierte Workflows für Datenverarbeitung, Reporting oder Compliance-Prüfungen (z. B. über Plattformen wie n8n oder Make.com)
- Bildgenerierung und visuelle KI für Marketing, Produktdesign oder Qualitätskontrolle
- RAG-Systeme (Retrieval Augmented Generation) für Vertragsanalyse oder Wissensdatenbanken
- KI-gestützte Analyse und Prognose in Produktion, Logistik oder Finanzen

Auch sollte bereits frühzeitig geklärt werden, ob das KI-System neben dem Einsatz im eigenen Unternehmen auch an Kunden weitervermarktet werden soll und auch kann. Dies betrifft vor allem der Unterlizenzierung.

Es sollte bereits im Zeitpunkt der Beschaffung darauf geachtet werden, dass das Recht zur Unterlizenzierung besteht, sofern ein KI-System in die Produkte für Kunden eingebaut werden soll.

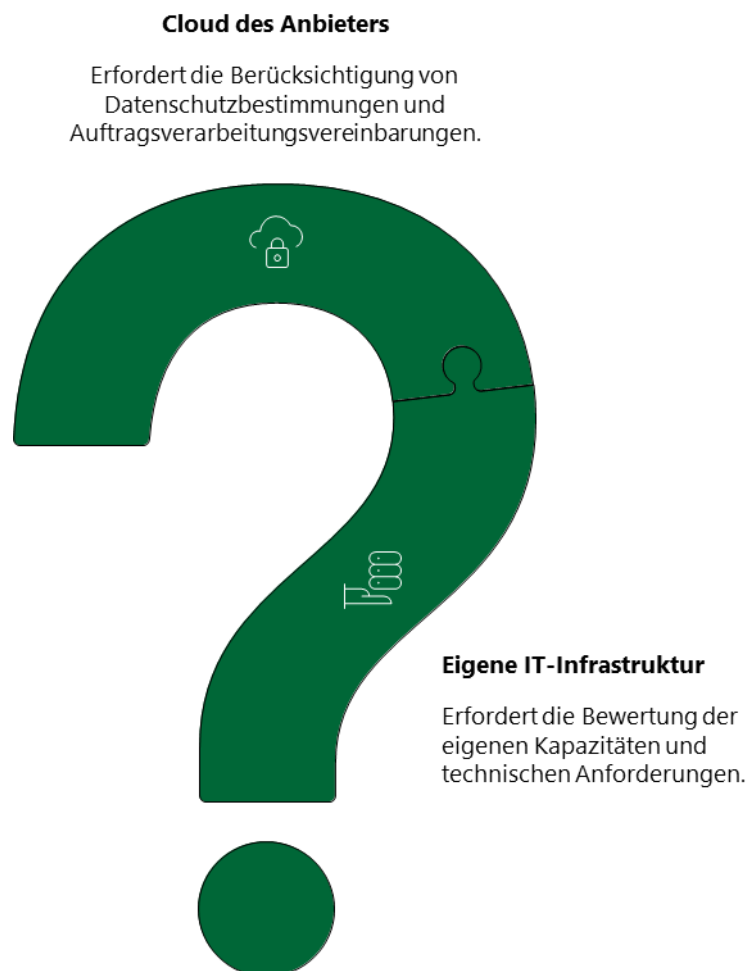
3.2 Ziele, Anforderungen und potenzielle Risiken identifizieren

Ziele, Anforderungen und potenzielle Risiken im Zusammenhang mit dem KI-System sind bereits im Auswahlprozess zu identifizieren. Dies gilt insbesondere auch im Hinblick auf die Auswirkungen auf betroffene Personen und deren Datenverarbeitung. Diese Bewertung bildet die Grundlage für die spätere Risikoklassifizierung gemäß AI Act und die Einhaltung der DSGVO-Anforderungen.

3.3 Betriebsmodell klären: Cloud oder eigene Infrastruktur?

Auch die Wahl des Betriebsmodells hat weitreichende Konsequenzen für Datenschutz, Kosten und Kontrolle und sollte bereits im Rahmen der Beschaffung berücksichtigt werden.

Wird das KI-System über die Cloud des Anbieters genutzt, als On-Premise-Lösung auf eigener Infrastruktur betrieben oder als hybrides Modell implementiert?



Soll das KI-System über eigene Ressourcen betrieben werden, sind auch die eigenen Kapazitäten zu berücksichtigen. Dies bedeutet, dass die technischen Anforderungen im Vorfeld zu klären sind, um die Eignung des KI-Systems für die eigene IT-Landschaft zu ermitteln.

Bei der Nutzung der Cloud des Anbieters ergeben sich weitere Fragestellungen, beispielsweise im Hinblick auf die DSGVO, wenn nämlich personenbezogene Daten mit dem KI-System verarbeitet werden ([siehe Ziffer 6 „Datenschutz“](#)).

SaaS vs. On-Premise

Cloud / SaaS	Self-Hosted / On-Premise
Schnelle Einführung, geringer Startaufwand	Volle Datensouveränität, keine Drittlandtransfers

Automatische Updates und Skalierung	Kontrolle über Modellversionen und Updates
Abhängigkeit vom Anbieter (Vendor Lock-in)	Höherer initialer Aufwand für Setup und Betrieb
Daten auf externen Servern (oft US)	Daten verlassen das Unternehmen nicht
Laufende Abonnementkosten	Einmalige Implementierung plus optionaler Managed Service
Berufsrechtlich oft problematisch	Kompatibel mit Berufsgeheimnisschutz

4 Bewertung und Auswahl von KI-Systemen

Der europäische Gesetzgeber geht davon aus, dass ein Unternehmen, das ein KI-System einsetzt bzw. nutzt, bei dessen Auswahl die gebührende Sorgfalt hat walten lassen, die vernünftigerweise erwartet werden kann.

„Der Begriff der gebührenden Sorgfalt bezieht sich auf die Anstrengungen, die eine mit normaler Umsicht handelnde oder vernünftige Person unternimmt, um unter Berücksichtigung der Umstände Schaden von anderen abzuwenden.“

Mit anderen Worten besteht hier die Pflicht, ein KI-System vorab im Hinblick auf bestimmte Anforderungen, die z.B. der AI Act, sektorspezifische Regelungen, Produkthaftung und Produktsicherheit, aber auch die Datenschutz-Grundverordnung (DSGVO) vorsehen, zu prüfen.

4.1 Bewertungskriterien für KI-Systeme

Für die Bewertung sind bestimmte Kriterien immer heranzuziehen.

4.1.1 Nachvollziehbarkeit und Transparenz

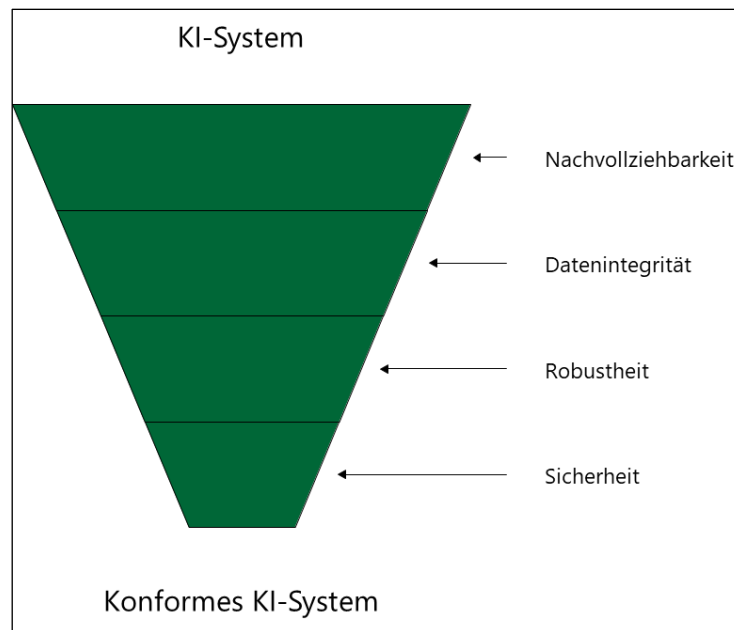
Dies betrifft zum einen die Nachvollziehbarkeit und Transparenz eines KI-Systems. Es geht hier vor allem um Erklärbarkeit und ob nachvollziehbar ist, wie das KI-System zu seinen Ergebnissen kommt. Nur so können die Beschäftigten und Kunden gemäß der gesetzlichen Informationspflichten aufgeklärt werden.

Wichtig ist aber auch, welche Daten für das Training, die Tests und die Validierung der KI genutzt wurden.

Wurden etwa personenbezogene Daten oder geschützte Daten zum Training verwendet, ist sicherzustellen, dass dies rechtmäßig erfolgt ist. Existiert zudem die Option, das Training zu korrigieren?

4.1.2 Robustheit und Sicherheit

Relevant sind auch die Robustheit des KI-Systems sowie dessen Schutz vor Angriffen. Wie werden Fehler kontrolliert und erkannt? Wie wird der Betreiber des KI-Systems darüber in Kenntnis gesetzt?



4.2 Due Diligence bei Hochrisiko-KI-Systemen

Insbesondere bei der Beschaffung von Hochrisiko-KI-Systemen² sollten beim Anbieter des Systems zumindest die folgenden Punkte vorab geklärt werden:

- **Registrierung:** Ist das Hochrisiko-KI-System in der von der EU-Kommission verwalteten Datenbank (Art. 49 AI Act) registriert?
- **Normenkonformität:** Hält das System alle einschlägigen Normen ein?
- **Konformitätsbewertung:** Wurde das System einer vorgeschriebenen Konformitätsbewertung gemäß Art. 43 AI Act unterzogen?
- **CE-Kennzeichnung:** Liegt eine CE-Kennzeichnung vor (Art. 49 Abs. 5 AI Act)?
- **Betriebsanleitung und Dokumentation:** Sind alle erforderlichen Informationen, Anleitungen und technischen Unterlagen vorhanden (Art. 13 AI Act)?
- **Transparenzpflichten:** Werden die Transparenzanforderungen erfüllt, insbesondere wenn natürliche Personen betroffen sind (Art. 13, Art. 52 AI Act)?
- **Monitoring und Meldepflichten:** Gibt es Prozesse zur Überwachung des Systems und zur Information der zuständigen Aufsichtsbehörden (Art. 61 ff. AI Act)?

4.3 Qualitätsindikator bei der KI-Beschaffung: ISO/IEC 42001

Auch eine ISO/IEC 42001-Zertifikat oder andere Zertifizierung kann bei der Auswahl eines KI-Anbieters ein wichtiger Qualitätsindikator sein. Die internationale Norm (veröffentlicht Dezember 2023) definiert Anforderungen an ein Managementsystem für Künstliche Intelligenz (AI Management System – AIMS).

4.3.1 Was bedeutet die Zertifizierung?

Ein nach ISO/IEC 42001 zertifiziertes Unternehmen hat ein strukturiertes Managementsystem für verantwortungsvollen KI-Einsatz etabliert:

- **Risiko- und Compliance-Management:** Systematische Identifikation und Steuerung von Risiken und gesetzlichen Anforderungen (z.B. DSGVO, AI Act)
- **Klare Verantwortlichkeiten:** Definierte Zuständigkeiten für den gesamten KI-Lebenszyklus
- **Transparenz:** Dokumentierte und nachvollziehbare Prozesse und Entscheidungen
- **Schulung:** Regelmäßige Sensibilisierung der Mitarbeitenden (AI Literacy gemäß Art. 4 AI Act)
- **Kontinuierliche Verbesserung:** Laufende Überprüfung und Anpassung des Managementsystems

4.3.2 Wichtige Einschränkung

Das Zertifikat bestätigt die Existenz eines KI-Managementsystems, nicht jedoch:

- Die technische Sicherheit einzelner KI-Produkte
- Die Rechtskonformität spezifischer KI-Systeme mit dem AI Act
- Die Konformitätsbewertung oder CE-Kennzeichnung bei Hochrisiko-KI

ISO/IEC 42001 ist ein Managementsystem-Standard (vergleichbar mit ISO 27001), kein Produktzertifikat. Bei der Beschaffung von Hochrisiko-KI-Systemen müssen zusätzlich die spezifischen Anforderungen des AI Act geprüft werden.

4.4 Qualitätsindikator bei der KI-Beschaffung: Woran erkennt man einen guten KI-Anbieter?

Bei komplexen KI-Projekten zeigt sich die Qualität eines Anbieters bereits in der Analysephase. Professionelle Partner gehen systematisch in drei Schritten vor:

- **Bestandsaufnahme (Discovery):**
Der Anbieter analysiert, wie Ihr Unternehmen heute arbeitet: Welche Daten fließen wohin? Wie nutzen Beschäftigte die vorhandenen Systeme? Wo entstehen Engpässe? Diese Erkenntnisse werden in übersichtlichen Prozessdiagrammen dargestellt, die zeigen, wo Automatisierung sinnvoll ist.
- **Potenzialanalyse (Transformation):**
Ihre Arbeitsabläufe werden in einzelne Schritte zerlegt. Dabei identifiziert der Anbieter die „Hebelpunkte“, also jene Bereiche, in denen KI den größten Nutzen bringt und den Aufwand am stärksten reduziert.
- **Lösungsdesign (Architektur):**
Aus der Analyse entsteht ein konkreter Umsetzungsplan. Dieser beschreibt genau, wie die

KI-Lösung funktioniert: Welche Aufgaben werden automatisiert, welche Datenquellen werden eingebunden und wie fügt sich alles in Ihre bestehende IT-Landschaft ein.

5 Vertragliche Absicherung

Bei der Beschaffung von KI-Systemen sind vertragliche Regelungen mit dem Anbieter entscheidend.

5.1 Leistungsbeschreibung und Zweckbestimmung

Ausgangspunkt für die vertragliche Betrachtung sollte immer die Leistungsbeschreibung und Zweckbestimmung des KI-Systems sein. Insbesondere die Leistungsbeschreibung gibt hier vor, wofür der Anbieter vertraglich einstehen will und wofür nicht.

Es sollte in jedem Fall auch geregelt werden, in welchem Rahmen das KI-System genutzt werden darf und welche Rechte das Unternehmen an den Ergebnissen hat, die durch das System produziert werden.

Eine Frage, die man sich auch stellen sollte, ist, ob das KI-System ohne größere Risiken im Hinblick auf Urheberrecht, Datenschutz, Produktsicherheitsgesetz und letztlich KI-Verordnung eingesetzt werden kann.

5.2 Besondere Anforderungen bei Hochrisiko-KI

Bei Hochrisiko-KI-Systemen (Kapitel III AI Act) müssen die gesetzlichen Verpflichtungen vertraglich festgeschrieben werden: Konformitätsbewertung (Art. 43), CE-Kennzeichnung (Art. 48), Registrierung (Art. 49), Post-Market-Monitoring (Art. 72), Meldepflichten (Art. 73).

Der Betreiber haftet für Verstöße gegen die KI-Verordnung, daher ist eine klare vertragliche Risikoverteilung wesentlich.

5.3 Vermeidung von Vendor Lock-in

Eine zu große Abhängigkeit vom Anbieter sollte vermieden werden. Vertraglich vereinbart werden sollten mindestens standardisierte Datenformate, Datenportabilität sowie Exit-Strategie mit Übergangsfristen.

5.4 Model Contractual Clauses for AI Procurement (MCC-AI)

Eine Orientierung für die Vertragsgestaltung bieten die Model Contractual Clauses for the public procurement of AI (MCC-AI), veröffentlicht von der EU Public Buyers Community (aktualisierte Version: 5. März 2025)³.

Es werden zwei Versionen zur Verfügung gestellt:

- **MCC-AI-High-Risk:** Für Hochrisiko-KI-Systeme (umfassende Klauseln zu Art. 6–51 AI Act)
- **MCC-AI-Light:** Für Nicht-Hochrisiko-KI mit Transparenzanforderungen

Wichtig ist dabei, dass die MCC-AI keine vollständigen Verträge darstellen, sondern als Anlage an bestehende Verträge (Kauf-/Werkvertrag, AVV) gedacht sind.

Nutzen für private Unternehmen

Obwohl primär für öffentliche Auftraggeber entwickelt, sind die MCC-AI auch für private Unternehmen nutzbar:

- Best-Practice-Standards für AI Act-konforme Beschaffung
- Risikoabsicherung und klare Haftungsverteilung
- Verhandlungsgrundlage mit KI-Anbietern
- Supply-Chain Due Diligence

Es besteht allerdings ein Anpassungsbedarf. Die Klauseln sind zum Teil spezifisch für den öffentlichen Sektor (z.B. Vergaberecht, öffentliche Transparenzpflichten) und müssen angepasst werden. Zudem kann deren Nutzung im privatrechtlichen Bereich unter die AGB-Kontrolle fallen.

6 Datenschutz

6.1 Einbindung des Datenschutzbeauftragten

Werden personenbezogene Daten mithilfe durch von KI-Systemen verarbeitet, sollte zwingend der betriebliche Datenschutzbeauftragte in Boot geholt werden.

6.2 Auftragsverarbeitung

Weiterhin muss mit dem Anbieter des Systems zwingend eine Vereinbarung zur Auftragsverarbeitung gemäß Art. 28 DSGVO abgeschlossen werden. Nur dadurch wird gewährleistet, dass der Anbieter personenbezogene Daten für die Zwecke des Auftraggebers und nicht für eigene Zwecke verarbeitet.

6.3 Drittlandsübermittlung

Daneben ist zu prüfen, wo die Datenverarbeitung stattfindet. Erfolgt diese außerhalb der EU oder des EWR, muss ein angemessenes Schutzniveau für die Datenverarbeitung gewährleistet sein. Dies kann zum Beispiel durch einen Angemessenheitsbeschluss der EU-Kommission gemäß Art. 45 DSGVO (z.B. UK, Schweiz, Japan) erfolgen oder durch EU-Standardvertragsklauseln gemäß Art. 46 Abs. 2 lit. c DSGVO.

6.4 Risikobewertung und Datenschutzfolgenabschätzung

Je nach Einsatzzweck kann es auch sein, dass für den Einsatz des zu beschaffenden KI-Systems eine sogenannte Datenschutz-Folgenabschätzung (DSFA) gemäß Art. 35 DSGVO durchgeführt werden muss. Dies bedeutet, dass hier die potenziellen Risiken, die von dem KI-System für Betroffenen ausgehen können, vorab zu identifizieren und zu dokumentieren sind. Dabei sind auch die Maßnahmen darzustellen, mit denen die Risiken minimiert werden können-

Z.B. muss eine DSFA beim Einsatz von künstlicher Intelligenz zur Verarbeitung personenbezogener Daten zur Steuerung der Interaktion mit den Betroffenen oder zur Bewertung persönlicher Aspekte der betroffenen Person durchgeführt werden. Typisches Einsatzfeld wäre hier z.B. der Kundensupport mittels künstlicher Intelligenz.

Beispiele:

„Ein Callcenter wertet automatisiert die Stimmungslage der Anrufer aus.

Ein Unternehmen setzt ein System ein, welches mit Kunden durch Konversation interagiert und für deren Beratung personenbezogene Daten durch eine künstliche Intelligenz verarbeitet werden.“⁴

6.5 DSGVO-Compliance bei KI: Trainingsdaten und Zweckbindung

Bei allen KI-Systemen, die personenbezogene Daten verarbeiten, stellt sich immer eine zentrale datenschutzrechtliche Frage:

Werden die eingegebenen Daten, insbesondere bei Large Language Models (LLMs), zu Trainingszwecken weiterverarbeitet, und ist dies mit der DSGVO vereinbar?

6.5.1 Zweckbindungsgrundsatz

Maßgeblich ist dabei insbesondere der sogenannte Zweckbindungsgrundsatz (Art. 5 Abs. 1 lit. b DSGVO):

„Personenbezogene Daten müssen für festgelegte, eindeutige und legitime Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden.“

6.5.2 Praktische Konsequenzen

Werden personenbezogene Daten z.B. zur Nutzung eines KI-Chatbots erhoben (Zweck: Beantwortung einer Kundenanfrage), ist eine Weiterverarbeitung zum Training des KI-Modells grundsätzlich nicht mit diesem ursprünglichen Zweck vereinbar, es sei denn, es liegt eine eigene Rechtsgrundlage vor.

6.6 Praxisbeispiel: Azure OpenAI und Abuse Monitoring – Datenschutzrechtliche Risiken

Ein weiteres praxisrelevantes Beispiel für datenschutzrechtliche Herausforderungen bei der KI-Beschaffung ist die Nutzung von Azure OpenAI von Microsoft.

6.6.1 Abuse Monitoring und Auftragsverarbeitung

Microsoft prüft bei der Nutzung von Azure OpenAI Chat-Inhalte zu Zwecken des Abuse Monitoring (Missbrauchserkennung). Diese Überprüfung kann unter Umständen nicht über eine Auftragsverarbeitung nach Art. 28 DSGVO abgedeckt sein. Das bedeutet:

Microsoft könnte eigenständigen Zugriff auf personenbezogene Daten in den Chats erhalten, für den das nutzende Unternehmen möglicherweise keine ausreichende Rechtsgrundlage hat.

6.6.2 Deaktivierungsmöglichkeiten und Prüfpflichten

Unternehmen können die Deaktivierung des Abuse Monitoring beantragen. Die Transparenz dieses Prozesses und dessen Umsetzung sollten von Unternehmen im Rahmen ihrer datenschutzrechtlichen Prüfpflichten überprüft werden.

6.6.3 Transparenz- und Informationspflichten

Nach den Anforderungen der DSGVO muss der Anbieter klar und transparent darlegen, zu welchen Zwecken und auf welcher Rechtsgrundlage personenbezogene Daten verarbeitet werden.

6.6.4 Informationspflichten der nutzenden Unternehmen

Unternehmen, die Azure OpenAI nutzen, sind verpflichtet, ihre Nutzenden über die Datenverarbeitung zu informieren und sicherzustellen, dass eine zulässige Rechtsgrundlage für die Übermittlung und Verarbeitung der Daten durch Microsoft besteht. Fehlt diese, besteht das Risiko eines Datenschutzverstoßes nach DSGVO.

6.6.5 Besondere Anforderungen bei Hochrisiko-KI-Systemen

Gerade bei Hochrisiko-KI-Systemen fordert der AI Act eine lückenlose Dokumentation und Nachvollziehbarkeit der Datenflüsse sowie die Möglichkeit, Betroffenenrechte effektiv wahrzunehmen.

6.6.6 Prüfung und Schutzmaßnahmen vor dem Einsatz

Unternehmen sollten daher vor dem Einsatz von Azure OpenAI oder ähnlichen Diensten sorgfältig prüfen, ob die vertraglichen und technischen Maßnahmen die Anforderungen des AI Act und der DSGVO erfüllen. Falls nicht, ist von einer Nutzung abzuraten, oder es müssen zusätzliche Schutzmaßnahmen implementiert werden (z.B. ergänzende Vereinbarungen mit Microsoft, technische Maßnahmen zur Datenminimierung, organisatorische Maßnahmen wie Schulungen und Nutzungsrichtlinien).

7 Kostenanalyse und Wirtschaftlichkeit

7.1 Token-basierte Preismodelle verstehen

Bei der Beschaffung von KI-Systemen ist das Kostenmanagement ein entscheidender Faktor. Die meisten API-basierten KI-Dienste rechnen auf Basis von Tokens ab: Einheiten, die etwa 0,75 Wörtern im Deutschen entsprechen.

Faustregel: 1 Token entspricht etwa 0,75 Wörtern im Deutschen (1.000 Tokens entsprechen ca. 750 Wörtern). Die meisten Anbieter unterscheiden zwischen Input-Tokens (Ihre Anfrage) und Output-Tokens (die KI-Antwort), wobei Output-Tokens typischerweise drei- bis fünfmal teurer sind.

7.2 Enormes Preisgefälle zwischen KI-Modellen

Die Token-Preise variieren erheblich. Zwischen Budget- und Premium-Modellen liegt ein Faktor von 20 bis 100:

Szenario: Tägliche Verarbeitung von 100 Millionen Tokens

Differenz: Über \$544.000 pro Jahr zwischen günstigstem und teuerstem Modell.

Wer KI-Leistungen einkauft, muss die Preisentwicklung der Anbieter kontinuierlich beobachten. Preismodelle, ob tokenbasiert, als Flatrate oder nutzungsabhängig, unterscheiden sich erheblich und ändern sich teilweise mehrfach pro Jahr. Das hat direkte Auswirkungen auf die Budgetplanung und die wirtschaftliche Bewertung laufender Verträge.

7.3 Die fünf häufigsten Kostenfallen bei der KI-Beschaffung

In nahezu jedem KI-Projekt treten dieselben Kostenrisiken auf. Wer sie kennt, kann sie vermeiden:

7.3.1 Falle 1: Pilotkosten sind nicht gleich Produktionskosten.

Die Pilotphase kostet fast nichts, aber der Vollbetrieb kann das 50- bis 100-fache betragen. Ohne realistische Hochrechnung droht eine Budgetüberschreitung.

7.3.2 Falle 2: Output-Tokens sind drei- bis fünfmal teurer als Input.

Viele Kalkulationen basieren nur auf Input-Preisen. Bei generativen Anwendungen (Texterstellung, Code-Generierung) dominieren die Output-Kosten.

7.3.3 Falle 3: Agenten-Workflows multiplizieren den Token-Verbrauch.

Wenn mehrere KI-Agenten zusammenarbeiten (z.B. Recherche, Analyse, Qualitätsprüfung), kann ein einzelner Auftrag 2.000 bis 5.000 Tokens pro Agenten-Schritt verbrauchen.

7.3.4 Falle 4: Kosten im Blick behalten

Die Preise für KI-Modelle verändern sich rasant, daher ist es ratsam, diese im Blick zu behalten und ggf. alte Verträge zu kündigen. OpenAI hat die Token-Kosten für GPT-4 zwischen März 2023 und August 2024 um bis zu 89 % gesenkt⁵; Anthropic reduzierte den Preis seines Flaggschiff-Modells Claude Opus Ende 2025 um rund 67 %⁶.

7.3.5 Falle 5: Versteckte Kosten.

Die API-Kosten machen typischerweise nur 5 bis 15 Prozent der Gesamtkosten aus. Integration, Schulung, Wartung und Compliance verursachen den Großteil der Kosten.

7.4 Proficiency-Kurve: Warum sich KI-Investitionen erst über die Zeit auszahlen

Ein häufiger Fehler bei der Wirtschaftlichkeitsberechnung liegt darin, dass Einkaufsteams den vollen ROI bereits im ersten Quartal erwarten. Die Praxis zeigt jedoch eine mehrstufige Proficiency-Kurve, d.h. die KI-Kompetenz wächst schrittweise mit zunehmender Erfahrung.

7.4.1 Adoption (Jahr 1)

Das KI-System befindet sich in der Lernphase. Mitarbeitende gewöhnen sich an neue Workflows. Erste Prozesse werden optimiert, aber der Produktivitätsgewinn ist moderat.

Zeiteinsparung: ca. 3 Stunden pro Woche pro aktivem Nutzer.

7.4.2 Optimierung (Jahr 2)

Kontinuierliches Training verbessert das Verhalten der KI-Agenten. Mitarbeitende nutzen das System routinierter. Prompts und Workflows sind auf die spezifischen Anforderungen abgestimmt.

Zeiteinsparung: Steigerung auf 5 Stunden pro Woche.

7.4.3 Mastery (Jahr 3)

Die Agenten sind vollständig auf die Anforderungen trainiert, kennen Branchenspezifika und Kundenpräferenzen. Prozesse sind durchgängig optimiert.

Zeiteinsparung: bis zu 7 Stunden pro Woche.

7.5 ROI-Beispiel: Beratungsunternehmen mit 20 Mitarbeitenden

7.5.1 Ausgangssituation

- 15 aktive KI-Nutzer im Unternehmen
- Interner Stundensatz: 120 EUR
- Betrachtungszeitraum: 3 Jahre

7.5.2 Produktivitätsgewinn im Zeitverlauf

Zeit- raum	Ø Zeitersparnis pro Nutzer	Eingesparte Stunden (15 Nutzer)
Jahr 1	3 Std./Woche	ca. 2.340 Std.
Jahr 2	5 Std./Woche	ca. 3.900 Std.

Zeit- raum	Ø Zeitersparnis pro Nutzer	Eingesparte Stunden (15 Nutzer)
Jahr 3	6 Std./Woche	ca. 4.680 Std.
Gesamt		ca. 10.920 Std.

Annahme: Steigende Effizienz durch wachsende KI-Kompetenz der Mitarbeitenden (Proficiency-Kurve)

Wirtschaftlichkeitsrechnung:

Monetärer Nutzen (3 Jahre):

10.920 Stunden × 120 EUR = 1.310.400 EUR

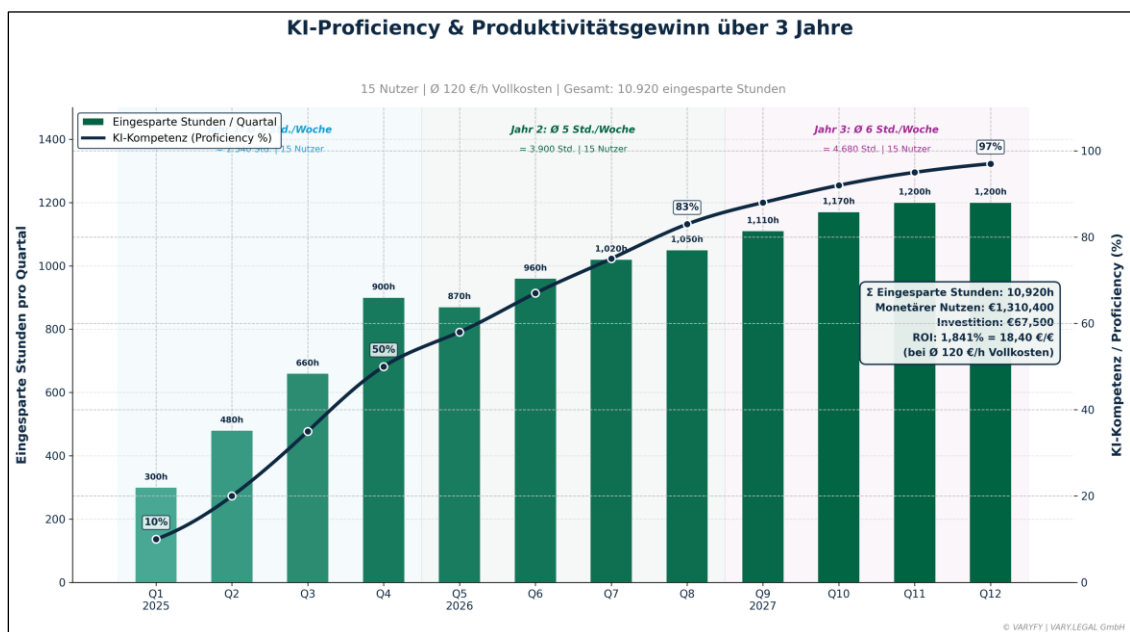
Gesamtinvestition (3 Jahre):

Setup + Implementierung + Managed Service = 67.500 EUR

Return on Investment (ROI):

$(1.310.400 \text{ EUR} - 67.500 \text{ EUR}) / 67.500 \text{ EUR} \times 100 = 1.841\%$

→ Jeder investierte Euro bringt über drei Jahre einen Gegenwert von **18,40 EUR**.



Dieser ROI basiert auf einer schrittweisen Kompetenzentwicklung. Die Zeitersparnis steigt kontinuierlich, da Mitarbeitende lernen, die KI effektiver einzusetzen. Eine KI-Einführung ist daher als langfristiger Kompetenzaufbau zu verstehen, nicht als sofortige Effizienzsteigerung.

7.6 Strategische Entscheidungskriterien und Kostenoptimierung

Die Modellwahl sollte auf den Qualitätsanforderungen des Anwendungsfalls, dem Datenvolumen und der Skalierung, Datenschutzanforderungen und dem verfügbaren Budget basieren.

7.7 Bewährte Strategien zur Kostenoptimierung

- Datenminimierung: Nur notwendige Daten senden
- Modell-Kaskadierung: Einfache Anfragen an Budget-Modell, komplexe an Premium-Modell. Einsparung: 40 bis 70 Prozent
- Prompt Caching: Häufig gestellte Anfragen cachen, einige Anbieter bieten natives Caching mit erheblichen Rabatten
- Batch-Verarbeitung: Anfragen bündeln statt einzeln senden
- Volumenrabatte verhandeln: Ab ca. \$5.000 monatlich sind Rabatte von 10 bis 30 Prozent üblich
- Abstraktionsschicht nutzen: Middleware-Lösungen (z.B. n8n, LiteLLM oder ähnliche Plattformen) für einfachen Modellwechsel ohne Code-Änderung
- Empfehlung: Kalkulieren Sie immer die TCO (Total Cost of Ownership) über mindestens drei Jahre, nicht nur die reinen API-Kosten. Die Proficiency-Kurve zeigt, dass der volle wirtschaftliche Nutzen erst mittelfristig eintritt.

7.8 Kostenrahmen: Individualprojekt versus Template-Ansatz

Für Einkäufer ist es wichtig, die grundlegend verschiedenen Preiskategorien bei KI-Implementierungen zu kennen:

Klassisches Individualprojekt	Template-basierter Ansatz
Individuelle Entwicklung von Grund auf	Wiederverwendbare Vorlage mit kundenspezifischer Konfiguration
Investition typisch ab 50.000 EUR	Einstieg typisch ab 15.000 EUR
Projektdauer 6 Monate und länger	Implementierung in 4 bis 8 Wochen
Oft Cloud-Abhängigkeit über SaaS	Self-Hosted-Option mit voller Datensouveränität möglich
Vendor Lock-in durch proprietäre Architektur	Open-Source-Basis mit voller Kontrolle

Template-basierte Ansätze senken die Einstiegshürde erheblich, weil bewährte Architekturen wiederverwendet werden. Die kundenspezifische Konfiguration (Prompts, Workflows, RAG-Anbindungen, API-Integrationen) gehört dem Kunden. Das Template selbst bleibt in der Regel IP des Anbieters und wird unter Nutzungslizenz bereitgestellt.

8 **Fazit**

Erfolgreiche KI-Beschaffung erfordert heute weit mehr als technische Expertise. Sie verlangt ein ganzheitliches Verständnis rechtlicher, ethischer und datenschutzrechtlicher Anforderungen sowie deren konsequente Umsetzung in allen Beschaffungsphasen.

- Unternehmen müssen bei der KI-Beschaffung die gebührende Sorgfalt nach AI Act und DSGVO walten lassen.
- Hochrisiko-KI-Systeme erfordern Registrierung, Konformitätsbewertung, CE-Kennzeichnung und umfassende Dokumentation.
- Training von KI-Modellen mit personenbezogenen Daten und Abuse Monitoring durch Anbieter bedürfen einer datenschutzrechtlichen Rechtsgrundlage.
- Die Model Contractual Clauses for AI (MCC-AI) bieten praktische Orientierung für vertragliche Regelungen.
- ISO/IEC 42001-Zertifikate belegen ein KI-Managementsystem, nicht aber die Rechtskonformität einzelner Produkte

© [2026] [VARY.LEGAL GmbH]. Dieses Werk ist lizenziert unter einer Creative Commons

Namensnennung - Nicht-kommerziell - Keine Bearbeitungen 4.0 International Lizenz.

Um eine Kopie dieser Lizenz zu sehen, besuchen Sie <http://creativecommons.org/licenses/by-nc-nd/4>

Disclaimer

Dieses Whitepaper dient ausschließlich Informationszwecken und stellt keine Rechts-, Steuer-, Anlage- oder sonstige Beratung dar. Die Informationen werden ohne Gewähr für Vollständigkeit, Richtigkeit oder Aktualität bereitgestellt. Die Nutzung der Informationen in diesem Dokument erfolgt auf eigenes Risiko. Es wird empfohlen, vor Entscheidungen fachkundigen Rat einzuholen.

VARY.FY (VARY.LEGAL GmbH) berät und unterstützt u.a. zur KI-Strategie und KI-Kompetenz. Wir übersetzen EU-Regulierung (AI Act, DSGVO) in praxistaugliche Lösungen für den Mittelstand: von der KI-Kompetenzschulung über die technische Implementierung bis zur rechtlichen Absicherung.

¹ Der am 1. August 2024 in Kraft getreten ist und dessen Regelungen schrittweise anwendbar werden

² Ein KI-System ist gemäß Artikel 6 AI Act eine Hochrisiko-KI, wenn es entweder (1) Teil eines regulierten Produkts nach Anhang I ist und eine Konformitätsbewertung durch Dritte erfordert, oder (2) in einem sensiblen Anwendungsbereich gemäß Anhang III eingesetzt werden soll. Zu Anhang III gehören Bereiche wie Biometrie, Personalmanagement, Bildung, Strafverfolgung, kritische Infrastrukturen und Zugang zu wesentlichen Diensten. Für diese Hochrisikosysteme gelten strenge Anforderungen nach den Artikeln 8 bis 15 der KI-Verordnung. Siehe hierzu z.B.: https://www.bundesnetzagentur.de/DE/Fachthemen/Digital/KI/9_Hochrisiko/start.html

³ <https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/updated-eu-ai-model-contractual-clauses>

⁴ „DSFA-Blacklist“ der Datenschutzkonferenz (DSK) für den nicht-öffentlichen Bereich, https://www.lfd.niedersachsen.de/startseite/datenschutzrecht/ds_gvo/liste_von_verarbeitungsvorgaengen_nach_art_35_abs_4_ds_gvo/muss-listen-zur-datenschutz-folgenabschätzung-179663.html

⁵ Vgl. OpenAI Pricing, <https://openai.com/api/pricing/> (abgerufen am Datum einfügen); Andrew Ng, X/Twitter, 29.08.2024; Neowin, „OpenAI cuts GPT-4o prices“, 06.08.2024.

⁶ Vgl. Anthropic Pricing, <https://platform.claude.com/docs/en/about-claude/pricing> (abgerufen am Datum einfügen); MindStudio, „What is Anthropic Claude 4.5“, 2025.