ChatGPT

EU AI Act: Anforderungen beim selbstgehosteten KI-Stack im Unternehmen

Unternehmen, die eigene KI-Lösungen (z. B. AI-Lab mit Ollama und Open-Source-LLMs wie LLaMA, Mistral) **on-premises** betreiben, müssen auch den **EU AI Act** (EU-KI-Verordnung, Stand 2024) beachten. Im Folgenden werden die wichtigsten Pflichten und Einstufungen erläutert – mit praxisnahen Empfehlungen und Auswirkungen auf die Systemarchitektur. Die **Zielgruppe** sind IT-Verantwortliche und Datenschutzbeauftragte, die sicherstellen wollen, dass der lokale KI-Stack rechtskonform und verantwortungsvoll betrieben wird.

Lokaler KI-Betrieb (On-Premises): Unternehmenspflichten und Rollen

Eigenbetrieb vs. Cloud: Der AI Act macht grundsätzlich **keinen Unterschied**, ob ein KI-System in der Cloud oder lokal im eigenen Rechenzentrum läuft. Entscheidend sind vielmehr Rolle und Risiko des Systems. Unternehmen, die KI-Systeme in eigener Verantwortung einsetzen (d. h. für betriebliche Zwecke nutzen), gelten als **"Betreiber"** im Sinne der Verordnung [1]. Das gilt auch, wenn z. B. ein Open-Source-Modell auf eigener Hardware oder fremder Infrastruktur (IaaS) betrieben wird [1]. Der Betrieb **intern** im Unternehmen befreit also nicht von den Vorgaben – im Gegenteil: Man muss selbst für die Einhaltung der Anforderungen sorgen, da keine Cloud-Anbieter als "Puffer" fungieren.

Provider vs. User: Wichtig ist zu klären, ob das Unternehmen als **Anbieter (Provider)** oder nur als **Betreiber (User)** auftritt. Ein *Anbieter* im Sinne des AI Act ist, wer ein KI-System **entwickelt oder entwickeln lässt und es unter eigenem Namen in Verkehr bringt oder in Betrieb nimmt** [2]. Ein Unternehmen, das intern ein eigenes KI-Tool aufsetzt (z. B. durch Kombination von Open-Source-Komponenten), kann somit selbst **Anbieter** dieses Systems sein – insbesondere wenn kein Dritter als Hersteller auftritt. Wird hingegen ein fertiges KI-System eines Drittanbieters (oder ein vortrainiertes Modell **unverändert**) genutzt, ist man primär **Betreiber** (und der ursprüngliche Anbieter bleibt verantwortlich). In der Praxis verschwimmen die Rollen oft, vor allem bei Open-Source: Da hier kein klassischer Hersteller mit CE-Kennzeichnung existiert, muss derjenige, der das System **in Betrieb nimmt**, viele Pflichten selbst übernehmen.

Allgemeine Pflichten beim KI-Einsatz: Unabhängig von der Risikostufe gelten einige **Grundanforderungen** für KI-Systeme. Unternehmen sollten insbesondere:

- **Verbotene Praktiken vermeiden:** Der AI Act verbietet bestimmte KI-Anwendungen komplett (Art. 5), z. B. manipulative **Beeinflussung** von Menschen durch KI (unterschwellige Manipulation), ausnutzende **Überwachung** verletzlicher Personen, **Social Scoring** oder ähnliche inakzeptable Risiken [3] [4]. Im Architektur-Design muss man daher sicherstellen, dass der KI-Stack nicht (versehentlich) für solche Zwecke eingesetzt wird. *Beispiel:* In einem AI-Lab sollten keine Funktionen implementiert werden, die Mitarbeiter heimlich bewerten oder Menschen sozial einstufen – das wäre unzulässig.
- **Transparenz wahren:** Wenn Nutzer (Mitarbeiter, Kunden) mit dem KI-System interagieren oder KI-generierte Inhalte erhalten, besteht eine **Informationspflicht**. Konkret verlangt Art. 52 AI Act die **Kennzeichnung von KI-generierten Inhalten** (z. B. Hinweise bei durch KI erzeugten Texten

oder Bildern) [5] [6]. Ebenso muss nach Art. 13(2) AI Act offengelegt werden, **wenn jemand mit einem KI-System interagiert** (z. B. ein Chatbot), damit klar ist, dass es sich nicht um einen Menschen handelt [5] [6]. *Praxis:* In einem internen Chatbot-Interface (wie OpenWebUI im AI-Lab) sollte deutlich sichtbar stehen, dass die Antworten von einer KI generiert werden. Generierte Dokumente oder Berichte könnten mit einem Fußnote-Hinweis "Automatisch mit KI erstellt" versehen werden, um Transparenz herzustellen.

- **Datenschutz beachten:** Beim Betrieb von KI-Systemen, die **personenbezogene Daten** verarbeiten, gelten weiterhin die Anforderungen der DSGVO. Ein on-premises Betrieb hat hier Vorteile (Daten verlassen nicht das Unternehmen), erfordert aber dennoch Maßnahmen wie Zugriffsbeschränkungen, Pseudonymisierung wo möglich und eine **Datenverarbeitungsübersicht**. Falls der KI-Stack z. B. Dokumente mit Mitarbeiterdaten analysiert, sollte eine **DSGVO-konforme Rechtsgrundlage** vorliegen und ggf. eine **Datenschutz-Folgenabschätzung (DSFA)** durchgeführt werden. Diese DSFA kann idealerweise mit der im AI Act geforderten **grundrechtlichen Folgenabschätzung** bei Hochrisiko-KI kombiniert werden (siehe unten).

Architektur-Tipp: Beim Design des **AI-Lab-Stacks** sollte man Compliance-"by-design" umsetzen. Dazu gehören Audit-Logs, Zugriffskontrollen, klare Datenflüsse und **keine versteckten Outputs**. Beispielsweise kann ein zentrales Logging (z. B. in PostgreSQL/PGVector) eingebaut werden, um nachzuverfolgen, welche Eingaben und Ausgaben die KI liefert – wichtig für Transparenz und zur Untersuchung etwaiger Fehlentscheidungen. Ebenso empfiehlt sich eine **Nutzer-Authentifizierung und Rollenverteilung** (wie im AI-Lab optional per JWT-Token gelöst), damit nur autorisierte Personen KI-Funktionen nutzen und sensible Ergebnisse sehen dürfen. Durch on-premises Hosting behält das Unternehmen die volle Datenhoheit, was die Einhaltung von Datenschutz- und Geheimhaltungsregeln erleichtert.

Hochrisiko-KI: Einstufung und Konsequenzen nach dem AI Act

Der AI Act teilt KI-Systeme anhand ihres Gefährdungspotenzials in Risikoklassen ein: **minimales Risiko**, **begrenztes (transparenzpflichtiges) Risiko**, **hohes Risiko** und **inakzeptables Risiko** [7] [8]. *Hochrisiko-KI-Systeme* unterliegen dabei den strengsten Auflagen. **Wann gilt ein On-Prem-KI-System als hochriskant?** Maßgeblich ist der *Einsatzzweck*: Anhang III der Verordnung listet konkrete Anwendungsbereiche, die als hochriskant gelten [6]. Beispiele für solche Hochrisiko-Einsatzfelder sind:

- **Personalwesen:** KI-Systeme zur **Personalauswahl, Beförderungsentscheidungen oder Kündigungen** gelten als Hochrisiko [9]. Ebenso erfasst sind Tools, die **Aufgaben auf Basis persönlicher Merkmale zuteilen** oder das **Verhalten und die Leistung von Mitarbeitern automatisiert überwachen** [10] [11]. (Dies ist relevant, falls z. B. der KI-Stack genutzt wird, um Bewerbungen zu filtern oder Mitarbeiterdaten auszuwerten.)
- **Bildung und Gesundheitswesen:** Systeme, die über den **Zugang zu Bildung** entscheiden (z. B. automatisierte Prüfungssysteme) oder **medizinische Diagnosen/Triage** stellen, fallen in Hochrisiko.
- **Kritische Infrastruktur und Verkehr:** KI zur Steuerung des Stromnetzes, der Verkehrsleittechnik etc. wäre hochriskant, was im Kontext eines Dokumenten- und Workflow-KI-Stacks selten zutrifft.
- **Recht und Verwaltung:** KI, die in **justiziellen Verfahren** oder zur **Behördenentscheidung** (z. B. Asylentscheidungen) eingesetzt wird, ist hochriskant.
- **Finanz- und Versicherungsentscheidungen:** Kreditwürdigkeitsprüfungen, Scoring-Systeme oder Versicherungs-Tarifierungen durch KI gelten ebenfalls als hochriskant, da sie über wichtige Lebensbereiche entscheiden.

Wichtig: **Nicht jede Nutzung von LLMs ist hochriskant.** Ein genereller Dokumenten-Chatbot oder Workflow-Assistent im Unternehmen zählt in der Regel **nicht** zu Anhang-III-Szenarien, solange er nicht für o. g. kritische Zwecke genutzt wird. Viele typische Anwendungsfälle (z. B. internes Wissensmanagement, Code-Generierung, Marketing-Textentwürfe) dürften maximal als begrenztes Risiko gelten. Allerdings kann sich der Risikostatus ändern, wenn das System in einem sensiblen Prozess verankert wird. *Beispiel:* Nutzt man das AI-Lab, um **automatisiert Bewerbungen** zu ranken oder Entscheidungen über Kundenkredite vorzubereiten, könnte daraus ein Hochrisiko-KI-System werden – mit entsprechenden Pflichten.

Pflichten bei Hochrisiko-Systemen: Wird der KI-Stack (oder ein Teil davon) als hochriskant eingestuft, greifen umfangreiche **Compliance-Anforderungen** des AI Act. Diese ähneln dem bekannten CE-Konformitätsprozess bei Maschinen und umfassen u. a. [12] [13]:

- **Risikomanagement:** Der Anbieter muss ein **Risikomanagement-System** etablieren (Art. 9), d. h. systematisch Risiken des KI-Modells analysieren, testen und mitigieren. In der Architektur sollte es z. B. Möglichkeiten zum Simulieren von Fehlverhalten geben (Testdaten, Sandbox) und ggf. eine Fail-Safe-Strategie, falls das Modell unerwartet agiert.
- **Daten- und Algorithmus-Governance:** Nach Art. 10 sind **Trainings-, Validierungs- und Testdaten** sorgfältig auszuwählen (relevant, repräsentativ, kein diskriminierender Bias). Für ein selbst gehostetes Modell heißt das: Man muss dokumentieren, auf welchen Daten es basiert und bei Feintuning mit Firmendaten sicherstellen, dass keine unerlaubten Daten (z. B. geschützte Merkmale ohne Rechtsgrundlage) einfließen [14]. Gegebenenfalls sind technische Vorkehrungen zu treffen, wenn besondere Kategorien personenbezogener Daten verarbeitet werden (z. B. Einwilligung einholen oder Daten anonymisieren).
- **Technische Dokumentation:** Anbieter Hochrisiko-KI müssen ausführliche **Dokumentationen (Technische Doku und Konformitätserklärung)** bereitstellen (Art. 11, 16). Für einen internen KI-Stack sollte man eine **Modell-Dokumentation ("Model Card")** führen: Welche Modelle und Versionen sind im Einsatz? Mit welchen Parametern wurden sie ggfs. nachtrainiert? Welche Genauigkeit/Zuverlässigkeit wurde festgestellt? Solche Informationen sollten in der Architektur zentral abgelegt werden (z. B. als Teil der AI-Lab-Dokumentation).
- **Transparenz und Informationen:** Auch Hochrisiko-KI müssen Nutzerinformationen bieten. Dazu gehören **Nutzungsanleitungen, Beschreibung des Zwecks, Leistung und Limitierungen** des Systems (Art. 13). Im Unternehmen heißt das, die Mitarbeiter, die mit dem System arbeiten, **müssen geschult werden** und Zugriff auf verständliche Anleitungen haben [15]. Beispielsweise sollte klar dokumentiert sein, dass der KI-Assistent keine garantierte Wahrheit liefert, biasbehaftet sein kann etc.
- **Human Oversight (menschliche Aufsicht):** Gemäß Art. 14 ist sicherzustellen, dass angemessene **menschliche Kontrolle** über das KI-System besteht. In der Praxis muss das Unternehmen Personen benennen, die die KI-Nutzung überwachen und eingreifen können. *Architektur-Umsetzung:* Workflows sollten so gestaltet sein, dass bei kritischen Entscheidungen ein **"Human-in-the-loop"** eingebunden ist. Etwa könnte das AI-Lab für wichtige Empfehlungen (z. B. Einstellungsentscheidungen) einen **Prüfschritt durch einen Menschen** vorsehen, bevor eine finale Aktion erfolgt. Außerdem sollte die Oberfläche dem Nutzer ermöglichen, KI-Entscheidungen zu **überstimmen oder das System abzuschalten** (eine Art Notabschaltung bei Fehlverhalten).
- **Genauigkeit, Robustheit, Cybersicherheit:** Hochrisiko-KI-Systeme müssen **robust und sicher** sein (Art. 15). Für die eigene Architektur bedeutet das: harte Testing-Routinen (z. B. Adversarial-Tests), regelmäßige **Updates/Patches** für Sicherheitslücken und ggf. Redundanzen einplanen. Beispiel: Wenn das KI-Modell in einer Workflow-Automatisierung ausfällt oder Unsinn liefert, sollte es Mechanismen geben, die einen Schaden verhindern (Fallback auf Standardprozess).

- **Automatisierte Aufzeichnungen (Logging):** Betreiber hochriskanter KI sind verpflichtet, **Protokolldaten aufzuzeichnen und vorzuhalten** (Art. 12, 29). In Art. 26 wird Betreibern empfohlen, automatisch erzeugte Logs **mindestens 6 Monate aufzubewahren** [16]. Daher muss der KI-Stack Logging-Komponenten enthalten, die Eingaben, Outputs und wichtige Entscheidungswege mitschneiden. Ein Praxis-Tipp ist, die Logs so zu speichern, dass sie vor Manipulation geschützt sind (z. B. in einer revisionssicheren Datenbank). Diese Logs dienen der späteren Nachvollziehbarkeit, z. B. falls ein Ergebnis angefochten oder ein **Incident Reporting** (Meldepflicht schwerer Vorfälle an Behörden) nötig wird.
- **Grundrechte-Folgenabschätzung:** Zusätzlich zu den technischen Maßnahmen verlangt Art. 29 i.V.m. Art. 26 vom Betreiber eine **Bewertung der Auswirkungen auf Grundrechte**, bevor ein Hochrisiko-KI-System eingesetzt wird [17] [18]. Dies ähnelt einer Datenschutz-Folgenabschätzung und sollte interdisziplinär erfolgen (Einbezug Datenschutz, Compliance, ggf. Betriebsrat). Vor Implementierung eines Hochrisiko-Anwendungsfalls mit AI-Lab wäre also zu prüfen: Welche Risiken für Betroffene (Mitarbeiter, Kunden) bestehen? Wie stellen wir sicher, dass keine Diskriminierung, Verletzung der Privatsphäre oder Gesundheit erfolgt? Die Ergebnisse fließen dann in die Konfiguration des Systems ein (z. B. Einschränkung bestimmter Funktionen, zusätzliche Kontrollen).

Hinweis zu Fristen: Die KI-Verordnung ist im August 2024 in Kraft getreten, aber viele Pflichten (v. a. für Hochrisiko-Systeme) werden **phasenweise bis 2026/2027** wirksam [19] [20]. Konkret gelten die Anforderungen für Hochrisiko-KI aus Art. 6 und Anhang III erst ab **August 2027** verpflichtend [20]. Unternehmen haben also eine Übergangsfrist, um ihre KI-Systeme anzupassen. Dennoch ist es **ratsam, schon jetzt bei neuen KI-Projekten die künftigen Vorgaben einzukalkulieren**, um böse Überraschungen zu vermeiden. Insbesondere bei Architekturentscheidungen, die lange wirken (z. B. Wahl einer bestimmten KI-Plattform oder Modellfamilie), sollte man „KI-Act-ready" planen.

Open-Source-Modelle: Auswahl, Transparenz und Governance

Open Source im AI Act: Der EU AI Act erkennt den Wert von Open-Source-KI ausdrücklich an und enthält bestimmte **Ausnahmen** für Open-Source-Anbieter [21]. **ABER:** Diese Ausnahmen sind **sehr eingeschränkt**. Zwar heißt es in Art. 2 AI Act, dass **frei verfügbare Open-Source-KI-Systeme** grundsätzlich nicht reguliert werden [22]. Sobald jedoch ein solches System **in der EU in Verkehr gebracht oder in Betrieb genommen wird** – insbesondere als Hochrisiko-System oder in direktem Kontakt mit Personen – greifen die Vorschriften wieder [22]. Für ein Unternehmen, das Open-Source-Modelle einsetzt, bedeutet das: Man kann sich **nicht darauf verlassen**, dass Open-Source-Nutzung automatisch vom AI Act ausgenommen ist. Die Regulierung „holt einen ein", sobald das Modell praktisch verwendet wird, vor allem bei höheren Risiken oder Nutzerinteraktionen.

Lizenz und Definition: Wichtig bei der Modellauswahl ist die **Lizenz**. Der AI Act definiert ziemlich genau, was *"Open Source"* bedeutet: Nämlich KI-Modelle, die unter einer **freien und quelloffenen Lizenz** veröffentlicht sind, die **Nutzung, Weitergabe, Veränderung und Untersuchung** erlaubt – inklusive Offenlegung der **Modellparameter, -architektur und -nutzung** [23]. Zudem muss der ursprüngliche Anbieter genannt werden und vergleichbare Lizenzbedingungen weitergegeben werden [24]. **Praxisbeispiel:** Ein Modell wie **Mistral 7B** (Apache-2.0-Lizenz) erfüllt diese Kriterien – Gewichte frei verfügbar, kommerzielle Nutzung erlaubt. Hingegen war das ursprüngliche **LLaMA** von Meta nur unter eingeschränkter Lizenz erhältlich (nur Forschungszwecke) und damit *nicht* "Open Source" im Sinne des AI Act [25]. Unternehmen sollten bevorzugt Modelle wählen, die tatsächlich offen lizenziert sind, um rechtliche Klarheit zu haben, welche Pflichten den ursprünglichen Entwickler treffen und welche bei einem selbst liegen.

Transparenzpflichten bei offenen Modellen: Offene Modelle kommen häufig ohne umfangreiche Dokumentation oder Garantie vom Anbieter. Dennoch fordert der AI Act **Transparenz auch bei der Nutzung solcher Modelle**. Insbesondere **Generative KI** (ein typischer Anwendungsfall von LLMs) unterliegt bestimmten Informationspflichten: Anbieter von allgemeinen KI-Modellen müssen z. B. eine **Zusammenfassung der verwendeten Trainingsdaten veröffentlichen** (Art. 53(1)(d)) und eine **Urheberrechts-Compliance-Policy** haben [25] [26]. Bei Open-Source-Modellen, die *nicht* unter "systemisches Risiko" fallen, gibt es Erleichterungen – sie sind von manchen Dokumentationspflichten ausgenommen [25] [27]. Allerdings greifen diese Ausnahmen für den **Endanwender** kaum: Wenn Ihr Unternehmen ein Open-Source-Modell in ein eigenes KI-System integriert, müssen Sie *als Systemanbieter* trotzdem die üblichen Nachweise erbringen. **Fazit:** Stellen Sie sicher, dass Sie über das gewählte Modell so viel wie möglich in Erfahrung bringen (Paper, Model Card, Community-Diskussionen). Dokumentieren Sie intern mindestens, **welche Trainingsdaten bekannt sind**, welche Qualität/Bias-Einschätzungen es gibt und welche Version/Weights genutzt werden. Diese Informationen sind nicht nur aus Compliance-Sicht wertvoll, sondern auch für die **Risikoabschätzung**: Ein unbekanntes Modell mit potenziell verzerrten Trainingsdaten könnte z. B. inadäquat für bestimmte Aufgaben (etwa HR-Entscheidungen) sein.

Governance und interne Kontrollen: Der Einsatz von Open-Source-LLMs erfordert **Governance-Maßnahmen**, da kein externer Anbieter für Sie die Verantwortung trägt. Praktische Empfehlungen:

- Richten Sie ein **KI-Gremium** oder zumindest klare Verantwortlichkeiten ein (z. B. benennt die IT-Abteilung einen "KI-Verantwortlichen"), der den Einsatz der Modelle überwacht.
- Führen Sie vor dem produktiven Einsatz eines neuen Modells einen **internen Audit/Review** durch: Entspricht das Modell unseren ethischen Standards? Liefert es in Tests verlässliche Ergebnisse? Gibt es bekannte Probleme (z. B. Halluzinationen, bestimmte Vorurteile in Antworten)? Dieser Review sollte dokumentiert werden.
- **Nutzungspolicy:** Definieren Sie schriftlich, wofür das Modell benutzt werden darf und wofür nicht. Insbesondere wenn das Modell **weiterlernen** kann (z. B. via Fine-Tuning oder durch Speicherung von Chat-Verläufen), sollten Regeln bestehen, welche Daten eingegeben werden dürfen. (Keine sensiblen personenbezogenen Daten in allgemeine LLM-Eingaben, falls nicht unbedingt nötig, etc.)
- **Monitoring:** Implementieren Sie Monitoring im KI-Stack, um die Performance und Auswirkungen laufend zu beobachten. Ein einfaches Beispiel ist ein Feedback-Mechanismus: Nutzer können Fehlantworten oder unerwünschtes Verhalten der KI melden. So kann man das Modell ggf. nachjustieren oder eingrenzen.
- **Notfallplan:** Überlegen Sie architektonisch, wie Sie reagieren, wenn sich herausstellt, dass ein verwendetes Open-Source-Modell **fehlerhaft oder unsicher** ist. (Etwa, wenn eine gravierende Sicherheitslücke bekannt wird oder regulatorische Änderungen das Modell plötzlich als riskant einstufen.) Ein guter Ansatz ist Modularität: In AI-Lab könnten Sie z. B. ein Modell austauschen (Container ersetzen) oder ein Feature deaktivieren, ohne das Gesamtsystem stillzulegen.

Zusammenarbeit mit der Community: Nutzen Sie die Stärke von Open Source! Halten Sie sich über Updates des Modells auf dem Laufenden – oft verbessern die Communities ihre Modelle schnell. Zudem können Sie eigene Verbesserungen zurückspielen. Beachten Sie aber: Wenn Sie ein Modell **selbst verändern und veröffentlichen**, könnten Sie rechtlich als (Mit-)Anbieter auftreten und müssten dann z. B. ebenfalls technische Dokumentation bereitstellen. Dieser Fall dürfte bei rein interner Nutzung selten relevant sein, ist aber im Hinterkopf zu behalten.

Eigenentwicklung, Modellanpassung oder reine Nutzung – was ändert sich?

Der EU AI Act differenziert klar nach dem **Grad der Kontrolle** über das KI-System. Daraus ergeben sich unterschiedliche Pflichten:

- **Eigenentwicklung (In-house-Modell):** Entwickelt ein Unternehmen ein KI-Modell von Grund auf selbst (oder trainiert ein Open-Source-Basismodell umfangreich mit eigenen Daten zu einem neuen Zweck), dann ist das Unternehmen der **Anbieter** dieses Systems. Alle oben genannten Anbieterpflichten – insbesondere bei Hochrisiko – liegen in der eigenen Verantwortung. In der Praxis bedeutet das, man muss z. B. die Konformitätsbewertung durchführen, ein CE-Kennzeichen anbringen (falls Hochrisiko) und das System ggf. bei der EU-Datenbank anmelden. Dies ist aufwändig, gibt einem aber auch volle Kontrolle. *Architekturimplikation:* Man sollte frühzeitig eine **Compliance-Dokumentation** mitführen, während man das System entwickelt (z. B. Architektur-Entscheidungen protokollieren, Trainingsdatenkatalog erstellen, Evaluationsberichte sammeln). Diese Unterlagen bilden die Basis für die gesetzlich geforderte technische Dokumentation [18] [28].
- **Anpassung fremder Modelle (Fine-Tuning, Modifikationen):** Hier wird ein bestehendes KI-Modell eines Dritten verändert – sei es durch Nachtraining (z. B. LLM mit Firmen-Daten feinjustieren) oder durch technische **Weiterentwicklung** (z. B. Einbau eines Modells in eine größere KI-Lösung mit neuem Zweck). Der AI Act bestimmt, dass **wesentliche Änderungen** an einem bereits in Verkehr gebrachten KI-System dazu führen, dass derjenige, der ändert, **zum neuen Anbieter** wird [29]. Gleiches gilt, wenn man die **Zweckbestimmung** eines KI-Systems so ändert, dass aus einem ehemals nicht-hochriskanten System plötzlich ein Hochrisiko-KI-System wird [30]. In diesen Fällen verliert der ursprüngliche Hersteller seine Anbieterrolle; die Verantwortung – und Pflichten – gehen vollständig auf den **modifizierenden Akteur** über [31]. Für Unternehmen heißt das: Wer ein Open-Source-Modell oder ein Zukauf-Modell so umbaut, dass es wesentlich anders funktioniert oder in einem streng regulierten Bereich eingesetzt wird, kann sich **nicht darauf berufen**, die Compliance läge noch beim Originator. *Praxis:* Bei größerem Fine-Tuning oder Integration eines KI-Moduls in eine eigene Anwendung sollte man die **rechtliche Rolle neu bewerten**. Wenn z. B. aus einem allgemeinen Textmodell durch Anreicherung mit proprietären Daten ein spezialisiertes Beratungs-KI-System wird, das Kundenentscheidungen beeinflusst (und damit evtl. hochriskant ist), wird man selbst zum Anbieter und muss alle Nachweise erbringen. Tipp: Sichern Sie sich vom ursprünglichen Anbieter (soweit vorhanden) **alle technischen Informationen**, die verfügbar sind – laut AI Act muss der ursprüngliche Anbieter dem neuen Anbieter angemessen zuarbeiten (Doku, technische Daten) [31]. Trotzdem bleibt die Hauptverantwortung bei Ihnen.
- **Reine Nutzung (unverändertes KI-Produkt nutzen):** Kauft oder nutzt das Unternehmen ein fertig verfügbares KI-System **ohne es technisch zu verändern**, bleibt man im Regelfall **Betreiber**. Die Herstellerpflichten (z. B. CE-Kennzeichnung, Risikoanalyse) liegen beim Anbieter des Produkts. Als Betreiber muss man aber trotzdem bestimmte Vorgaben einhalten, vor allem bei Hochrisiko: Art. 26 AI Act verpflichtet Betreiber u. a., die **ordnungsgemäße Verwendung** sicherzustellen, die vom Anbieter mitgelieferten **Instruktionen zu befolgen**, für menschliche Überwachung im Betrieb zu sorgen und (falls relevant) die **Mitbestimmungsgremien** zu informieren [32] [33]. Beispiel: Bei Nutzung eines extern eingekauften HR-KI-Tools muss das Unternehmen seine HR-Mitarbeiter schulen, eine DSFA/Grundrechte-Abschätzung machen und den Betriebsrat über den Einsatz informieren [34]. Für die Architektur des AI-Lab hieße "reine Nutzung" z. B., wenn man ein Modell wie OpenAI's GPT-4 über API einbindet: Hier müsste OpenAI als Anbieter die EU-Vorgaben erfüllen, während man selbst vor allem auf korrekte Integration und Nutzung achten muss. Dennoch bleibt die **Verantwortung für den Endeinsatz**

beim Unternehmen – das heißt, falls das KI-System falsch verwendet wird (z. B. in einem Kontext, für den es nicht vorgesehen ist), haftet der Betreiber. Daher sollte auch beim reinen Zukauf die interne Governance sicherstellen, dass das KI-System *nur im vorgesehenen Rahmen* verwendet wird.

Architektur-Überlegung: Je nachdem, welche Rolle man einnimmt, kann es sinnvoll sein, unterschiedliche **Architekturoptionen** zu wählen. Wenn man etwa die Anbieterrolle scheut, könnte man bevorzugen, **fertige KI-Komponenten** mit Herstellerunterstützung zu nutzen (z. B. ein kommerzielles LLM mit Zusicherung der Compliance-Unterlagen). Will man maximale Unabhängigkeit und Flexibilität – z. B. Open-Source-Modelle frei kombinieren – muss man dafür sorgen, dass die eigene Infrastruktur alle nötigen Kontrollpunkte abdeckt (Logging, Monitoring, Sicherheitsupdates etc.), da man sich nicht auf einen Vendor verlassen kann. In jedem Fall gilt: **Rolle klären, bevor man ein KI-System beschafft oder baut** – das gehört zu den Vorüberlegungen im KI-Projekt [35] [36] . Eine klare Rollenverteilung hilft, die richtigen Verträge (bei Drittanbietern) aufzusetzen und intern die notwendigen Ressourcen (für Dokumentation, Schulung, Monitoring) einzuplanen.

Fazit und Empfehlungen für die KI-Lab-Architektur

Fazit: Der EU AI Act stellt auch für lokal betriebene KI-Stacks klare Anforderungen. Unternehmen müssen – abhängig vom Einsatz – Transparenz schaffen, Risiken managen und ggf. strenge Nachweispflichten erfüllen. Ein selbst gehostetes **AI-Lab** kann dabei Vor- und Nachteile haben: Einerseits behält man die Daten unter Kontrolle und kann technische Maßnahmen eigenständig umsetzen. Andererseits trägt man auch direkt die Compliance-Verantwortung und kann sich nicht auf einen Dienstleister abwälzen.

Empfehlungen (Summary):

- **Risikoanalyse vorab:** Prüfen Sie frühzeitig, in welche Risikokategorie Ihr geplanter KI-Anwendungsfall fällt. Vermeiden Sie möglichst Anwendungsfelder, die offenkundig Hochrisiko sind, es sei denn, Sie sind bereit für den entsprechenden Aufwand (Konformitätsbewertung, CE-Kennzeichnung etc.). Nutzen Sie die Übergangszeit bis 2027, um Pilotprojekte durchzuführen und Erfahrung mit der AI-Act-Compliance zu sammeln.
- **Architektur "compliance-ready" gestalten:** Bauen Sie Logging, Monitoring, Benutzerrechte und Schnittstellen für Human Oversight von Anfang an in Ihren KI-Stack ein. Ein KI-System, das von Haus aus Protokolle speichert, erklärbare Teilergebnisse liefert (wo möglich) und Kontrollmöglichkeiten bietet, wird die Anforderungen aus AI Act **leichter erfüllen**. Beispiel AI-Lab: Durch die Kombination von n8n-Workflows mit menschlichen Freigabeschritten und die Speicherung aller Zwischenergebnisse (z. B. OCR-Ergebnisse in PostgreSQL) lässt sich später zeigen, wie eine Entscheidung zustande kam – das erleichtert sowohl interne Audits als auch externe Nachweispflichten.
- **Dokumentation und Schulung nicht vergessen:** Technische Architektur ist das eine, aber **Organisationsprozesse** sind ebenso wichtig. Erstellen Sie für Ihren KI-Stack eine **Nutzerdokumentation** (wer darf was, wie funktioniert es) und schulen Sie die Anwender. Mitarbeiter sollten die Grenzen der KI kennen und wissen, wie sie etwaige Fehler korrigieren oder melden können. Eine informierte Nutzerschaft reduziert Fehlanwendungen und sorgt dafür, dass Transparenzpflichten (Art. 13/52) praktisch gelebt werden.
- **Kontinuierliche Governance:** Etablieren Sie einen Prozess, um **regelmäßig** die KI-Nutzung zu evaluieren – gerade weil sich die Gesetzeslage und die Technologien weiterentwickeln. Passen Sie Ihre Policies an neue Regulierungsschwerpunkte an (die EU AI Act-Auslegung wird sich in den nächsten Jahren konkretisieren). Auch **Security-Updates** und Model-Upgrades sollten geplant erfolgen, um stets einem "anerkannten Stand der Technik" zu genügen [37] . Dies kann z. B.

bedeuten, das AI-Lab vierteljährlich auf neue Versionen der Open-Source-Komponenten zu aktualisieren und diese Updates zu testen, bevor sie produktiv gehen.

- **Interdisziplinäre Zusammenarbeit:** Die Verantwortung für einen KI-Stack liegt nicht nur bei der IT. Beziehen Sie **Datenschutz, Compliance und Fachabteilungen** mit ein – etwa in einem Lenkungsausschuss. So stellen Sie sicher, dass sowohl technische als auch rechtliche Aspekte berücksichtigt werden (z. B. kann der Datenschutz frühzeitig warnen, wenn spezielle Daten heikel sind; die Rechtsabteilung kann helfen, die Dokumentation AI-Act-konform aufzusetzen).

Mit diesen Maßnahmen kann ein Unternehmen die **Chancen eines selbstgehosteten KI-Stacks** nutzen, ohne die rechtlichen Fallstricke zu übersehen. Der EU AI Act mag komplex wirken, lässt sich aber durch proaktive Planung und gutes Architekturdesign in den betrieblichen Alltag integrieren – für einen innovativen **und** verantwortungsvollen KI-Einsatz im Unternehmen.

Quellen: Die obigen Informationen basieren auf der finalen Fassung der EU-KI-Verordnung (Stand Juni 2024) sowie einschlägigen Leitfäden [1] [9] [18] [29] [23]. Diese gewährleisten, dass die Empfehlungen **rechtssicher und praxisnah** sind, um Ihr AI-Lab fit für die KI-Regulierung zu machen.

---

[1] [3] [4] [7] [8] [9] [10] [11] [15] [16] [18] [20] [23] [24] [28] [32] [33] [34] 06-2025 KI-Verordnung Broschüre PRINT FINAL.indd
https://digitalzentrum-lmo.de/wp-content/uploads/2025/10/06-2025-KI-Verordnung-Broschuere-PRINT-FINAL.pdf

[2] [21] What Open Source Developers Need to Know about the EU AI Act
https://linuxfoundation.eu/newsroom/ai-act-explainer

[5] [6] [12] [13] [14] [17] [29] [30] [31] [35] [36] [37] Generative KI im Unternehmen
https://www.bitkom.org/sites/main/files/2024-02/Bitkom-Leitfaden-Generative-KI-im-Unternehmen.pdf

[19] AI Act | Shaping Europe's digital future - European Union
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[22] The EU AI Act: Application to Open-Source Projects
https://www.orrick.com/en/Insights/2024/09/The-EU-AI-Act-Application-to-Open-Source-Projects

[25] [26] [27] What do the open source exemptions for GPAI models mean for you? The EU AI Act Guidelines provide some clarity, Natalie Donovan
https://thelens.slaughterandmay.com/post/102kzax/what-do-the-open-source-exemptions-for-gpai-models-mean-for-you-the-eu-ai-act-gu