

# UX for AI

Erfolgreich mit nutzerzentrierten  
KI-Anwendungen und -Ökosystemen





Zusammenfassung

# Ihr Wegweiser für ein funktionierendes KI-Ökosystem

**UX for AI holt das Passende und Beste aus KI heraus – für spürbare Wertbeiträge heute & ein adaptives Ökosystem morgen.**

Unternehmen haben GenAI – was fehlt, ist der Sprung vom Pilot zur Wirkung im Arbeitsalltag.

## Worum es geht

Wir identifizieren Use Cases, prototypisieren und verproben sie. Und wir messen von Anfang an, wie die Lösung wirkt. So entstehen Anwendungen, die Menschen befähigen statt belasten – und Schritt für Schritt ein KI-Ökosystem.

### Die 3 +1 Prinzipien

1. Relevanz: Kontext erkannt, Daten vorhanden
  2. Vertrauen durch Transparenz und Kontrolle
  3. Integration in bestehende Workflows
- +1 Kontinuierliche Evaluation

### Ergebnis

Schnellere Wertnachweise (erste KPIs in Wochen).  
Höhere Akzeptanz (bessere Qualität, weniger Verifikationsaufwand).  
Sichere Skalierung (vom Einzeltool zum Ökosystem).

### Was Sie davon haben

Dieses Whitepaper zeigt kompakt, wie Sie Ihre KI-Strategie durch strikten Anwenderfokus und kontinuierliche Evaluation auf sichere Füße stellen — von der Use-Case-Auswahl über prototypisches Verproben bis zur Wirkung im Betrieb.



# Inhalt

- 01 | WARUM KI-PROJEKTE SCHEITERN
- 02 | WIE KI-ANWENDUNGEN GELINGEN
- 03 | UX für AI – DREI SCHRITTE ZUM ERFOLG
- 04 | ADAPTIVE KI-ÖKOSYSTEME BAUEN
- 05 | WIE SIE IHRE KI-STRATEGIE VERFEINERN



Status Quo

# Warum KI-Projekte scheitern

## KI ist in den Unternehmen angekommen

Fast alle experimentieren mit KI und versuchen, die Chancen der neuen Technologie gewinnbringend zu nutzen. Tools wie ChatGPT und Copilot sind aus dem Arbeitsalltag nicht mehr wegzudenken – trotzdem zahlen sich viele KI-Initiativen bisher nicht aus.

Aktuelle Studien wie der Report „State of AI in Business 2025“ des MIT zeigen:

**Trotz Investitionen im zweistelligen Milliardenbereich liefern rund 95 % der Gen AI-Projekte keinen messbaren Business-Return, und nur 5 % schaffen den Sprung in skalierte, wertstiftende Nutzung.** Häufig bleiben Vorhaben in Piloten stecken – die Studie spricht von einem „GenAI Divide“.

These: Die Adoption ist hoch, die Wertrealisierung bleibt jedoch aus – nicht wegen der KI-Modelle, sondern aufgrund der Vorgehensweise. Woran liegt das genau? Zwei aktuelle Studien vom MIT und McKinsey geben Auskunft darüber.

## Gründe für den ausbleibenden Erfolg von KI-Initiativen

- Auswahl ungeeigneter Use Cases. Viele Vorhaben liefern zu geringe Werthebel oder passen nicht zur Arbeitsrealität und zu bestehenden Prozessen.
- Kein Umgestaltung der Arbeitsprozesse. KI wird als Zusatz-Tool eingeführt, statt als Workflow-Upgrade, das Aufgaben, Rollen und Handover neu denkt.
- Schwache Integration und Data-Fit. Fehlende Anbindung an Systeme, Policies und Domänenwissen verhindert Wirkung im Tagesgeschäft.
- Fehlende bzw. zu späte Evaluation des Modell-Outputs. Der notwendige Aufwand zur Überprüfung der Ergebnisse bleibt hoch; Qualität und Risiken werden zu wenig gemessen.
- Buy-vs.-Build ohne klare Kriterien. Eigenbauten überfordern oft Organisationen, während Partnerschaften und spezialisierte Lösungen schneller performen.
- Shadow IT und Tool-Zoo. Mitarbeitende nutzen generische Tools produktiv, während offizielle Lösungen nicht überzeugen.
- Fehlende Governance und KPI-Fokus. Es fehlen klare Verantwortungen, Messpunkte sowie Entscheidungs- und Eskalationspfade.





Das Ziel

# Wie KI-Anwendungen gelingen

Was also tun?

Echte Wertbeiträge entstehen, wenn KI-Lösungen an Menschen und ihre Arbeitsabläufe andocken.

Entscheidend ist der richtige Ansatz: Wir müssen den Blick konsequent auf die Anwender richten, ihre Arbeitsrealität ganzheitlich betrachten und Workflows systematisch neu gestalten – stets mit klar definiertem Business-Outcome.

Ein ganzheitlicher UX for AI-Ansatz richtet sich positiv am Menschen aus: Er automatisiert nicht nur Aufgaben, sondern augmentiert Kreativität, indem er Kontexte versteht, gute Vorschläge macht und Exploration sicher ermöglicht.

**Jede KI-Anwendung muss 3+1 Qualitätskriterien erfüllen:**

Relevanz, Vertrauen, Integration und Evaluation.

*"In this new paradigm, the best interfaces  
might feel less like tools and more like  
collaborators, less like pathways and  
more like playgrounds."*

Jakob Nielsen



3 +1 Prinzipien

# Wie KI-Anwendungen gelingen

1

## Relevanz für den Anwendungsfall

Die Anwendung versteht den konkreten Kontext und hat die nötigen Daten zur Hand, um präzise Antworten zu liefern.

- Wissensquellen anbinden über Retrieval/Grounding
- Kontext strukturiert übergeben
- Zugriffsrechte sauber regeln
- Policy-konform arbeiten

2

## Vertrauen by Design

Nutzer:innen behalten die Kontrolle, sehen Grenzen und Herkunft der Antworten und können jederzeit eingreifen.

- Quellen und Konfidenz anzeigen
- Aktionen reversibel halten
- Eskalationspfade vorsehen (Human-in-the-Loop)
- Fehler defensiv gestalten

3

## Nahtlose Integration

Die Lösung fügt sich in den bestehenden Arbeitsablauf ein und nimmt Arbeit ab, statt neue zu schaffen.

- Übergaben in Fachsysteme
- Automatische Protokollierung
- Rollen, Rechte und Audits
- Beobachtbarkeit/Telemetrie

+1

## Kontinuierliche Evaluation

Die Qualität der Ergebnisse wird laufend gemessen und verbessert; das System lernt aus Feedback.

- Offline-Tests mit Golden Sets
- Online-Metriken
- Human- und AI-Feedback-Schleifen
- Drift-Monitoring
- Guardrails
- Regelmäßiges Red-Teaming



# Unscheinbar, aber mächtig

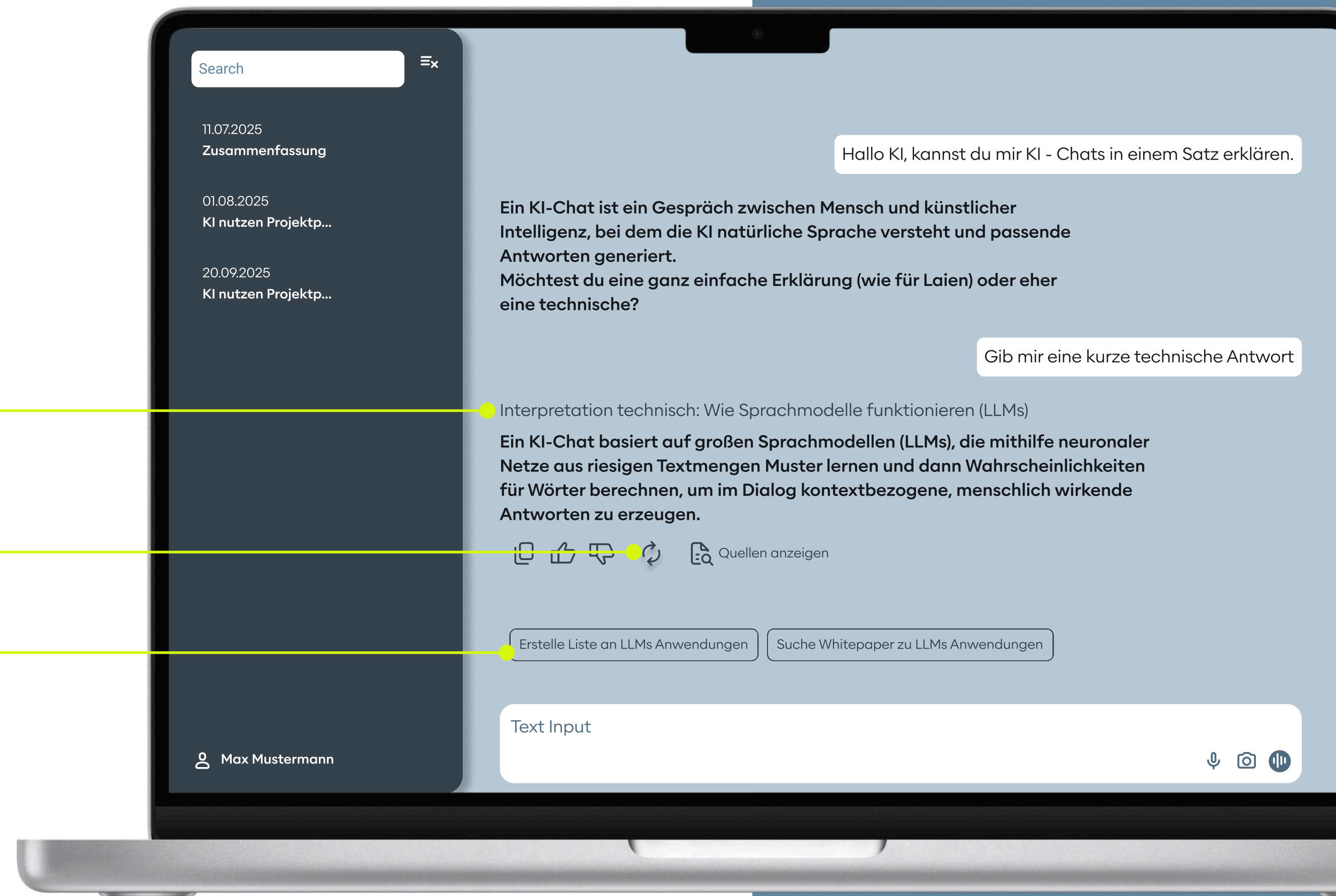
Kleine Elemente mit großer Wirkung: Warum funktionieren Anwendungen wie ChatGPT relativ gut? Weil hinter jedem kleinen Schalter im Interface eine gut überlegte Funktion steckt.

Spezielle Arten des Feedbacks und hilfreiche Symbole helfen den Anwendern, Chat-Anwendungen effizienter zu nutzen. UX Design bedeutet nicht nur, Interfaces zu gestalten, sondern eben auch: aus den Sprachmodellen das Maximum rauszuholen.

**Restating:** Die KI beschreibt, wie es den Input verstanden hat.

**Regenerate Button:** Einfache Möglichkeit, die Anfrage nochmals auszuführen.

**Next Steps:** Das System schlägt passende Anschlussaktionen vor.



# UX for AI – Drei Schritte zum Erfolg

## Vom Warum und Was zum Wie

**Unser UX-for-AI-Vorgehen führt in drei Schritten von der Idee bis zum Wirknachweis.**

**Der rote Faden dabei:** eine kontinuierliche Evaluationsschleife (Evaluation Loop) von Anfang an.

**Kontinuierliche Evaluation:** Wir messen Qualität von Anfang an – erst als Plan (Schritt 1), dann in kurzen Testzyklen (Schritt 2) und anschließend im Live-Betrieb (Schritt 3).

### Understand & Define

### Design & Prototype

### Implement & Grow

## Schritt 1 – Understand & Define

**Ziel:** Problem und Werthebel schärfen, Risiken & Datenlage verstehen, gemeinsames Zielbild fixieren

### Aktivitäten

Workflow-Analyse: Aufgaben, Brüche, Datenflüsse im Ist-Prozess

Risiko- & Compliance-Analyse: fachlich, rechtlich, ethisch

Qualitäts- & Erfolgsdefinition: KPIs, Qualitätskriterien, Akzeptanz

Zielbild visualisieren: Storyboard & Digital Twin der Lösung

**Methoden (Überblick):** Stakeholder/Jobs-Mapping, Journey/Process Walkthroughs, Value-Matrix, Storyboard & Digital Twin, Discovery-Workshops

**Ergebnisse:** Priorisierte Use-Cases, Storyboard & Digital Twin, KPI- & Qualitätsdefinition, Risiko-/Daten-Landkarte

### Evaluation Loop

Kriterien und Messpunkte festlegen

Testplan skizzieren

Golden-Set definieren und testen





## Schritt 2 — Design & Prototype

**Ziel:** Interaktion formen, mit realem Kontext testen und Qualität messbar machen

### Aktivitäten

Interaktions- & Flow-Design  
Rapid Prototyping und nutzernahe Tests  
Vertrauens- & Fehlermanagement (Transparenz, Reversibilität, Handover)  
Messaufbau finalisieren

**Methoden (Überblick):** Conversational Design, Rapid Prototyping (Skizze → klickbarer Prototyp → RAG-PoC), RITE-Testing, Wizard-of-Oz, Bewertungsrubriken

**Ergebnisse:** Validierter Prototyp (mit Erkenntnissen), Golden-Set & Rubriken (final definiert), Design-Richtlinien (Transparenz/Undo/Handover), 90-Tage-Plan/Go-No-Go

### Evaluation Loop

Kurze Testzyklen (z.B. 2–3 Runden) mit realistischen Szenarien, Baseline für zentrale Metriken entsteht

## Schritt 3 — Implement & Grow

**Ziel:** In den Betrieb bringen, integrieren, absichern und skalieren

### Aktivitäten

Integration & Datenfluss in Systeme und Domänenwissen  
Leitplanken & Betrieb aufsetzen  
Enablement & Change für die Teams  
Skalieren & Standardisieren

**Methoden (Überblick):** Grounding, API-Integration, Zugriffskontrolle (Rollen/Rechte), Leitplanken & Richtlinien (Guardrails/Policy), Performance Tracking (Tracing, Latenz, Kosten) & Nutzungsmetriken, A/B-Tests, Drift-Monitoring, Standards & Wiederverwendung (z.B. Skills-Registry)

**Ergebnisse:** MVP/Release mit angebundenen Systemen & Daten, Betrieb mit Leitplanken, Runbooks & Performance-Dashboards, Adoption & KPI-Monitoring live, Standards/Policies und Roadmap/Portfolio für die Skalierung

### Evaluation Loop

Online-Metriken laufen (Qualität, Nutzung, Performance), Drift-Monitoring & A/B-Tests steuern das Tuning im Betrieb



## KI ist ein Teamsport

UX, Data und IT/Plattform arbeiten **im Takt**:

Jede UX-Iteration schärft Anforderungen an das KI-Modell.  
Jeder RITE-Spike legt **Datenlücken, Bias oder Policy-Grenzen** offen;  
daraus entstehen Design-Workarounds und **neue Ideen** – die  
nächste Runde beginnt.

In diesem „neuen Normal“ ist **UX der Kitt** zwischen Kund:innen,  
Business und Technologie:

Das Kernteam löst Probleme **gemeinsam und in Echtzeit**, testet  
schnell mit **echten Daten** und bringt Lösungen in den Betrieb – statt  
Berichte zu produzieren.

Problemanalyse /  
Evaluation

Ideen und Patente

UX = Kitt

Effizienz /  
Geschwindigkeit

Kosteneinsparung



Ausblick

# Adaptive KI-Ökosysteme bauen

## Von der einzelnen App zur Plattform

Mit "UX for AI" bauen nicht nur einzelne Apps, sondern ein System aus Fähigkeiten, Daten und Regeln. Komponenten teilen Kontext und arbeiten zusammen – so entsteht aus Einzelerfolgen ein lernendes Ökosystem. In Zukunft greifen **vier KI-Anwendungstypen** auf einer gemeinsamen Basis nahtlos ineinander, passen sich an die Menschen und ihre Aufgaben an.

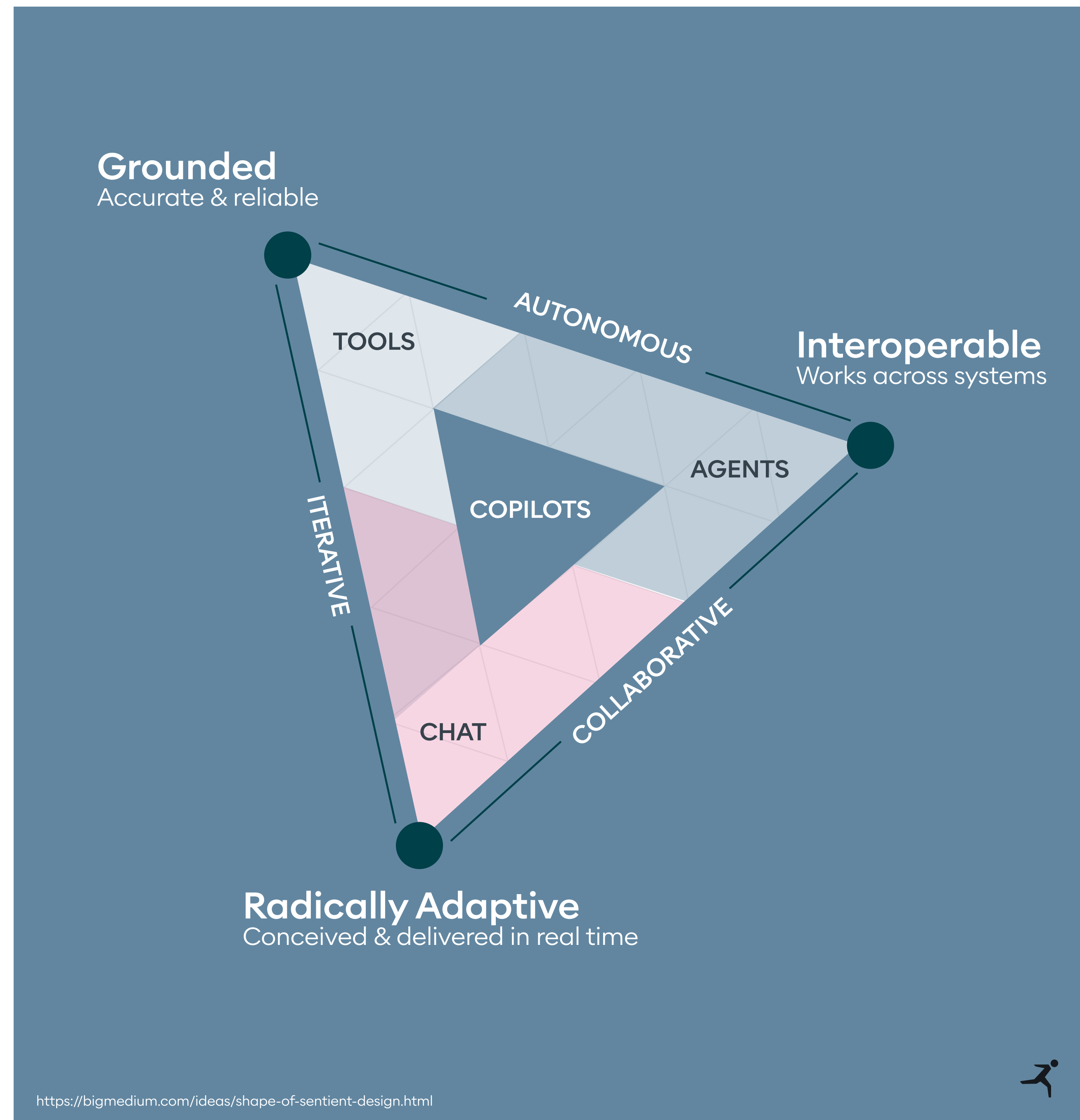
**Tool:** klare Funktion, schneller Nutznachweis

**Copilot:** Assistenz im Flow, bessere Qualität

**Agent:** mehrschrittig mit Handover, kontrollierte Autonomie

**Chat:** Exploration & Wissensverdichtung

**Weiterlesen:** Sentient Design beschreibt, wie Interfaces Kontext & Intention erkennen und sich im Moment anpassen. Spannend als Zielbild – an dieser Stelle führt das (leider) zu weit.





# Gutes Design für produktive Anwendungen

## KI-Anwendungen sind mehr als Chats

Je nach Anwendungsart und Prozess-Status ist es sinnvoll, den Aufbau des Interfaces zu ändern, also zum Beispiel die Größe des Chatfensters und sein Zusammenspiel mit anderen Elementen zu variieren.

Auch der klassische Copilot kann seine Struktur dem Status anpassen: mit Chat in der Randspalte oder als großes Overlay. Individuelle KI-Tools und Chats arbeiten tendenziell im Full-Screen Mode.

Agenten passen ihre Interfaces je nach Stand der Aufgabenerledigung kontinuierlich an den aktuellen Zustand an.

Was wann am besten passt, hängt vom Nutzungskontext und den präsentierten Inhaltstypen ab. UX for AI gestaltet passende UI Patterns für alle Lagen.





Diese Vision dient als Kompass für die Entwicklung funktionierender KI-Ökosysteme. Dabei sind folgende Elemente entscheidend:

#### **Design-Leitlinien**

Erfolgreiche KI-Ökosysteme starten mit klaren UX-Prinzipien: Menschen geben Intentionen ein, sehen Fähigkeiten und Grenzen, können jederzeit eingreifen und erhalten hilfreiche Erklärungen. Eskalationspunkte (Handover) und Privacy-by-Default sichern Verantwortung und Vertrauen. Diese Leitplanken reduzieren Fehlerkosten und erhöhen die Akzeptanz.

#### **Archetypen als Landkarte**

Wir wählen die passende Erlebnisform je Journey: Tools für präzise Funktionen, Copilots für kontextstarke Assistenz im Flow, Agents für mehrschrittige Aufgaben mit kontrolliertem Handover und Chat für Exploration & Wissensverdichtung. Diese Archetypen helfen, Wert schnell freizuschalten – ohne alles gleichzeitig bauen zu müssen.

#### **Plattform-Bausteine:**

Damit alle Komponenten zuverlässig zusammenspielen, braucht es gemeinsame Infrastruktur: ein einheitliches Data-Grounding für Domänenwissen, eine Skills/Tools-Registry zur Wiederverwendung von Fähigkeiten, einen Policy & Compliance Layer für einheitliche Regeln sowie Performance-Tracking für Betrieb und Qualität.

**Ergebnis: So entsteht ein Ökosystem-Blueprint mit Prioritäten, Standards und Metriken – als Grundlage für Roadmaps und Investitionen.**



Jetzt handeln

# Wie Sie Ihre KI-Strategie verfeinern

**UX for AI** verwandelt verstreute Piloten in wirksame Anwendungen und lernfähige Plattformen – mit Fokus auf Intention, Vertrauen, Workflow-Neugestaltung und Evaluation. Wenn Sie Ihre KI-Strategie auf solide Füße stellen möchten, nutzen Sie diesen Ansatz, um Klarheit über geeignete Use Cases und passende AI Patterns zu gewinnen.

Wir bieten passgenaue Unterstützung in verschiedenen Formaten und Größen. Vom Schnupper-Workshop über ein AI Maturity Assessment bis zur gemeinsamen Arbeit an KI-Projekten steigen wir gerne in Ihre Projekte ein.

**Die Zukunft wird spannend und hat schon längst begonnen. Lassen Sie uns gemeinsam daran arbeiten!**

*"AI isn't here to execute directions but to be a co-creator that takes humans to new heights."*

Jakob Nielsen / Eras of UX

## Mögliche Einstiegsformate

**Ziel:** Unternehmen und Akteure kennenlernen, Status der KI-Adoption ermitteln, wertstiftende Projekte definieren und aufplanen.



### Schnupperangebot: AI Ideation Workshop

1/2 Tag gemeinsames Brainstorming zur Identifikation geeigneter Use Cases für den Einsatz von KI in Ihrem Unternehmen.



### AI Strategy Workshop

1-2-tägiger Workshop mit individueller Vorbereitung.  
Ergebnis: Priorisierte Use Cases mit Backlog und provisorischer Projektplan.



### UX for AI-Projektbegleitung

Begleitung von AI-Projekten aus UX-Perspektive.  
User Research, Use Case-Definition, Rapid Prototyping, Testing und laufende Evaluation.





Kontakt

# Warum wir?

Als AI-First-Agentur bedienen wir alle Themen aus diesem Whitepaper – von Research & Design über Evaluation & Governance bis zur Integration in Fachsysteme.

Als Teil der Reply-Gruppe haben wir für jedes technologische Setup (AI, Data, MLOps, Data, Security) die passenden Partner:innen an Bord.



„Want to learn more?“

Vereinbaren Sie einen Termin für ein Gespräch mit unseren Spezialisten:

**Telefon:** 069. 9 99 99 37 – 0 | **E-Mail:** [triplesense@reply.de](mailto:triplesense@reply.de)



**Web:** <https://www.triplesensereply.de/>



# Glossar & Quellen

## Glossar

- Conversational Design — Gezieltes Entwerfen von Dialogen/Prompts, Zuständen und Rückfragen.
- Data- & Capability-Mapping — Welche Daten/Fähigkeiten vorhanden sind bzw. fehlen; Basis für Machbarkeit & Priorisierung.
- Design of Experiments (DoE) — Geplante Versuchsreihen zum effizienten Tuning und Verständnis von Faktorwirkungen.
- Digital Twin (der Lösung) — Skizze der künftigen Lösung als System-Zwilling (Oberflächen, Daten, Fähigkeiten, Regeln).
- Drift-Erkennung/Monitoring — Veränderungen bei Daten oder Modellleistung früh erkennen und gegensteuern.
- Evaluation-as-a-Service — Zentrale Services für Tests, Scoring, Golden-Sets & Red-Teaming.
- Golden Sets — Kuratierte Testfälle mit Soll-Antworten; Grundlage für reproduzierbare Qualitätsmessung.
- Guardrails (Leitplanken) — Ein-/Ausgabeprüfungen, Filter und Aktionsgrenzen für sichere Ergebnisse.
- Handover & Eskalation — Definierte Übergaben & Wege bei Unsicherheit/Fehlern (Mensch – Agent – System).
- Identity / RBAC — Rollenbasierte Zugriffsrechte auf Daten und Funktionen.
- KPI-Steuerung — Lead-/Lag-KPIs plus regelmäßige Reviews zur Kurskontrolle.
- Maturitäts-Pfad (Tool-Fit → Ökosystem-Fit) — Stufenweise Skalierung mit klaren Gates (Nutzen-Beweis, Leitplanken, Tracking).
- Performance Tracking — Tracing, Metriken und Logs zu Qualität, Latenz, Kosten im Betrieb.
- Policy & Compliance Layer — Ausführbare Regeln (PII, Safety, Regulatorik) als zentrale Schutzschicht.
- RAG (Retrieval-Augmented Generation) — Kombination aus Retrieval und Generierung für fundierte, zitierbare Antworten.
- Red-Teaming — System gezielt „angreifen“, um Schwachstellen und Missbrauchsrisiken aufzudecken.
- Retrieval / Grounding — Domänenwissen zur Laufzeit einbinden; erhöht Relevanz und reduziert Halluzinationen.
- Risk & Ethics Canvas / Compliance-Check — Risiken, Pflichten und Leitplanken systematisch erfassen.
- RITE-Testing — Rapid Iterative Testing & Evaluation: kurze Testzyklen mit sofortigen Fixes.
- Sentient Design — Interfaces erkennen Kontext & Intention und passen sich adaptiv an (Vision/Zielbild).
- Skills/Tools-Registry — Verzeichnis ausführbarer Fähigkeiten mit Metadaten/Rechten zur Wiederverwendung.
- Task-Success-Rate — Anteil erfolgreich erledigter Aufgaben (Kern-Outcome-Metrik).
- Time-to-First-Draft (TtFD) — Zeit bis zum brauchbaren ersten Entwurf (Tempo-Leitmetrik).
- Trust-Patterns — UI-Bausteine für Verständnisprüfung, Begründung/Transparenz und schnelle Nachbesserung.
- Verification Tax — Zusätzlicher Aufwand für händische Ergebnisprüfung; sollte durch Design/Eval sinken.
- Wizard-of-Oz (WoZ) — Prototyping, bei dem Menschen „heimlich“ Systemintelligenz simulieren, um realistisch zu testen.

## Quellen

- MIT / Project NANDA (2025): „The GenAI Divide: State of AI in Business 2025“.
- McKinsey & Company (2024): „The state of AI in early 2024: Gen AI adoption spikes and starts to generate value“.
- Jakob Nielsen: „A New AI: Creation as Exploration and Discovery“.
- Josh Clark / Veronika Kindred: „The Shape of Sentient Design“.
- Greg Nudelman: „UX for AI: A Framework for Designing AI-Driven Products“. (Wiley)
- Lewis et al. (2020): „Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks“. (arXiv / NeurIPS)
- Medlock et al. (2002): „Using the RITE Method to Improve Products: A Definition and a Case Study“

