REPORT

Von der KI-Idee zum Pilotprojekt. Ein Leitfaden

Wie aus einer KI-Idee ein Pilotprojekt wird, das eine echte Entscheidung ermöglicht. Skalieren, überarbeiten oder bewusst stoppen.

Report
Geschrieben von Emanuel Stadler

Optimentum

Über diesen Report

Dieser Report ist die ausführliche Fassung des gleichnamigen Beitrags auf optimentum.eu. Er führt durch jede Phase, mit Kriterien, Arbeitsblättern, Checklisten und Vorlagen für die eigene Arbeit. Beispiele sind veranschaulichend und beschreiben keine Kundenprojekte. Rechtliche Hinweise ersetzen keine rechtliche Beratung.

Report
Geschrieben von Emanuel Stadler

Optimentum GmbH · Tegetthoffstraße 3/10, 1010 Wien · FN 686237y, Handelsgericht Wien
optimentum.eu · office@optimentum.eu

Inhalt

Zusammenfassung

01 Warum Pilotprojekte scheitern

02 Die richtige Aufgabe wählen

03 Entscheidungsfrage und Kriterien

04 Daten, Recht und Sicherheit

05 Mensch und Verantwortung

06 Den Pilotbetrieb gestalten

07 Auswerten und entscheiden

08 Vom Pilot in den Betrieb

Anhang

ZUSAMMENFASSUNG

KERNERGEBNISSE

01 Ein Pilotprojekt ist ein Instrument für eine Entscheidung. Ohne vorab definierte Frage liefert es Eindrücke, aber keine Grundlage.

02 Ob ein Pilotprojekt trägt, entscheiden Daten, Zuständigkeiten und Abläufe ebenso wie die Technik.

03 Wer Fragen zu Informationssicherheit und Datenschutz erst am Ende stellt, riskiert, dass ein Pilotprojekt kurz vor dem Betrieb stecken bleibt.

04 Menschliche Kontrolle wird mit dem Ablauf gestaltet. Sie bestimmt Risiko und Nutzen gleichermaßen.

05 Die Voraussetzungen für den Betrieb gehören in die Planung des Pilotprojekts, nicht an sein Ende.

DAS WICHTIGSTE IN KÜRZE

- Legen Sie vor dem Start fest, welche Frage das Pilotprojekt beantworten soll.
- Formulieren Sie Erfolgs- und Abbruchkriterien so, dass sie sich prüfen lassen.
- Testen Sie mit echten, auch schwierigen Fällen, nicht nur mit vorbereiteten Beispielen.
- Beenden Sie das Pilotprojekt mit einer Entscheidung, nicht mit einer Präsentation.

KAPITEL 01

Warum Pilotprojekte scheitern

Von der KI-Idee zum Pilotprojekt. Ein Leitfaden

Ein Prototyp für eine KI-Anwendung ist heute in wenigen Tagen gebaut. Er zeigt, dass ein Modell eine Aufgabe grundsätzlich lösen kann. Ob die Lösung im Alltag trägt, ob sie sich verantwortungsvoll betreiben lässt und ob sich der Aufwand lohnt, zeigt er nicht. Diese Lücke soll ein Pilotprojekt schließen.

In der Praxis gelingt das oft nicht. Viele Pilotprojekte enden mit einem positiven Eindruck, einer Präsentation und ohne Entscheidung. Die Lösung bleibt in einem Zwischenzustand, in dem sie weder eingesetzt noch beendet wird. Dieses Kapitel beschreibt, woran das liegt und was ein Pilotprojekt stattdessen leisten muss.

Prototyp, Proof of Concept und Pilotprojekt

Die drei Begriffe werden oft gleichbedeutend verwendet. Sie beantworten aber unterschiedliche Fragen, und wer sie verwechselt, erwartet von einem Vorhaben etwas, das es nicht liefern kann.

Ein **Prototyp** macht eine Idee greifbar. Er dient dazu, eine Vorstellung zu teilen und Rückmeldungen zu sammeln. Ein Prototyp darf unvollständig sein, er arbeitet meist mit wenigen, oft vorbereiteten Beispielen und läuft außerhalb der tatsächlichen Systeme.

Ein **Proof of Concept** prüft, ob etwas technisch möglich ist. Kann ein Modell aus diesen Dokumenten die benötigten Angaben verlässlich herauslesen? Lässt sich das Quellsystem anbinden? Ein Proof of Concept beantwortet eine eng gefasste technische Frage mit Ja oder Nein.

Ein **Pilotprojekt** prüft, ob eine Lösung im tatsächlichen Ablauf Nutzen stiftet und sich verantwortungsvoll betreiben lässt. Es arbeitet mit echten Fällen, mit den Menschen, die die Aufgabe heute erledigen, und mit Kriterien, die vorab feststehen. Sein Ergebnis ist eine Entscheidung über die Fortführung.

Merkmal	Prototyp	Proof of Concept	Pilotprojekt
Leitfrage	Wie könnte es aussehen?	Ist es technisch möglich?	Lohnt es sich im Alltag, und lässt es sich verantwortungsvoll betreiben?
Fälle	Wenige, oft vorbereitete Beispiele	Ausgewählte Testdaten	Echte, repräsentative Fälle, auch schwierige
Beteiligte	Kleines Projektteam	Technisches Team	Fachbereich, Nutzerinnen und Nutzer, IT, Sicherheit, Datenschutz, Entscheider
Umgebung	Getrennt vom Betrieb	Test- oder Laborumgebung	Im Ablauf oder parallel dazu

Merk-mal	Prototyp	Proof of Concept	Pilotprojekt
Kriteri-en	Meist keine	Technische Machbarkeit	Qualität, Prüfaufwand, Akzeptanz, Betrieb, Risiko
Ergeb-nis	Eindruck und Rückmeldungen	Ja oder Nein zur technischen Frage	Entscheidung: skalieren, überarbeiten oder be-wusst stoppen

Die drei Formen bauen aufeinander auf, sind aber nicht zwingend nacheinander nötig. Ist die technische Machbarkeit offensichtlich, kann ein Proof of Concept entfallen. Ein Pilotprojekt lässt sich dagegen nicht durch einen besonders gelungenen Prototyp ersetzen.

Typische Muster des Scheiterns

Pilotprojekte scheitern selten an der Technik allein. Häufiger scheitern sie an Fragen, die vor dem Start nicht gestellt wurden. Die folgenden Muster treten einzeln oder gemeinsam auf.

Die Frage fehlt

Das Pilotprojekt startet mit dem Auftrag, eine Technologie auszuprobieren. Welche Entscheidung am Ende getroffen werden soll, ist nicht festgelegt. Die Ergebnisse lassen sich dann auf viele Arten deuten, und jede Seite findet ihre Bestätigung.

Vorbereitete Beispiele statt echter Fälle

Die Lösung wird mit sorgfältig ausgewählten, sauberen Beispielen getestet. Unvollständige Eingaben, Sonderfälle und schlecht lesbare Dokumente fehlen. Die Lösung sieht im Pilotprojekt besser aus, als sie im Alltag sein wird.

Der Vergleich fehlt

Niemand hat erfasst, wie lange die Aufgabe heute dauert und wie viele Fehler dabei entstehen. Am Ende steht eine Zahl ohne Bezugspunkt. Ob die Lösung besser ist als der heutige Ablauf, lässt sich nicht sagen.

Der Umfang wächst still

Während des Pilotprojekts kommen weitere Dokumentarten, Teams oder Teilaufgaben hinzu. Jede Erweiterung wirkt klein. In Summe misst das Pilotprojekt am Ende etwas anderes als geplant, und die Kriterien passen nicht mehr.

Sicherheit und Datenschutz kommen zu spät

Fragen zu Datenzugriff, Verarbeitungsort und Rechtsgrundlage werden erst gestellt, wenn die Lösung in den Betrieb gehen soll. Dann sind Architektur und Anbieter bereits gewählt. Das Pilotprojekt bleibt genau an dem Punkt stecken, an dem es Nutzen stiften sollte.

Niemand ist zuständig

Das Pilotprojekt wird von einem Innovationsteam oder einem externen Dienstleister getragen. Eine prozessverantwortliche Person im Fachbereich fehlt. Nach dem Pilotprojekt ist unklar, wer die Lösung übernimmt und wer über ihre Zukunft entscheidet.

Die Präsentation ersetzt die Entscheidung

Das Pilotprojekt endet mit einer Vorstellung der Ergebnisse vor einem größeren Kreis. Es gibt Zustimmung und Anregungen, aber keinen Beschluss. Die Lösung läuft weiter, ohne Budget, ohne Betriebsverantwortung und ohne Ende.

Wofür ein Pilotprojekt da ist

Ein Pilotprojekt ist kein verkleinerter Rollout und keine Vorführung. Es ist ein Instrument, um eine Entscheidung vorzubereiten. Diese Entscheidung betrifft eine Investition, eine Veränderung von Abläufen und eine Verantwortung für den Betrieb.

> Ein Pilotprojekt ist ein Instrument für eine Entscheidung. Ohne vorab definierte Frage liefert es Eindrücke, aber keine Grundlage.

Daraus folgen drei Anforderungen. Erstens muss vor dem Start feststehen, welche Frage beantwortet werden soll und woran sich die Antwort messen lässt. Zweitens muss das Pilotprojekt unter Bedingungen laufen, die dem späteren Alltag nahekommen: echte Fälle, echte Nutzerinnen und Nutzer, echte Systeme oder eine realistische Nachbildung. Drittens muss am Ende eine Person oder ein Gremium mit Entscheidungsbefugnis eine von drei Möglichkeiten wählen: skalieren, überarbeiten oder bewusst stoppen.

Ein bewusst gestopptes Pilotprojekt ist dabei kein Misserfolg. Es hat eine Frage beantwortet und verhindert, dass weiteres Budget in eine Lösung fließt, die nicht trägt. Misslungen ist ein Pilotprojekt erst dann, wenn es keine Entscheidung ermöglicht.

Wie dieser Leitfaden aufgebaut ist

Die folgenden Kapitel beschreiben den Weg von der Auswahl der Aufgabe bis zum Übergang in den Betrieb. Sie folgen der Reihenfolge, in der die Fragen beantwortet werden sollten. Kapitel 2 und 3 klären, was geprüft wird und woran. Kapitel 4 und 5 behandeln Daten, Recht, Sicherheit und die Rolle des Menschen. Kapitel 6 beschreibt den Pilotbetrieb selbst, Kapitel 7 die Auswertung und Entscheidung, Kapitel 8 den Schritt in den Betrieb.

Der Anhang enthält Checklisten für jede Phase, Vorlagen zum Ausfüllen und ein Glossar. Die Vorlagen sind bewusst schlicht gehalten. Sie lassen sich in bestehende Projektwerkzeuge übernehmen und an die eigene Organisation anpassen.

Der Leitfaden setzt dort an, wo ein KI-Anwendungsfall grundsätzlich ausgewählt ist. Die Priorisierung zwischen mehreren Anwendungsfällen wird nur berührt, soweit sie die Wahl der konkreten Aufgabe betrifft.

Auf einen Blick

- Ein Prototyp zeigt, was möglich ist. Ein Pilotprojekt zeigt, ob es sich im Alltag lohnt und verantwortungsvoll betreiben lässt.
- Pilotprojekte scheitern meist an fehlenden Fragen, fehlenden Vergleichen und fehlender Zuständigkeit, seltener an der Technik.
- Das Ergebnis eines Pilotprojekts ist eine Entscheidung, keine Präsentation.
- Bewusst stoppen ist ein gleichwertiges Ergebnis, wenn es begründet ist.

KAPITEL 02

Die richtige Aufgabe wählen

Von der KI-Idee zum Pilotprojekt. Ein Leitfaden

Am Anfang eines Pilotprojekts steht eine konkrete Aufgabe, nicht eine Technologie. Auch wenn der Anwendungsfall bereits feststeht, ist er meist zu breit, um ihn in einem Pilotprojekt sinnvoll zu prüfen. „KI im Kundenservice“ ist ein Anwendungsfeld. „Eingehende E-Mails einer Kategorie zuordnen und an das zuständige Team weiterleiten“ ist eine Aufgabe.

Dieses Kapitel zeigt, wie Sie aus einem Anwendungsfall eine prüfbare Aufgabe machen, nach welchen Kriterien Sie zwischen mehreren Kandidaten wählen und wie Sie das Ergebnis in einem Aufgabensteckbrief festhalten.

Von der Idee zur Aufgabe

Eine Aufgabe ist dann konkret genug, wenn Außenstehende sie ohne Rückfrage verstehen. Beschreiben Sie dafür fünf Punkte:

- **Wer** erledigt die Aufgabe heute, und wie viele Personen sind beteiligt?
- **Welche Eingaben** liegen vor, in welcher Form und aus welchen Systemen?
- **Welches Ergebnis** entsteht, und in welcher Form wird es weitergegeben?
- **Wer arbeitet** mit dem Ergebnis weiter, und was geschieht, wenn es fehlerhaft ist?
- **Wie oft** fällt die Aufgabe an, und wie stark schwankt das Aufkommen?

Ein Beispiel: Eine Versicherung prüft eingehende Schadensmeldungen. Heute liest eine Sachbearbeiterin jede Meldung, ordnet sie einer Schadenart zu, prüft die Vollständigkeit der Unterlagen und fordert fehlende Angaben nach. Für ein Pilotprojekt ist diese Beschreibung noch zu breit. Eine geeignete Aufgabe wäre etwa: „Bei eingehenden Meldungen zu Sachschäden fehlende Pflichtangaben erkennen und einen Entwurf für die Nachforderung erstellen.“ Eingabe, Ergebnis und die weiterarbeitende Person sind damit klar.

Hilfreich ist, diese Beschreibung wie eine Anforderung zu behandeln. Sie wird mit den beteiligten Stakeholdern abgestimmt, schriftlich festgehalten und nur bewusst geändert.

Auswahlkriterien

Oft stehen mehrere Aufgaben zur Wahl. Die folgenden Kriterien helfen, die geeignetste für ein Pilotprojekt zu finden. Geeignet heißt dabei nicht, dass die Aufgabe den größten denkbaren Nutzen verspricht. Geeignet heißt, dass das Pilotprojekt eine klare Antwort liefern kann.

Nutzen

Die Aufgabe sollte einen spürbaren Anteil an Zeit, Wartezeit oder Fehlern verursachen. Fragen Sie, was sich für die beteiligten Menschen und für die nachgelagerten Schritte ändert, wenn die Aufgabe schneller oder zuverlässiger erledigt wird. Ein Nutzen, der sich nur in einer Kostenrechnung zeigt, überzeugt die Beteiligten selten.

Machbarkeit

Die nötigen Eingaben müssen digital vorliegen oder mit vertretbarem Aufwand digitalisiert werden können. Die Systeme, aus denen die Daten kommen und in die das Ergebnis fließt, sollten erreichbar sein. Prüfen Sie auch, ob die Aufgabe genügend häufig anfällt, um im Pilotzeitraum ausreichend Fälle zu sammeln.

Risiko und Fehlertoleranz

Wie schwer wiegt ein fehlerhaftes Ergebnis? Lässt es sich rückgängig machen, und wie schnell wird es bemerkt? Für ein erstes Pilotprojekt eignen sich Aufgaben, bei denen ein Fehler begrenzte Folgen hat und vor der Weiterverwendung erkannt werden kann. Aufgaben, die Entscheidungen über Personen betreffen, Zahlungen auslösen oder nach außen kommunizieren, verlangen mehr Kontrolle und einen längeren Vorlauf.

Messbarkeit

Das Ergebnis der Aufgabe muss sich beurteilen lassen. Am einfachsten ist das, wenn es eine richtige Antwort gibt, etwa eine Kategorie oder einen Betrag. Schwieriger ist es bei Texten wie Zusammenfassungen oder Antwortentwürfen. Dort braucht es vorab vereinbarte Bewertungsmaßstäbe und Personen, die sie anwenden.

Verantwortung und Rückhalt

Für die Aufgabe braucht es eine prozessverantwortliche Person, die das Pilotprojekt fachlich trägt und am Ende mitentscheidet. Ebenso wichtig ist ein Team, das bereit ist, im Pilotbetrieb mitzuarbeiten. Eine Aufgabe ohne diese Voraussetzungen ist für ein Pilotprojekt ungeeignet, auch wenn sie fachlich vielversprechend ist.

KI oder klassische Automatisierung

Nicht jede Aufgabe braucht KI. Wenn Regeln klar und Eingaben vorhersehbar sind, ist eine klassische Automatisierung häufig einfacher zu betreiben und zu prüfen. Sie verhält sich bei gleicher Eingabe gleich, und ihre Logik lässt sich lesen.

KI kann sinnvoll sein, wenn unstrukturierte Inhalte verarbeitet oder variable Situationen eingeordnet werden müssen: freie Texte, Dokumente unterschiedlicher Herkunft, Anfragen in verschiedener Formulierung. Häufig ist eine Kombination am besten. Die KI liest und ordnet ein, die nachfolgenden Schritte laufen regelbasiert.

Zerlegen Sie die Aufgabe deshalb in Schritte und prüfen Sie für jeden Schritt, ob ein Modell, eine Regel oder ein Mensch ihn am besten erledigt. Das Pilotprojekt konzentriert sich dann auf die Schritte, bei denen der Einsatz von KI tatsächlich eine offene Frage ist.

Bewertungsmatrix

Wenn mehrere Aufgaben zur Wahl stehen, hilft eine einfache Matrix. Bewerten Sie jede Aufgabe je Kriterium auf einer Skala von 1 bis 3 und gewichten Sie die Kriterien nach den Prioritäten Ihrer Organisation. Die Gewichtung legen Sie vor der Bewertung fest, damit sie nicht an ein gewünschtes Ergebnis angepasst wird.

Kriterium	Leitfrage	1 = gering	2 = mittel	3 = hoch	Gewichtung
Nutzen	Wie stark entlastet eine bessere Lösung die Beteiligten und die Folgeschritte?	Kaum spürbar	Spürbar für ein Team	Spürbar über mehrere Schritte	festlegen
Machbarkeit	Liegen Eingaben digital vor, und sind die Systeme erreichbar?	Daten fehlen oder sind schwer zugänglich	Mit Aufwand erreichbar	Verfügbar und zugänglich	festlegen
Fehlertoleranz	Wie begrenzt und erkennbar sind die Folgen eines Fehlers?	Schwere, schwer erkennbare Folgen	Begrenzte Folgen, mit Prüfung erkennbar	Geringe Folgen, leicht korrigierbar	festlegen
Messbarkeit	Lässt sich das Ergebnis eindeutig beurteilen?	Nur subjektiv	Mit vereinbarten Maßstäben	Eindeutig richtig oder falsch	festlegen
Fallaufkommen	Fallen im Pilotzeitraum genügend Fälle an?	Selten	Regelmäßig	Häufig	festlegen
Verantwortung	Gibt es eine prozessverantwortliche Person und ein mitwirkendes Team?	Offen	Benannt, aber wenig verfügbar	Benannt und verfügbar	festlegen

Die Matrix ersetzt keine Entscheidung. Sie macht aber sichtbar, warum eine Aufgabe gewählt wurde, und sie zeigt, wo Schwächen liegen, die im Pilotprojekt besonders beobachtet werden sollten. Eine Aufgabe mit hohem Nutzen und geringer Messbarkeit kann trotzdem die richtige sein, wenn vorab geklärt wird, wie die Ergebnisse beurteilt werden.

Umfang eingrenzen

Grenzen Sie den Umfang bewusst ein. Ein Pilotprojekt für einen Prozessschritt in einem Team liefert klarere Erkenntnisse als eines, das mehrere Abläufe zugleich verändern soll. Legen Sie ausdrücklich fest, was nicht dazugehört: weitere Dokumentarten, weitere Standorte, weitere Sprachen, angrenzende Prozessschritte.

Diese Abgrenzung schützt das Pilotprojekt vor stillem Wachstum. Sie erleichtert außerdem die Kommunikation. Wer fragt, warum eine naheliegende Erweiterung fehlt, bekommt eine klare Antwort: Sie ist bewusst ausgeschlossen und kann nach der Entscheidung geprüft werden.

ARBEITSBLATT: AUFGABENSTECKBRIEF

Feld	Inhalt
Bezeichnung der Aufgabe	…
Anwendungsfall, zu dem die Aufgabe gehört	…
Heutiger Ablauf in wenigen Sätzen	…
Eingaben (Form, Herkunft, System)	…
Ergebnis (Form, Empfänger)	…
Wer arbeitet mit dem Ergebnis weiter	…
Häufigkeit und Schwankung des Aufkommens	…
Folgen eines fehlerhaften Ergebnisses	…
Schritte für KI, Regel und Mensch	…
Nicht im Umfang	…
Prozessverantwortliche Person	…
Mitwirkendes Team	…
Bewertung in der Matrix	…
Offene Fragen	…

Auf einen Blick

- Ein Anwendungsfall ist noch keine Aufgabe. Beschreiben Sie Eingaben, Ergebnis und Beteiligte so genau, dass Außenstehende sie verstehen.
- Wählen Sie die Aufgabe, bei der das Pilotprojekt eine klare Antwort liefern kann, nicht zwingend die mit dem größten denkbaren Nutzen.
- Prüfen Sie je Schritt, ob ein Modell, eine Regel oder ein Mensch ihn am besten erledigt.
- Halten Sie schriftlich fest, was nicht zum Umfang gehört.

KAPITEL 03

Entscheidungsfrage und Kriterien

Von der KI-Idee zum Pilotprojekt. Ein Leitfaden

Bevor das Pilotprojekt beginnt, steht fest, welche Frage es beantworten soll und woran die Antwort gemessen wird. Das klingt selbstverständlich. In der Praxis werden Kriterien trotzdem häufig erst nach den ersten Ergebnissen formuliert und dann unbewusst an diese angepasst.

Dieses Kapitel beschreibt, wie Sie die Entscheidungsfrage formulieren, den heutigen Ablauf als Vergleichspunkt erfassen und Erfolgs- sowie Abbruchkriterien festlegen, die sich prüfen lassen.

Die eine Frage

Ein Pilotprojekt beantwortet eine Frage. Sie richtet sich an die Person oder das Gremium, das am Ende entscheidet, und sie lässt sich mit den drei möglichen Ergebnissen beantworten: skalieren, überarbeiten oder bewusst stoppen.

Eine gute Entscheidungsfrage benennt die Aufgabe, den Vergleichsmaßstab und die Bedingungen des Betriebs. Ein Beispiel: „Kann die Lösung bei Sachschadenmeldungen fehlende Pflichtangaben so zuverlässig erkennen, dass der Prüfaufwand gegenüber heute sinkt, ohne dass mehr unvollständige Fälle in die Bearbeitung gelangen, und lässt sie sich im bestehenden System betreiben?“

Schwache Fragen erkennen Sie an ihrer Offenheit. „Was kann KI im Schadenprozess leisten?“ lässt sich nicht beantworten. „Funktioniert das Modell?“ ist eine technische Frage, die in einen Proof of Concept gehört. „Sind die Nutzerinnen zufrieden?“ greift einen Aspekt heraus und blendet Aufwand und Risiko aus.

Halten Sie die Frage schriftlich fest und stimmen Sie sie mit der entscheidenden Person ab. Wenn diese Person die Frage nicht als ihre eigene erkennt, wird sie am Ende auch der Antwort nicht folgen.

Den heutigen Ablauf erfassen

Ohne Ausgangspunkt fehlt der Vergleich. Bevor das Pilotprojekt beginnt, erfassen Sie deshalb, wie die Aufgabe heute erledigt wird. Diese Erfassung heißt Baseline. Sie muss nicht perfekt sein, aber sie muss auf die gleiche Weise gemessen werden wie später die Lösung.

Typische Größen der Baseline sind:

- **Aufwand je Fall:** die Arbeitszeit, die eine Person für einen Fall aufwendet.
- **Durchlaufzeit:** die Zeit vom Eingang bis zum fertigen Ergebnis, einschließlich Wartezeiten.
- **Fehlerquote:** der Anteil der Fälle, in denen das Ergebnis später korrigiert werden muss.
- **Nacharbeit:** Rückfragen, Nachforderungen und Korrekturen in nachgelagerten Schritten.
- **Aufkommen:** die Zahl der Fälle je Zeiteinheit und ihre Schwankung.

Für die Erfassung eignen sich drei Wege. Erstens eine Auswertung vorhandener Systemdaten, etwa Zeitstempel in einem Ticket- oder Vorgangssystem. Zweitens eine kurze, begleitete Zeiterfassung über einen festgelegten Zeitraum. Drittens eine nachträgliche Prüfung einer Stichprobe bereits erledigter Fälle, um die Fehlerquote zu bestimmen.

Achten Sie darauf, dass die Baseline die gleiche Fallmischung abbildet, die auch im Pilotprojekt vorkommt. Wird die Baseline in einer ruhigen Phase gemessen und das Pilotprojekt in einer Spitzenzeit, ist der Vergleich verzerrt.

Die Baseline hat einen zweiten Nutzen. Sie zeigt, wie gut der heutige Ablauf tatsächlich ist. Manchmal stellt sich dabei heraus, dass auch Menschen bei der Aufgabe häufiger voneinander abweichen als angenommen. Das ist eine wichtige Erkenntnis für die Frage, welche Qualität von der Lösung realistisch zu erwarten ist.

Erfolgskriterien formulieren

Erfolgskriterien werden vor dem Start festgelegt. Sie sollten prüfbar sein, ähnlich wie Akzeptanzkriterien in der Softwareentwicklung, und in der Sprache des Fachbereichs formuliert. Ein gutes Kriterium beschreibt, was gemessen wird, wie es gemessen wird, welcher Wert erreicht werden soll und woher die Daten kommen.

Qualität des Ergebnisses

Wie oft ist das Ergebnis richtig, vollständig und verwendbar? Messen Sie Qualität im Vergleich zum heutigen Ablauf, nicht gegen ein abstraktes Ideal. Unterscheiden Sie nach Fehlerarten. Ein übersehener Pflichtfall wiegt meist schwerer als ein unnötiger Hinweis. Legen Sie fest, welche Fehlerarten wie gewichtet werden.

Prüfaufwand

Wie viel Zeit kostet die menschliche Prüfung der Ergebnisse? Dieses Kriterium wird oft unterschätzt. Eine Lösung, deren Ergebnisse vollständig nachgeprüft werden müssen, spart weniger Zeit als erwartet. Messen Sie deshalb den Gesamtaufwand je Fall, einschließlich Prüfung und Korrektur.

Akzeptanz

Arbeiten die Menschen mit der Lösung, und vertrauen sie ihr in angemessenem Maß? Akzeptanz lässt sich über kurze, wiederholte Befragungen erfassen, aber auch über Beobachtung: Wie oft wird das Ergebnis übernommen, wie oft verworfen, wie oft umgangen? Beides zusammen ergibt ein verlässlicheres Bild als eine Befragung allein.

Betrieb und Kosten

Lässt sich die Lösung im vorgesehenen Umfeld betreiben, und steht der erwartete laufende Aufwand in einem vertretbaren Verhältnis zum Nutzen? Dieses Kriterium lässt sich im Pilotprojekt nur teilweise messen. Erfassen Sie zumindest den Verbrauch je Fall, den Betreuungsaufwand und die offenen Voraussetzungen für den Betrieb.

Abbruchkriterien

Ebenso wichtig wie Erfolgskriterien sind Abbruchkriterien. Sie legen fest, unter welchen Bedingungen das Pilotprojekt vorzeitig endet. Ohne sie läuft ein Pilotprojekt auch dann weiter, wenn früh erkennbar ist, dass es seine Frage nicht beantworten kann.

Typische Abbruchbedingungen sind:

- Die nötigen Daten sind nicht oder nicht in ausreichender Qualität verfügbar.
- Fehler einer kritischen Art treten auf und werden nicht zuverlässig erkannt.
- Datenschutz, Informationssicherheit oder Recht untersagen die vorgesehene Verarbeitung.
- Das Fallaufkommen reicht nicht aus, um die Kriterien zu prüfen.
- Die mitwirkenden Personen stehen nicht mehr zur Verfügung.

Ein Abbruch ist keine Niederlage. Er beantwortet die Entscheidungsfrage früher als geplant und spart Aufwand. Damit er tatsächlich stattfindet, muss vorab feststehen, wer ihn feststellt und wer ihn beschließt.

Beispielkatalog

Die folgende Tabelle zeigt, wie ein Kriterienkatalog aufgebaut sein kann. Die Schwellen sind bewusst offen gelassen. Sie hängen von der Aufgabe, der Baseline und der Risikobereitschaft Ihrer Organisation ab und werden vor dem Start gemeinsam festgelegt.

Kriterium	Messung	Schwelle	Quelle
Erkennungsrate kritischer Fälle	Anteil der Fälle mit fehlender Pflichtangabe, die die Lösung erkennt, geprüft an einem Referenzset	festlegen, mindestens auf Niveau der Baseline	Referenzset, fachliche Prüfung
Unnötige Hinweise	Anteil der Hinweise, die sich bei Prüfung als unbegründet erweisen	festlegen	Prüfprotokoll
Gesamtaufwand je Fall	Arbeitszeit einschließlich Prüfung und Korrektur, verglichen mit der Baseline	festlegen	Zeiterfassung, Systemdaten

Kriterium	Messung	Schwelle	Quelle
Durchlaufzeit	Zeit vom Eingang bis zur Nachforderung	festlegen	Zeitstempel im Vorgangssystem
Übernahme der Ergebnisse	Anteil der Ergebnisse, die ohne oder mit geringer Änderung übernommen werden	festlegen	Protokoll der Lösung
Akzeptanz	Einschätzung der mitwirkenden Personen in wiederholten Kurzbefragungen	festlegen	Befragung
Verbrauch je Fall	Kosten für Modellaufrufe und Infrastruktur je Fall	festlegen	Abrechnung, Monitoring
Abbruch: kritische Fehler	Kritische Fehler, die in der Prüfung nicht erkannt wurden	festlegen, Auftreten führt zur Prüfung des Abbruchs	Stichprobe, Rückmeldungen nachgelagerter Stellen
Abbruch: Datenverfügbarkeit	Anteil der Fälle, bei denen benötigte Daten fehlen	festlegen	Protokoll der Lösung

Kriterien abstimmen und festschreiben

Die Kriterien werden mit der prozessverantwortlichen Person, der entscheidenden Person und den Vertreterinnen und Vertretern von IT, Informationssicherheit und Datenschutz abgestimmt. Danach werden sie festgeschrieben, mit Datum und Version.

Änderungen während des Pilotprojekts sind möglich, aber nicht beiläufig. Wer eine Schwelle ändern will, begründet das, dokumentiert es und holt die Zustimmung der entscheidenden Person ein. So bleibt nachvollziehbar, woran das Pilotprojekt am Ende gemessen wird.

Auf einen Blick

- Formulieren Sie eine Entscheidungsfrage, die Aufgabe, Vergleichsmaßstab und Betriebsbedingungen benennt.
- Erfassen Sie den heutigen Ablauf auf die gleiche Weise, wie Sie später die Lösung messen.
- Kriterien beschreiben Messung, Schwelle und Quelle. Die Schwellen legen Sie vor dem Start fest.
- Abbruchkriterien sind ebenso verbindlich wie Erfolgskriterien.
- Änderungen an Kriterien werden begründet, dokumentiert und bewusst freigegeben.

KAPITEL 04

Daten, Recht und Sicherheit

Von der KI-Idee zum Pilotprojekt. Ein Leitfaden

Ein Pilotprojekt ist nur so aussagekräftig wie die Fälle, mit denen es arbeitet. Gleichzeitig ist es nur so tragfähig wie die Klärung, ob diese Daten genutzt werden dürfen und wie sie geschützt sind. Beide Fragen gehören an den Anfang.

Wer Fragen zu Informationssicherheit und Datenschutz erst am Ende stellt, riskiert, dass ein Pilotprojekt kurz vor dem Betrieb stecken bleibt. Dieses Kapitel beschreibt, was vor dem Start geklärt sein sollte, und gibt einen Überblick über die Risikologik des EU AI Act. Die rechtliche Bewertung im Einzelfall bleibt bei Ihren Rechtsexpertinnen und Rechtsexperten sowie bei Ihrer Datenschutzbeauftragten oder Ihrem Datenschutzbeauftragten.

Der Datencheck

Prüfen Sie vor dem Start drei Dinge: ob die Daten verfügbar sind, ob sie die nötige Qualität haben und ob Sie sie für diesen Zweck nutzen dürfen.

Verfügbarkeit

Wo liegen die Daten, in welchem System und in welcher Form? Wer kann den Zugriff freigeben, und wie lange dauert das erfahrungsgemäß? Lassen sich die Daten automatisiert abrufen, oder braucht es manuelle Exporte? Manuelle Zwischenschritte sind im Pilotprojekt vertretbar, müssen aber als offene Voraussetzung für den Betrieb festgehalten werden.

Qualität

Sind die Daten vollständig, aktuell und einheitlich? Wie hoch ist der Anteil schlecht lesbarer Dokumente, fehlender Felder oder widersprüchlicher Angaben? Ziehen Sie eine Stichprobe echter Fälle und sehen Sie sich diese gemeinsam mit den Menschen an, die heute damit arbeiten. Sie kennen die Eigenheiten der Daten besser als jede Dokumentation.

Nutzungsrechte

Dürfen die Daten für diesen Zweck verarbeitet werden? Das betrifft personenbezogene Daten ebenso wie vertraulich eingestufte Geschäftsinformationen, Daten von Vertragspartnern und Inhalte, an denen Dritte Rechte halten. Prüfen Sie auch, ob Verträge mit Kunden oder Lieferanten die Weitergabe an externe Dienste einschränken.

Prüfpunkt	Leitfrage	Ergeb-nis	Zu-ständig
Quelle	Aus welchen Systemen kommen die Daten?	…	…
Zugriff	Wer gibt den Zugriff frei, und ist er für das Pilotprojekt beantragt?	…	…

Prüfpunkt	Leitfrage	Ergeb-nis	Zu-ständig
Form	Liegen die Daten strukturiert, als Dokument oder als freier Text vor?	…	…
Vollständigkeit	Wie häufig fehlen Angaben, die für die Aufgabe nötig sind?	…	…
Repräsentativität	Bildet die Stichprobe die tatsächliche Fallmischung ab?	…	…
Personenbezug	Enthalten die Daten personenbezogene oder besonders schutzwürdige Angaben?	…	…
Vertraulichkeit	Wie sind die Daten nach der internen Klassifizierung eingestuft?	…	…
Vertragliche Einschränkungen	Gibt es Verträge, die die Verarbeitung durch Dritte einschränken?	…	…
Aufbewahrung	Wie lange dürfen Eingaben, Ergebnisse und Protokolle gespeichert werden?	…	…

Datenschutz

Sobald personenbezogene Daten verarbeitet werden, gilt die Datenschutz-Grundverordnung. Für ein Pilotprojekt sind vor allem folgende Fragen zu klären, gemeinsam mit Ihrer Datenschutzorganisation:

- **Zweck und Rechtsgrundlage:** Auf welcher Grundlage werden die Daten verarbeitet, und deckt der ursprüngliche Erhebungszweck die neue Verarbeitung ab?
- **Datenminimierung:** Braucht die Lösung tatsächlich alle Angaben, oder lassen sich Felder weglassen, kürzen oder pseudonymisieren?
- **Auftragsverarbeitung:** Wird ein externer Dienst eingesetzt, braucht es die entsprechenden Verträge. Klären Sie auch, ob Eingaben vom Anbieter gespeichert oder für eigene Zwecke verwendet werden.
- **Verarbeitungsort:** Wo werden die Daten verarbeitet und gespeichert, und findet eine Übermittlung in Drittländer statt?
- **Folgenabschätzung:** Bei voraussichtlich hohem Risiko für die betroffenen Personen kann eine Datenschutz-Folgenabschätzung erforderlich sein.
- **Automatisierte Entscheidungen:** Trifft die Lösung Entscheidungen mit rechtlicher oder ähnlich erheblicher Wirkung für Personen, gelten besondere Anforderungen. Menschliche Prüfung ist dann nicht nur eine Frage der Qualität.

- **Betroffenenrechte und Transparenz:** Wie werden betroffene Personen informiert, und wie werden Auskunfts- oder Löschanfragen auch für Protokolle und Zwischenergebnisse erfüllt?

Für die Vorbereitung sind sensible Daten oft nicht nötig. Anonymisierte oder eigens erstellte Beispiele reichen, um Aufgabe, Ablauf und technische Anbindung zu klären. Echte Daten kommen erst zum Einsatz, wenn die Verarbeitung freigegeben ist.

Informationssicherheit

Eine KI-Lösung ist ein neues System mit neuen Schnittstellen. Sie unterliegt den gleichen Anforderungen wie jede andere Anwendung, bringt aber einige eigene Fragen mit. Klären Sie mit Ihrer Informationssicherheit mindestens:

- **Anbieter und Betriebsmodell:** Wird ein Modell über einen externen Dienst genutzt, in einer eigenen Cloud-Umgebung oder lokal betrieben? Welche Zusagen macht der Anbieter zu Speicherung, Zugriff und Sicherheit?
- **Zugriffe und Berechtigungen:** Auf welche Systeme und Daten greift die Lösung zu? Erhält sie nur die Rechte, die sie für die Aufgabe braucht? Das gilt besonders, wenn die Lösung Werkzeuge aufruft oder Daten verändert.
- **Eingaben aus unsicheren Quellen:** E-Mails, hochgeladene Dokumente und Webinhalte können Anweisungen enthalten, die das Modell zu unerwünschtem Verhalten verleiten. Legen Sie fest, wie solche Eingaben behandelt und welche Handlungen ohne Freigabe ausgeschlossen werden.
- **Ausgaben:** Kann die Lösung vertrauliche Inhalte an Personen weitergeben, die sie nicht sehen dürfen? Wie wird verhindert, dass Ergebnisse ungeprüft nach außen gelangen?
- **Protokollierung:** Was wird protokolliert, wo und wie lange? Protokolle sind für Nachvollziehbarkeit nötig, enthalten aber oft selbst schützenswerte Daten.
- **Einbindung in bestehende Prozesse:** Wie ist die Lösung an Identitätsverwaltung, Schwachstellenmanagement und Notfallprozesse angebunden?

Im Pilotprojekt dürfen einzelne Anforderungen vereinfacht umgesetzt werden, wenn das Risiko begrenzt ist und die Vereinfachung dokumentiert wird. Entscheidend ist, dass die Anforderungen für den Betrieb bekannt sind und in die Bewertung einfließen.

Der EU AI Act im Überblick

Der EU AI Act, die Verordnung (EU) 2024/1689 in ihrer jeweils geltenden Fassung, knüpft Pflichten an die Rolle eines Unternehmens und an Art und Risiko der Anwendung. Anbieter entwickeln beispielsweise KI-Systeme und bringen sie in Verkehr; Betreiber setzen sie in eigener

Verantwortung ein. Die folgende Übersicht dient der ersten Orientierung. Ihre Rechtsexperten prüfen die Einordnung und die geltenden Pflichten im Einzelfall.

Stand 2. Oktober 2026: Nach dem Umsetzungszeitplan der Europäischen Kommission gelten Verbote bestimmter KI-Praktiken bereits seit 2. Februar 2025. Die Transparenzregeln gelten grundsätzlich seit 2. August 2026, mit Übergangsfristen für bestimmte Bestandsangebote. Die Hochrisikoregeln gelten für Systeme nach Anhang III ab 2. Dezember 2027 und für produktgebundene Systeme nach Anhang I ab 2. August 2028. Diese Auswahl wichtiger Termine ersetzt keine Prüfung der jeweiligen Übergangsregeln und Ausnahmen. Quelle: Europäische Kommission, AI Act Service Desk, „Timeline for the Implementation of the EU AI Act“, abgerufen am 2. Oktober 2026.

Risiko-klasse	Kennzeichen	Folge im Überblick
Verbo-tene Prakti-ken	Anwendungen, die der Gesetzgeber als unvertretbar einstuft, etwa bestimmte Formen der Manipulation oder der Bewertung von Personen nach sozialem Verhalten	Einsatz ist unzulässig
Hohes Risiko	Anwendungen in festgelegten Bereichen, etwa bestimmte Einsätze in Personalwesen, Zugang zu wesentlichen Leistungen oder als Sicherheitsbau-teil von Produkten	Umfangreiche Anforderungen, unter ande-rem an Risikomanagement, Datenqualität, Dokumentation, Protokollierung und men-schliche Aufsicht
Trans-parenz pflich-ten	Anwendungen, die mit Menschen interagieren oder Inhalte erzeugen	Offenlegung, dass KI im Spiel ist, oder Kennzeichnung erzeugter Inhalte
Gerin-ges Risiko	Die meisten übrigen Anwendungen	Keine spezifischen Pflichten aus der Risikologik, allgemeine Anforderungen bleiben bestehen

Für das Pilotprojekt folgen daraus drei praktische Schritte. Erstens nehmen Sie die geplante Anwendung in ein Inventar der KI-Anwendungen auf, sofern es eines gibt, oder legen den ersten Eintrag an. Zweitens nehmen Sie eine erste Einordnung nach Rolle und Risikoklasse vor und lassen sie von Ihren Rechtsexperten bestätigen. Drittens berücksichtigen Sie die daraus folgenden Anforderungen bereits in der Gestaltung des Pilotprojekts. Anforderungen an Protokollierung und menschliche Aufsicht lassen sich schwer nachrüsten.

Unabhängig von der rechtlichen Einordnung brauchen die beteiligten Personen Kenntnisse über Einsatzgrenzen, Fehlerquellen und menschliche Kontrolle. Schulen Sie sie deshalb vor dem Pilotbetrieb anhand der konkreten Aufgabe. Welche rechtlichen Schulungs- und Kompetenzanforderungen für Ihre Rolle und Anwendung gelten, prüfen Ihre Rechtsexperten anhand der aktuellen Rechtslage.

Früh einbinden

Datenschutz, Informationssicherheit und Recht werden nicht am Ende um Zustimmung gebeten, sondern am Anfang um Mitwirkung. Das hat einen praktischen Grund. Ihre Anforderungen bestimmen Architektur, Anbieterwahl und Kontrollpunkte. Werden sie spät gestellt, müssen diese Entscheidungen neu getroffen werden.

Bewährt hat sich ein kurzer, gemeinsamer Termin vor dem Start. Dort werden Aufgabe, Datenfluss und vorgesehene Anbieter vorgestellt, offene Fragen gesammelt und Zuständigkeiten für ihre Klärung vereinbart. Ergebnis ist eine Liste mit Freigaben, die vor dem Einsatz echter Daten vorliegen müssen, und eine Liste mit Anforderungen, die erst für den Betrieb erfüllt sein müssen.

> Informationssicherheit und Datenschutz gehören an den Anfang. Spät gestellte Fragen halten ein Pilotprojekt genau dann auf, wenn es in den Betrieb gehen soll.

Auf einen Blick

- Prüfen Sie Verfügbarkeit, Qualität und Nutzungsrechte der Daten, bevor Sie Kriterien endgültig festlegen.
- Klären Sie Zweck, Rechtsgrundlage, Auftragsverarbeitung und Verarbeitungsort mit Ihrer Datenschutzorganisation.
- Behandeln Sie Eingaben aus unsicheren Quellen und die Berechtigungen der Lösung als eigene Sicherheitsfragen.
- Ordnen Sie die Anwendung früh nach den Risikoklassen des EU AI Act ein. Die rechtliche Bewertung treffen Ihre Rechtsexperten.
- Arbeiten Sie in der Vorbereitung mit anonymisierten oder eigens erstellten Beispielen.

KAPITEL 05

Mensch und Verantwortung

Von der KI-Idee zum Pilotprojekt. Ein Leitfaden

Menschliche Kontrolle ist Teil des Ablaufs, keine nachträgliche Absicherung. Sie wird gemeinsam mit der Lösung gestaltet und bestimmt Risiko und Nutzen gleichermaßen. Zu viel Kontrolle frisst die Zeitersparnis, zu wenig lässt Fehler unbemerkt durch.

Dieses Kapitel beschreibt drei Grundmuster menschlicher Kontrolle, wie Sie das passende wählen und wie Sie Verantwortung im Pilotprojekt eindeutig verteilen.

Drei Muster der Kontrolle

Mensch prüft jedes Ergebnis

Die Lösung erstellt einen Vorschlag, ein Mensch prüft ihn und gibt ihn frei, ändert ihn oder verwirft ihn. Kein Ergebnis wird ohne Freigabe weiterverwendet. Dieses Muster eignet sich für den Beginn eines Pilotprojekts, für Aufgaben mit schweren Fehlerfolgen und für alle Schritte, die nach außen wirken. Der Nutzen liegt dann in schnellerer Vorbereitung, nicht im Wegfall der Prüfung.

Mensch prüft eine Stichprobe

Die Ergebnisse werden weiterverwendet, ein festgelegter Anteil wird nachträglich geprüft. Dieses Muster eignet sich, wenn Fehler begrenzte Folgen haben, sich später korrigieren lassen und die Qualität über einen längeren Zeitraum stabil war. Die Stichprobe muss zufällig gezogen werden und groß genug sein, um Veränderungen zu erkennen.

Mensch prüft Ausnahmen

Die Lösung bearbeitet eindeutige Fälle selbstständig und leitet unsichere, ungewöhnliche oder regelwidrige Fälle an einen Menschen weiter. Dieses Muster verspricht den größten Nutzen, verlangt aber die verlässlichste Erkennung von Ausnahmen. Es setzt voraus, dass die Lösung ihre Unsicherheit belastbar anzeigt oder dass klare Regeln festlegen, wann ein Fall weitergeleitet wird.

Muster	Geeignet, wenn	Nutzen entsteht durch	Worauf zu achten ist
Jedes Ergebnis	Fehlerfolgen schwer, Wirkung nach außen, frühe Phase	Schnellere Vorbereitung, einheitlichere Ergebnisse	Prüfaufwand kann die Zeitersparnis aufzehren. Gewöhnung führt zu oberflächlicher Prüfung.
Stichprobe	Fehlerfolgen begrenzt und korrigierbar, Qualität stabil belegt	Wegfall der Einzelprüfung	Stichprobe muss zufällig und ausreichend sein. Fehler wirken bis zur Entdeckung.

Muster	Geeignet, wenn	Nutzen entsteht durch	Worauf zu achten ist
Ausnahmen	Eindeutige Fälle klar abgrenzbar, Unsicherheit verlässlich erkennbar	Menschen bearbeiten nur schwierige Fälle	Falsch eingeordnete Ausnahmen bleiben unbemerkt. Weiterleitungsregeln brauchen eigene Prüfung.

Das passende Muster wählen

Maßgeblich sind drei Fragen: Wie schwer wiegen die Folgen eines Fehlers? Lässt er sich rückgängig machen? Wie schnell wird er bemerkt? Lesende Schritte wie Einordnen oder Zusammenfassen lassen sich meist freier gestalten als Schritte, die Daten ändern, Zahlungen auslösen oder nach außen kommunizieren.

Beginnen Sie im Pilotprojekt in der Regel mit dem strengeren Muster. So entsteht die Datengrundlage, um später zu entscheiden, ob ein leichteres Muster vertretbar ist. Kontrolle lässt sich schrittweise lockern, wenn Auswertungen eine stabile Qualität zeigen. Legen Sie die Bedingungen dafür vorab fest.

Ein Hinweis zur Wirksamkeit: Wer über längere Zeit überwiegend richtige Vorschläge prüft, prüft mit der Zeit weniger genau. Diese Gewöhnung ist menschlich und lässt sich nicht durch Appelle beheben. Hilfreich sind klare Prüfpunkte statt einer allgemeinen Freigabe, eine Anzeige der Stellen, an denen die Lösung unsicher ist, und gelegentliche bekannte Testfälle, mit denen sich die Aufmerksamkeit der Prüfung selbst überprüfen lässt.

Verantwortung festlegen

Für jedes Ergebnis muss klar sein, wer es verantwortet. Die Lösung trägt keine Verantwortung. Verantwortlich bleibt die Person oder Stelle, die das Ergebnis freigibt oder deren Ablauf es nutzt. Klären Sie vor dem Start drei Fragen:

- Welche Ergebnisse dürfen ohne Prüfung weiterverwendet werden, und welche nicht?
- Wer ist verantwortlich, wenn ein fehlerhaftes Ergebnis unbemerkt bleibt?
- Wie fließen Korrekturen zurück, damit die Lösung verbessert werden kann?

Die Antworten gehören in den Pilotsteckbrief. Sie sind auch deshalb wichtig, weil sie für den Betrieb weitergelten. Eine Verantwortung, die im Pilotprojekt ungeklärt bleibt, wird im Betrieb selten nachträglich sauber geregelt.

Rollen im Pilotprojekt

Die folgende Übersicht verteilt die wichtigsten Aufgaben auf Rollen. Sie folgt der bekannten RACI-Logik: R steht für die Durchführung, A für die abschließende Verantwortung, C für die Beteiligung vor einer Entscheidung, I für die Information danach. Je Aufgabe gibt es genau eine Rolle mit A.

Aufgabe	Entscheider	Prozessverantwortung	Fachanwender	Projektleitung	IT und Architektur	Informationssicherheit	Datenschutz
Entscheidungsfrage festlegen	A	R	C	R	C	I	I
Kriterien und Schwellen festlegen	A	R	C	R	C	C	C
Daten und Zugriff freigeben	I	A	I	R	R	C	C
Kontrollmuster festlegen	C	A	C	R	C	C	C
Ergebnisse prüfen und korrigieren	I	A	R	I	I	I	I
Beobachtungen dokumentieren	I	C	R	A	R	I	I
Abbruch feststellen	A	R	C	R	C	C	C
Auswertung erstellen	C	C	C	A	R	C	C
Entscheidung treffen	A	R	I	C	C	C	C

Die Rollen können je nach Organisation zusammenfallen. Wichtig ist, dass jede Rolle mit einer Person besetzt ist, die für die Dauer des Pilotprojekts verfügbar ist. Eine kleine, feste Gruppe arbeitet verlässlicher als eine große, wechselnde.

Korrekturen zurückführen

Jede menschliche Korrektur ist eine Information über die Lösung. Sie zeigt, wo die Lösung irrt, welche Fälle schwierig sind und wo Anweisungen, Daten oder Regeln angepasst werden müssen. Diese Information geht verloren, wenn Korrekturen nur im Ergebnis landen.

Legen Sie deshalb fest, wie Korrekturen erfasst werden: mit dem ursprünglichen Ergebnis, der korrigierten Fassung und einer kurzen Kategorie für den Fehler. Das kostet wenige Sekunden je Fall, wenn es in die Oberfläche eingebaut ist, und liefert die wichtigste Grundlage für Auswertung und Verbesserung.

> Menschliche Kontrolle wird mit dem Ablauf gestaltet. Sie bestimmt Risiko und Nutzen gleichermaßen.

Auf einen Blick

- Wählen Sie zwischen Prüfung jedes Ergebnisses, Stichprobe und Ausnahmeprüfung nach Fehlerfolgen, Umkehrbarkeit und Erkennbarkeit.
- Beginnen Sie mit dem strengeren Muster und legen Sie vorab fest, unter welchen Bedingungen Kontrolle gelockert wird.
- Verantwortung für Ergebnisse liegt immer bei Menschen. Halten Sie fest, bei wem.
- Erfassen Sie Korrekturen strukturiert. Sie sind die wichtigste Grundlage für Auswertung und Verbesserung.

KAPITEL 06

Den Pilotbetrieb gestalten

Von der KI-Idee zum Pilotprojekt. Ein Leitfaden

Das Pilotprojekt läuft in einem festen Zeitraum, mit echten Fällen und mit den Menschen, die die Aufgabe heute erledigen. Die Vorbereitung aus den bisherigen Kapiteln bestimmt, was geprüft wird. Dieses Kapitel beschreibt, wie der Pilotbetrieb organisiert wird, damit am Ende belastbare Ergebnisse vorliegen.

Das Team

Zum Kernteam gehören die prozessverantwortliche Person, zwei bis wenige Fachanwenderinnen und Fachanwender, eine Projektleitung und die technische Umsetzung. Informationssicherheit, Datenschutz und die entscheidende Person sind nicht laufend beteiligt, aber zu festen Zeitpunkten eingebunden.

Vereinbaren Sie mit den Vorgesetzten der Fachanwender, wie viel Zeit diese für das Pilotprojekt aufwenden. Ein Pilotprojekt, das neben dem vollen Tagesgeschäft laufen soll, erhält die wenige Zeit, die übrig bleibt, und liefert entsprechend dünne Ergebnisse.

Zeitrahmen und Phasen

Der Zeitraum wird vor dem Start festgelegt. Er sollte so lang sein, dass genügend echte Fälle durchlaufen, um die Kriterien zu prüfen, und so kurz, dass Aufmerksamkeit und Beteiligung erhalten bleiben. Wie viele Fälle genügen, hängt von den Kriterien ab: Wer seltene, kritische Fehler erkennen will, braucht mehr Fälle als jemand, der eine durchschnittliche Bearbeitungszeit vergleicht. Eine Verlängerung ist möglich, braucht aber eine bewusste Entscheidung und eine Begründung.

Bewährt hat sich eine Gliederung in vier Phasen:

Phase	Ziel	Ergebnis	Zeitrahmen
Vorbereitung	Aufgabe, Kriterien, Daten, Freigaben und Referenzset klären, Lösung aufbauen	Pilotsteckbrief, Kriterienkatalog, Freigaben, einsatzbereite Lösung	einige Wochen, abhängig von Freigaben und Anbindung
Parallelbetrieb	Lösung läuft neben dem heutigen Ablauf, Ergebnisse werden verglichen, aber nicht verwendet	Vergleich mit dem heutigen Ablauf ohne Risiko für den Betrieb	einige Tage bis wenige Wochen
Begleiteter Echtbetrieb	Ergebnisse werden unter dem vereinbarten Kontrollmuster im Ablauf verwendet	Messwerte zu Qualität, Prüfaufwand und Akzeptanz unter realen Bedingungen	einige Wochen

Phase	Ziel	Ergebnis	Zeitrahmen
Auswer-tung	Ergebnisse gegen Kriterien stellen, Entscheidungsvorlage erstellen	Entscheidungsvorlage und Entscheidungstermin	wenige Wochen

Ob der Parallelbetrieb nötig ist, hängt vom Risiko ab. Bei Aufgaben mit schweren Fehlerfolgen kann das gesamte Pilotprojekt parallel laufen. Bei Aufgaben mit geringen Folgen kann der begleitete Echtbetrieb früh beginnen, solange jedes Ergebnis geprüft wird.

Echte und schwierige Testfälle

Verwenden Sie echte, repräsentative Fälle, auch die schwierigen und unvollständigen. Sorgfältig ausgewählte Beispiele lassen eine Lösung besser aussehen, als sie im Alltag sein wird.

Stellen Sie vor dem Start ein Referenzset zusammen. Das ist eine feste Sammlung von Fällen mit bekanntem richtigem Ergebnis, das Fachleute festgelegt haben. Es dient dazu, die Lösung vor und während des Pilotprojekts gleichbleibend zu prüfen, etwa nach jeder Änderung an Anweisungen oder Modell. Das Referenzset sollte enthalten:

- **gewöhnliche Fälle** in der Mischung, in der sie im Alltag vorkommen,
- **Grenzfälle**, bei denen auch erfahrene Personen nachdenken müssen,
- **unvollständige und fehlerhafte Eingaben**, etwa schlecht lesbare Dokumente oder fehlende Felder,
- **Fälle außerhalb des Umfangs**, bei denen die Lösung erkennen soll, dass sie nicht zuständig ist,
- **kritische Fälle**, bei denen ein Fehler schwere Folgen hätte,
- **Eingaben, die die Lösung zu unerwünschtem Verhalten verleiten könnten**, sofern die Aufgabe Inhalte aus unsicheren Quellen verarbeitet.

Die Menschen, die die Aufgabe heute erledigen, wissen am besten, welche Fälle schwierig sind. Bitten Sie sie gezielt darum, solche Fälle beizusteuern. Das Referenzset wird während des Pilotprojekts nicht still verändert. Neue Fälle kommen in eine Ergänzung, damit Vergleiche über die Zeit möglich bleiben.

Feedback und Zwischentermine

Vereinbaren Sie feste, kurze Zwischentermine in regelmäßigem Abstand. Dort werden drei Fragen geprüft: Bleiben die Kriterien erreichbar? Ist eine Abbruchbedingung eingetreten? Gibt es Beobachtungen, die eine Anpassung nahelegen?

Neben den Terminen braucht es einen einfachen Weg für laufende Rückmeldungen. Fachanwender sollten eine auffällige Beobachtung in wenigen Sekunden festhalten können, direkt in der Lösung oder in einem gemeinsamen Protokoll. Rückmeldungen, die erst im Zwischentermin aus dem Gedächtnis gesammelt werden, sind unvollständig.

Wird die Lösung während des Pilotprojekts verbessert, etwa durch geänderte Anweisungen, prüfen Sie jede Änderung zuerst am Referenzset. So wird sichtbar, ob eine Verbesserung an einer Stelle zu Verschlechterungen an anderer führt.

Dokumentation

Dokumentieren Sie laufend, nicht am Ende. Drei Protokolle reichen in der Regel aus:

- **Ergebnisprotokoll:** Eingabe, Ergebnis der Lösung, Prüfentscheidung, Korrektur und Fehlerkategorie je Fall. Es wird möglichst automatisch geführt.
- **Beobachtungsprotokoll:** auffällige, unerwartete oder wiederkehrende Beobachtungen, auch solche, die in keines der Kriterien passen.
- **Änderungsprotokoll:** jede Änderung an Lösung, Anweisungen, Modell, Daten, Kontrollmuster, Umfang oder Kriterien, mit Datum, Grund und Freigabe.

Diese Protokolle sind die Grundlage der Auswertung. Sie sind zugleich eine Vorübung für den Betrieb, in dem Nachvollziehbarkeit dauerhaft gebraucht wird.

Änderungen am Umfang

Änderungen am Umfang sind erlaubt, aber nicht beiläufig. Wird die Aufgabe während des Pilotprojekts erweitert oder verändert, entscheiden die Verantwortlichen das bewusst und passen die Kriterien an. Sonst misst das Pilotprojekt am Ende etwas anderes als geplant.

Ein Beispiel: Die Fachanwender schlagen vor, neben Sachschäden auch Meldungen einer weiteren Sparte aufzunehmen, weil die Lösung dort ebenfalls helfen könnte. Die Anregung ist berechtigt. Sie wird im Beobachtungsprotokoll festgehalten und für die Entscheidung vorgemerkt. Aufgenommen wird sie nur, wenn die entscheidende Person zustimmt und Kriterien, Referenzset und Zeitrahmen angepasst werden.

Kommunikation

Ein Pilotprojekt wird beobachtet, auch von Menschen, die nicht beteiligt sind. Erwartungen entstehen schnell, in beide Richtungen. Kommunizieren Sie deshalb vorab und sachlich: welche Aufgabe geprüft wird, warum, wie lange, wer beteiligt ist und dass am Ende eine Entscheidung steht, die auch ein Stopp sein kann.

Für die beteiligten Fachanwender ist eine Frage besonders wichtig: Was bedeutet das Pilotprojekt für ihre Arbeit? Beantworten Sie sie offen. Menschen, die befürchten, sich mit ihrer Mitwirkung selbst überflüssig zu machen, liefern keine ehrlichen Rückmeldungen. Wenn Mitbestimmungsgremien betroffen sind, beziehen Sie diese frühzeitig ein.

Auf einen Blick

- Legen Sie Zeitrahmen, Phasen und die Zeit der Fachanwender vor dem Start fest.
- Prüfen Sie mit einem festen Referenzset, das gewöhnliche, schwierige, unvollständige und kritische Fälle enthält.
- Halten Sie feste Zwischentermine und einen einfachen Weg für laufende Rückmeldungen bereit.
- Führen Sie Ergebnis-, Beobachtungs- und Änderungsprotokoll laufend.
- Ändern Sie den Umfang nur bewusst und passen Sie die Kriterien dann an.

KAPITEL 07

Auswerten und entscheiden

Von der KI-Idee zum Pilotprojekt. Ein Leitfaden

Das Pilotprojekt endet mit einem Entscheidungstermin, nicht mit einer Präsentation. Vorab steht fest, wer entscheidet und auf welcher Grundlage. Die Ergebnisse werden den vereinbarten Kriterien gegenübergestellt, einschließlich der Beobachtungen, die in keines der Kriterien passen.

Dieses Kapitel beschreibt, wie Sie die Auswertung aufbereiten, welche drei Ergebnisse möglich sind und wie Sie den Entscheidungstermin so gestalten, dass er tatsächlich zu einem Beschluss führt.

Gegen die Kriterien auswerten

Stellen Sie jedes Kriterium aus dem Katalog mit dem gemessenen Wert, der vereinbarten Schwelle und einer kurzen Einordnung dar. Verwenden Sie dafür drei Stufen: erfüllt, teilweise erfüllt, nicht erfüllt. Die Einordnung „teilweise“ braucht eine Begründung, etwa dass die Schwelle nur für einen Teil der Fälle erreicht wurde.

Achten Sie auf vier Punkte:

- **Vergleichbarkeit:** Messwerte der Lösung werden mit der Baseline verglichen, die auf gleiche Weise erhoben wurde. Unterschiede in Fallmischung oder Zeitraum werden benannt.
- **Aufschlüsselung:** Durchschnittswerte verdecken Probleme. Schlüsseln Sie Ergebnisse nach Fallarten auf, besonders nach kritischen Fällen und Grenzfällen.
- **Entwicklung über die Zeit:** Hat sich die Qualität im Verlauf verändert, etwa nach Änderungen an der Lösung? Ist die Akzeptanz gestiegen oder gesunken?
- **Unsicherheit:** Bei wenigen Fällen sind Messwerte unsicher. Benennen Sie das, statt Genauigkeit vorzutäuschen.

Beobachtungen außerhalb der Kriterien

Nicht alles Wichtige passt in einen Kriterienkatalog. Vielleicht hat sich gezeigt, dass ein vorgelagerter Schritt die eigentliche Ursache für Nacharbeit ist. Vielleicht haben Fachanwender die Lösung für eine Teilaufgabe genutzt, die nicht vorgesehen war. Vielleicht ist der Aufwand für eine Freigabe im Betrieb deutlich höher als gedacht.

Diese Beobachtungen gehören in die Auswertung, in einen eigenen Abschnitt. Sie ersetzen die Kriterien nicht und rechtfertigen keine nachträgliche Umdeutung. Sie können aber die Entscheidung beeinflussen, etwa ob ein Überarbeiten sinnvoller ist als ein Skalieren.

Die drei möglichen Ergebnisse

Drei Ergebnisse sind möglich, und jedes ist gleichwertig, wenn es begründet ist.

Skalieren

Die Kriterien sind erfüllt, und die Voraussetzungen für den Betrieb lassen sich schaffen. Skalieren bedeutet nicht, die Lösung sofort überall einzusetzen. Es bedeutet, den Übergang in den Betrieb zu beauftragen, mit Budget, Verantwortung und einem Plan für den Rollout. Offene Betriebsvoraussetzungen werden als Bedingungen benannt.

Überarbeiten

Der Ansatz ist tragfähig, aber Aufgabe, Daten oder Kontrollpunkte müssen angepasst werden. Daraus folgt ein neues, eng umrissenes Pilotprojekt oder eine Verlängerung mit geänderten Kriterien. Überarbeiten ist nur dann ein sinnvolles Ergebnis, wenn konkret benannt wird, was sich ändert und welche Frage danach beantwortet werden soll. Ein Überarbeiten ohne klare Änderung ist ein verdecktes Weiterlaufen.

Bewusst stoppen

Nutzen, Risiko und Aufwand passen nicht zusammen. Gründe und Erkenntnisse werden festgehalten, damit sie in künftige Bewertungen einfließen. Dazu gehören auch die Bedingungen, unter denen eine erneute Prüfung sinnvoll wäre, etwa bessere Datenqualität oder veränderte technische Möglichkeiten.

Gerade das bewusste Stoppen braucht den Rückhalt der Führung. Ohne ihn werden Pilotprojekte eher verlängert als beendet.

> Ein beendetes Pilotprojekt mit klarer Begründung ist mehr wert als eines, das ohne Entscheidung weiterläuft.

Der Entscheidungstermin

Der Entscheidungstermin ist ein Arbeitstermin mit einem kleinen Kreis. Teilnehmen sollten die entscheidende Person, die prozessverantwortliche Person, die Projektleitung und, je nach offenen Punkten, Vertreter von IT, Informationssicherheit und Datenschutz.

Die Entscheidungsvorlage wird einige Tage vorher verteilt. Der Termin selbst dient nicht der Vorstellung der Ergebnisse, sondern der Klärung offener Fragen und dem Beschluss. Am Ende stehen das gewählte Ergebnis, die Begründung, die nächsten Schritte mit Verantwortlichen und gegebenenfalls die Bedingungen, an die das Ergebnis geknüpft ist.

Die Entscheidungsvorlage folgt einer festen Gliederung:

Abschnitt	Inhalt	Leitfrage
Entscheidungsfrage	Die vor dem Start festgelegte Frage	Was sollte das Pilotprojekt beantworten?
Zusammenfassung	Empfehlung in wenigen Sätzen	Welches Ergebnis wird empfohlen, und warum?
Kriterien	Messwert, Schwelle und Einordnung je Kriterium	Wurden die Kriterien erfüllt?
Beobachtungen	Wichtige Erkenntnisse außerhalb der Kriterien	Was hat sich zusätzlich gezeigt?
Risiken	Offene Risiken aus Qualität, Sicherheit, Datenschutz und Recht	Was könnte im Betrieb schiefgehen?
Betriebsvoraussetzungen	Offene Punkte für den Betrieb, grobe Kostenarten	Was braucht es, um die Lösung zu betreiben?
Optionen	Skalieren, überarbeiten, bewusst stoppen, jeweils mit Folgen	Was bedeutet jede Option?
Beschluss	Gewähltes Ergebnis, Bedingungen, nächste Schritte	Was wird entschieden, und wer tut was?

Eine ausfüllbare Fassung finden Sie im Anhang.

Nach der Entscheidung

Halten Sie den Beschluss schriftlich fest und teilen Sie ihn mit allen Beteiligten, auch mit den Fachanwendern. Sie haben Zeit und Erfahrung eingebracht und sollten wissen, was daraus folgt.

Sichern Sie außerdem die Grundlagen: Referenzset, Protokolle, Kriterienkatalog und Auswertung. Bei einem Skalieren bilden sie den Ausgangspunkt für den Betrieb. Bei einem Stopp sind sie die Grundlage, um die Frage später mit weniger Aufwand erneut zu prüfen.

Auf einen Blick

- Stellen Sie jedes Kriterium mit Messwert, Schwelle und Einordnung dar und schlüsseln Sie nach Fallarten auf.
- Beobachtungen außerhalb der Kriterien gehören in die Auswertung, rechtfertigen aber keine nachträgliche Umdeutung.
- Skalieren, überarbeiten und bewusst stoppen sind gleichwertig, wenn sie begründet sind.
- Der Entscheidungstermin dient dem Beschluss, nicht der Präsentation.

KAPITEL 08

Vom Pilot in den Betrieb

Von der KI-Idee zum Pilotprojekt. Ein Leitfaden

Der Übergang in den Betrieb beginnt nicht nach dem Pilotprojekt, sondern in seiner Planung. Im Pilotprojekt wird vieles von Hand erledigt, das im Betrieb verlässlich und dauerhaft funktionieren muss: Datenabruf, Prüfung der Qualität, Freigabe von Änderungen, Betreuung der Nutzer.

Dieses Kapitel beschreibt die Voraussetzungen für den Betrieb, die technische Kontrollschicht um die Lösung, die Verteilung der Verantwortung, die laufenden Kostenarten und einen schrittweisen Rollout.

Voraussetzungen für den Betrieb

Wer skalieren will, sollte die folgenden Punkte schon während des Pilotprojekts klären:

- eine fachliche und eine technische Verantwortung für den laufenden Betrieb,
- die Anbindung an bestehende Systeme statt manueller Zwischenschritte,
- eine regelmäßige Prüfung, ob die Qualität der Ergebnisse stabil bleibt,
- Regeln dafür, wer Änderungen an der Lösung freigibt,
- Schulung und Begleitung der Menschen, die mit der Lösung arbeiten,
- eine realistische Sicht auf die laufenden Kosten.

Dazu kommen die Anforderungen aus Informationssicherheit, Datenschutz und gegebenenfalls dem EU AI Act, die im Pilotprojekt vereinfacht oder zurückgestellt wurden. Die Liste aus dem gemeinsamen Termin zu Beginn wird jetzt abgearbeitet.

Das AI Harness

Als AI Harness bezeichnen wir die technische Kontrollschicht um Modelle und Agenten. Sie verbindet organisatorische Vorgaben mit Kontrollen und Nachweisen im Betrieb. Ihre Wirksamkeit hängt von der Umsetzung und den geprüften Anwendungsfällen ab. Das Harness umfasst hier fünf Bausteine.

Leitplanken

Leitplanken prüfen Eingaben, Ausgaben und geplante Aktionen. Je nach Zweck kombinieren sie feste Regeln, Berechtigungsprüfungen und modellgestützte Bewertungen. Sie helfen, unerlaubte Datenweitergabe oder Aktionen zu begrenzen. Zugriffsrechte und Freigaben werden außerhalb des Sprachmodells durchgesetzt. Prüfen Sie die Wirksamkeit an normalen und gezielt manipulierten Fällen und dokumentieren Sie verbleibende Grenzen.

Evaluierungen

Evaluierungen prüfen die Qualität der Lösung systematisch an einem festen Satz von Fällen. Das Referenzset aus dem Pilotprojekt ist ihr Ausgangspunkt. Vor jeder Änderung an Modell, Anweisungen oder Daten wird evaluiert, und die Änderung wird nur freigegeben, wenn die Qualität nicht unter die vereinbarte Schwelle fällt. Das Referenzset wird im Betrieb um neue, typische und schwierige Fälle ergänzt.

Protokollierung

Die Protokollierung hält fest, welche Eingabe zu welchem Ergebnis geführt hat, mit welcher Version von Modell und Anweisungen und mit welcher menschlichen Prüfentscheidung. Sie ermöglicht es, einzelne Fälle nachzuvollziehen und Nachweise zu erbringen. Umfang und Aufbewahrungsdauer werden mit Datenschutz und Informationssicherheit abgestimmt.

Monitoring

Das Monitoring beobachtet die Lösung laufend: Qualität in Stichproben, Anteil der Korrekturen, Anteil weitergeleiteter Ausnahmen, Antwortzeiten, Fehler, Verbrauch. Es macht sichtbar, wenn sich die Qualität schleichend verändert, etwa weil sich die Eingaben verändern oder ein Anbieter sein Modell aktualisiert. Für jede Kennzahl ist festgelegt, ab welchem Wert wer benachrichtigt wird.

Änderungsprozess

Modelle, Anweisungen und Konfigurationen werden versioniert. Jede Änderung durchläuft einen festen Weg: Vorschlag, Evaluierung, Freigabe, Einführung, Beobachtung. Ebenso ist festgelegt, wie auf eine frühere Version zurückgewechselt wird und wie die Lösung angehalten und der Ablauf von Hand fortgeführt wird.

Baustein	Zweck	Leitfrage für den Betrieb
Leitplanken	Eingaben, Ausgaben und Handlungen kontrollieren	Welche Vorgänge müssen geprüft, blockiert oder freigegeben werden?
Evaluierungen	Qualität vor Änderungen prüfen	Woran erkennen wir, dass eine Änderung die Qualität verschlechtert?
Protokollierung	Fälle nachvollziehen und Nachweise erbringen	Können wir erklären, wie ein bestimmtes Ergebnis zustande kam?
Monitoring	Veränderungen im laufenden Betrieb erkennen	Bemerken wir einen Qualitätsverlust, bevor er Schaden anrichtet?
Änderungsprozess	Änderungen kontrolliert einführen und zurücknehmen	Wer gibt Änderungen frei, und wie kehren wir zur letzten guten Version zurück?

Das Harness muss nicht vollständig vor dem ersten Einsatz stehen. Sein Umfang richtet sich nach dem Risiko der Anwendung. Protokollierung, ein Rückfallweg und eine Evaluierung vor Änderungen gehören aber zu jeder Lösung im Betrieb.

> Die Voraussetzungen für den Betrieb gehören in die Planung des Pilotprojekts, nicht an sein Ende.

Verantwortung im Betrieb

Im Betrieb braucht die Lösung zwei feste Verantwortungen. Die fachliche Verantwortung liegt im Fachbereich, meist bei der prozessverantwortlichen Person aus dem Pilotprojekt. Sie entscheidet über Umfang, Kontrollmuster und Qualitätsschwellen. Die technische Verantwortung liegt in der IT oder bei einem Produktteam. Sie betreibt die Lösung, das Harness und die Anbindungen.

Legen Sie außerdem fest, wer Änderungen freigibt, wer bei Auffälligkeiten im Monitoring reagiert, wer Rückmeldungen der Nutzer sammelt und wer die Lösung im Inventar der KI-Anwendungen pflegt. Eine Lösung ohne klare Betriebsverantwortung verliert mit der Zeit an Qualität, ohne dass es jemand bemerkt.

Laufende Kosten

Die laufenden Kosten einer KI-Lösung setzen sich anders zusammen als bei klassischer Software. Ein Teil hängt direkt vom Fallaufkommen ab, ein anderer vom Aufwand, die Qualität zu halten. Die folgende Übersicht nennt die üblichen Kostenarten. Ihre Höhe ermitteln Sie aus den Messwerten des Pilotprojekts und den Angeboten Ihrer Anbieter.

Kostenart	Treiber	Hinweis
Modellnutzung	Fallaufkommen, Länge der Eingaben und Ausgaben, Wahl des Modells	Aus dem Verbrauch je Fall im Pilotprojekt hochrechnen, Schwankungen berücksichtigen
Infrastruktur	Hosting, Speicher, Protokolle, Netzwerk	Aufbewahrungsdauer der Protokolle beeinflusst den Speicherbedarf
Integration und Wartung	Anzahl und Stabilität der Schnittstellen	Änderungen an angebundenen Systemen erzeugen Folgeaufwand
Harness und Monitoring	Umfang der Leitplanken, Evaluierungen und Kennzahlen	Mit dem Risiko der Anwendung abstimmen

Kostenart	Treiber	Hinweis
Menschliche Prüfung	Kontrollmuster, Anteil der Ausnahmen, Korrekturaufwand	Oft der größte laufende Posten und im Pilotprojekt messbar
Pflege der Qualität	Erweiterung des Referenzsets, Anpassung von Anweisungen, Modellwechsel	Auch ohne neue Funktionen fällt Aufwand an
Schulung und Begleitung	Anzahl der Nutzer, Fluktuation, Änderungen am Ablauf	Neue Mitarbeitende brauchen Einführung
Governance und Nachweis	Inventar, Prüfungen, Dokumentation	Je nach Risikoklasse unterschiedlich umfangreich

Rollout

Führen Sie die Lösung schrittweise ein. Beginnen Sie mit dem Team aus dem Pilotprojekt, erweitern Sie auf weitere Teams oder Standorte und prüfen Sie nach jedem Schritt, ob die Kennzahlen stabil bleiben. Jede Erweiterung des Umfangs, etwa auf neue Fallarten oder Sprachen, wird wie eine Änderung behandelt: mit Evaluierung und Freigabe.

Für jeden Schritt gilt: Der Rückfallweg ist bekannt und erprobt. Die Menschen, die mit der Lösung arbeiten, sind geschult, kennen ihre Prüfaufgabe und wissen, wie sie Auffälligkeiten melden. Die Fachanwender aus dem Pilotprojekt sind dabei wertvolle Ansprechpersonen für neue Teams.

Planen Sie außerdem einen festen Termin, an dem die Lösung nach einer ersten Betriebsphase erneut bewertet wird. Dort wird geprüft, ob Nutzen, Kosten und Risiko den Erwartungen aus der Entscheidung entsprechen, und ob das Kontrollmuster angepasst werden kann.

Auf einen Blick

- Klären Sie Betriebsvoraussetzungen während des Pilotprojekts, nicht danach.
- Das AI Harness verbindet Kontrollen, Prüfungen und Nachweise im Betrieb. Seine Wirksamkeit und Grenzen müssen überprüft werden.
- Die Lösung braucht eine fachliche und eine technische Verantwortung.
- Laufende Kosten entstehen auch durch menschliche Prüfung und Pflege der Qualität, nicht nur durch Modellnutzung.
- Führen Sie die Lösung schrittweise ein, mit erprobtem Rückfallweg.

Anhang

Von der KI-Idee zum Pilotprojekt. Ein Leitfaden

Der Anhang fasst die Arbeitsschritte dieses Leitfadens in Checklisten und Vorlagen zusammen. Die Vorlagen lassen sich unverändert verwenden oder in bestehende Projektwerkzeuge übernehmen.

Checklisten

Aufgabe wählen

- [ ] Die Aufgabe ist so beschrieben, dass Außenstehende sie ohne Rückfrage verstehen.
- [ ] Eingaben, Ergebnis und die weiterarbeitenden Personen sind benannt.
- [ ] Für jeden Schritt ist geprüft, ob ein Modell, eine Regel oder ein Mensch ihn am besten erledigt.
- [ ] Die Aufgabe ist in der Bewertungsmatrix bewertet, die Gewichtung wurde vorab festgelegt.
- [ ] Der Umfang ist eingegrenzt, und was nicht dazugehört, ist schriftlich festgehalten.
- [ ] Eine prozessverantwortliche Person und ein mitwirkendes Team sind benannt.

Frage und Kriterien festlegen

- [ ] Die Entscheidungsfrage ist schriftlich festgehalten und mit der entscheidenden Person abgestimmt.
- [ ] Die Baseline des heutigen Ablaufs ist erfasst, auf die gleiche Weise wie die spätere Messung.
- [ ] Erfolgskriterien zu Qualität, Prüfaufwand, Akzeptanz und Betrieb sind mit Messung, Schwelle und Quelle beschrieben.
- [ ] Abbruchkriterien sind festgelegt, ebenso wer einen Abbruch feststellt und beschließt.
- [ ] Der Kriterienkatalog ist mit Datum und Version festgeschrieben.

Daten, Recht und Sicherheit klären

- [ ] Verfügbarkeit, Qualität und Nutzungsrechte der Daten sind an einer Stichprobe geprüft.
- [ ] Zweck, Rechtsgrundlage, Auftragsverarbeitung und Verarbeitungsort sind mit dem Datenschutz geklärt.
- [ ] Anbieter, Berechtigungen, Umgang mit unsicheren Eingaben und Protokollierung sind mit der Informationssicherheit abgestimmt.
- [ ] Die Anwendung ist im Inventar erfasst und nach den Risikoklassen des EU AI Act vorläufig eingeordnet, die rechtliche Bewertung ist beauftragt.
- [ ] Freigaben vor dem Einsatz echter Daten liegen vor.
- [ ] Anforderungen, die erst für den Betrieb erfüllt sein müssen, sind gelistet.

Mensch und Verantwortung festlegen

- [ ] Das Kontrollmuster ist gewählt und begründet.
- [ ] Die Bedingungen, unter denen Kontrolle gelockert werden darf, sind festgelegt.
- [ ] Für jedes Ergebnis ist klar, wer es verantwortet.
- [ ] Die Rollen sind besetzt, und je Aufgabe gibt es genau eine abschließend verantwortliche Rolle.
- [ ] Korrekturen werden strukturiert mit Fehlerkategorie erfasst.

Pilotbetrieb durchführen

- [ ] Zeitrahmen und Phasen sind festgelegt, die Zeit der Fachanwender ist vereinbart.
- [ ] Das Referenzset enthält gewöhnliche, schwierige, unvollständige, fachfremde und kritische Fälle.
- [ ] Zwischentermine sind terminiert.
- [ ] Ergebnis-, Beobachtungs- und Änderungsprotokoll werden laufend geführt.
- [ ] Jede Änderung an der Lösung wird vorab am Referenzset geprüft.
- [ ] Die Beteiligten und die betroffenen Bereiche sind informiert, Mitbestimmungsgremien sind eingebunden, wo nötig.

Auswerten und entscheiden

- [ ] Jedes Kriterium ist mit Messwert, Schwelle und Einordnung dargestellt.
- [ ] Die Ergebnisse sind nach Fallarten aufgeschlüsselt.
- [ ] Beobachtungen außerhalb der Kriterien sind in einem eigenen Abschnitt festgehalten.
- [ ] Die Entscheidungsvorlage ist vor dem Termin verteilt.
- [ ] Der Beschluss mit Begründung, Bedingungen und nächsten Schritten ist schriftlich festgehalten und allen Beteiligten mitgeteilt.

In den Betrieb überführen

- [ ] Fachliche und technische Verantwortung für den Betrieb sind benannt.
- [ ] Manuelle Zwischenschritte sind durch Anbindungen an bestehende Systeme ersetzt.
- [ ] Leitplanken, Evaluierungen, Protokollierung, Monitoring und Änderungsprozess sind dem Risiko entsprechend umgesetzt.
- [ ] Der Rückfallweg ist beschrieben und erprobt.
- [ ] Die laufenden Kostenarten sind erfasst und im Budget berücksichtigt.
- [ ] Nutzerinnen und Nutzer sind geschult, ein Termin für die erneute Bewertung ist festgelegt.

Vorlagen

Pilotsteckbrief

Feld	Inhalt
Bezeichnung des Pilotprojekts	
Aufgabe	
Entscheidungsfrage	
Umfang	
Nicht im Umfang	
Entscheidende Person	
Prozessverantwortliche Person	
Projektleitung	
Fachanwenderinnen und Fachanwender	
IT und Architektur	
Informationssicherheit	
Datenschutz	
Datenquellen und Freigaben	
Kontrollmuster	
Verantwortung für Ergebnisse	
Zeitrahmen und Phasen	
Zwischentermine	
Entscheidungstermin	
Offene Voraussetzungen für den Betrieb	

Kriterienkatalog

Nr.	Kriterium	Art	Messung	Baseline	Schwelle	Quelle	Verantwortlich
1		Erfolg					
2		Erfolg					
3		Erfolg					
4		Erfolg					
5		Erfolg					
6		Abbruch					
7		Abbruch					
8		Abbruch					

Version: … Festgeschrieben am: … Freigegeben durch: …

Entscheidungsvorlage

Abschnitt	Inhalt
Pilotprojekt	
Entscheidungsfrage	
Empfehlung in wenigen Sätzen	
Kriterien: erfüllt	
Kriterien: teilweise erfüllt, mit Begründung	
Kriterien: nicht erfüllt	
Beobachtungen außerhalb der Kriterien	
Offene Risiken	
Offene Voraussetzungen für den Betrieb	
Laufende Kostenarten	
Folgen bei Skalieren	

Abschnitt	Inhalt
Folgen bei Überarbeiten	
Folgen bei bewusstem Stoppen	
Beschluss	
Bedingungen	
Nächste Schritte und Verantwortliche	
Datum und Unterschrift der entscheidenden Person	

Glossar

Begriff	Bedeutung
Abbruch-kriterium	Vorab festgelegte Bedingung, unter der ein Pilotprojekt vorzeitig endet.
AI Harness	Technische Kontrollschicht um Modelle und Agenten, mit Leitplanken, Evaluierungen, Protokollierung, Monitoring, Freigaben und Versionierung.
Agent	KI-System, das eine Aufgabe in Teilschritte zerlegt, dafür Werkzeuge und Systeme aufruft und sein Vorgehen anhand von Zwischenergebnissen anpasst.
Baseline	Erfassung des heutigen Ablaufs, an der die Ergebnisse des Pilotprojekts gemessen werden.
Entsch-eidungs-frage	Die eine Frage, die das Pilotprojekt für die entscheidende Person beantworten soll.
Erfolgs-kriterium	Prüfbare Anforderung mit Messung, Schwelle und Quelle, die vor dem Start festgelegt wird.
EU AI Act	Verordnung (EU) 2024/1689 über künstliche Intelligenz in ihrer jeweils geltenden Fassung. Pflichten und Anwendungstermine hängen von Rolle, Anwendung und Übergangsregeln ab. Den datierten Überblick enthält Kapitel 4.
Evaluie-rung	Systematische Prüfung der Qualität einer Lösung an einem festen Satz von Fällen, insbesondere vor Änderungen.
Kontroll-muster	Festgelegte Form menschlicher Prüfung: jedes Ergebnis, Stichprobe oder Ausnahmen.

Begriff	Bedeutung
Leitplanke	Technische Kontrolle für Eingaben, Ausgaben oder Aktionen, etwa durch Regeln, Berechtigungsprüfungen oder modellgestützte Bewertungen. Wirksamkeit und Grenzen müssen geprüft werden.
Monitoring	Laufende Beobachtung von Qualität, Verhalten und Verbrauch einer Lösung im Betrieb.
Parallelbetrieb	Phase, in der die Lösung neben dem heutigen Ablauf läuft und ihre Ergebnisse verglichen, aber nicht verwendet werden.
Pilotprojekt	Zeitlich begrenzte Erprobung einer Lösung im tatsächlichen Ablauf mit echten Fällen und vorab festgelegten Kriterien, die zu einer Entscheidung führt.
Proof of Concept	Prüfung, ob etwas technisch möglich ist.
Referenzset	Feste Sammlung von Fällen mit bekanntem richtigem Ergebnis, an der die Lösung gleichbleibend geprüft wird.

Sprechen wir über Ihr Vorhaben.

Optimentum begleitet Unternehmen in Strategie, Technologie und Transformation. Schreiben Sie uns, welche Entscheidung bei Ihnen ansteht.

office@optimentum.eu
optimentum.eu

Optimentum GmbH · Tegetthoffstraße 3/10, 1010 Wien · FN 686237y, Handelsgericht Wien