DDIM.

PRAXISBEISPIEL / CASE STUDY – DIGITALE TRANSFORMATION & KI

Wenn vier bis fünf Millionen Token noch kein Wissen sind

Wie agentische KI, systematische Validierung und Context Engineering helfen, hochkomplexe Regulatorik beherrschbar zu machen

Marc Kresin
Leiter der DDIM-Fachgruppe Digitale Transformation & KI

> **Information wird zunehmend zur Commodity. Belastbares Wissen nicht.**

15.007
validierte Wissensobjekte

9.515
atomisierte regulatorische Anforderungen

2.025
Exception-, Conflict- und Gate-Einträge

16
Human-/Counsel-Gates

Von Millionen Informationen zu kontrolliertem Wissen

15,26 Mio.
Zeichen finale Textbasis

≈ 4–5 Mio.
LLM-Tokens, abhängig vom Tokenizer

15.007
validierte Wissensobjekte

9.515
atomisierte regulatorische Anforderungen

214
CEE-Fiches

346
kanonische Quellen

386
nachvollziehbare Source-/Artifact-Lineage-Objekte

2.025
Exception-, Conflict- und Gate-Einträge

16
Human-/Counsel-Gates

20
dokumentierte Validierungsbatches der Fiche-Schicht

3
unabhängige Red-Check-Runden

111/111
finale QA-Prüfungen bestanden

38/38
Closure-Gate-Prüfungen bestanden

27/27
Assembly-QA-Prüfungen bestanden

0
materielle Findings im finalen unabhängigen Red Check

KERNAUSSAGE

Nicht möglichst viel Information in die KI laden.

Sondern Wissen so validieren und strukturieren, dass für jede Aufgabe der richtige Kontext bereitsteht.

ZITAT MARC KRESIN:

> Der entscheidende KI-Hebel in komplexen Umfeldern liegt nicht darin, immer mehr Informationen zu erzeugen. Er liegt darin, aus verfügbaren Informationen belastbares Wissen zu machen – und dieses Wissen für Menschen und Maschinen im richtigen Kontext verfügbar zu halten.

In einem aktuellen Transformationsmandat in Frankreich beschäftige ich mich mit einem Thema, das außerhalb der Energiebranche vermutlich nur wenige kennen: **CEE – Certificats d'Économies d'Énergie**.

Das CEE-System ist ein französisches regulatorisches Instrument zur Förderung und zum Nachweis von Energieeinsparungen. Für verpflichtete Energieunternehmen und andere betroffene Marktteilnehmer ist es kein freiwilliges Zusatzprogramm, sondern regulatorische Realität.

Und diese Realität ist ausgesprochen komplex.

Das System umfasst regulatorische Vorgaben, standardisierte Energieeffizienzmaßnahmen, Nachweispflichten, Kontrollen, Fristen, Marktmechanismen, Rollen, Dokumentationsanforderungen, Due-Diligence-Fragen, Fraud-Risiken und zahlreiche Sonderkonstellationen.

Im aktuellen Mandat wurde das operative Umfeld zunächst gemeinsam in einer Taskforce stabilisiert. Parallel entsteht ein zukünftiges **Target Operating Model**. Perspektivisch soll eine digitale und KI-gestützte Plattform die Arbeit unterstützen.

Bevor man aber Prozesse digitalisiert, automatisiert oder KI-Agenten daraufsetzt, muss eine fundamentalere Frage beantwortet werden:

Worauf soll die KI eigentlich arbeiten?

Denn eine schlechte, widersprüchliche oder nicht nachvollziehbare Wissensbasis wird durch KI nicht besser.

Sie wird nur schneller skaliert.

Information ist nicht mehr der Engpass

Noch vor wenigen Jahren bestand ein erheblicher Teil professioneller Recherche darin, Informationen überhaupt zu finden.

Welches Gesetz ist relevant? Welche Behörde hat welche Regel veröffentlicht? Wo liegen die Marktdaten? Welche Quelle ist aktuell? Welche Detailanforderung gilt für welchen Fall?

Generative KI und moderne Recherchewerkzeuge verändern diese Situation fundamental. Öffentliche Informationen lassen sich heute mit einer Geschwindigkeit erschließen, die vor wenigen Jahren kaum vorstellbar gewesen wäre.

Damit verschiebt sich jedoch der Engpass.

Information wird zunehmend zur Commodity. Belastbares Wissen nicht.

Ein Large Language Model kann in Sekunden eine hervorragend klingende Antwort erzeugen. Sprachliche Qualität ist aber kein Beweis für faktische Qualität.

Gerade in regulierten Märkten können scheinbar kleine Unterschiede entscheidend sein:

- Ist eine Regel bereits in Kraft oder nur angekündigt?
- Ist eine Quelle eine regulatorische Primärquelle oder zitiert sie lediglich eine andere Quelle?
- Für B2C, B2B oder einen Teilmarkt?
- Ist ein Wert eine beobachtete Tatsache, eine Berechnung, eine Prognose oder eine Annahme?
- Ist eine Aussage öffentlich belegbar oder braucht sie eine unternehmensinterne bzw. juristische Bewertung?
- Bezieht sich eine Zahl auf Kunden, Verträge, Lieferstellen oder Volumen?
- Gilt eine Aussage für Strom oder Gas?
- Für welchen Zeitraum?
- Gibt es widersprüchliche Quellen?

Wer diese Unterschiede nicht kontrolliert, baut auf eine Faktenbasis, die präzise aussieht, aber strukturell fragil ist.

Was „große Wissensmenge“ konkret bedeutet

Im Projekt entstand am Ende nicht einfach eine Sammlung von Dokumenten, sondern eine konsolidierte, validierte und maschinenlesbare Wissensbasis.

Allein ihre finale textbasierte Schicht umfasst rund **15,26 Millionen Zeichen**.

Je nach verwendetem Tokenizer entspricht das einer Größenordnung von ungefähr **vier bis fünf Millionen LLM-Tokens** – grob mehrere vollständige Bibeln Text.

Die Tokenzahl vermittelt allerdings nur ein Gefühl für die Größenordnung.

Interessanter ist, was aus diesem Informationsraum tatsächlich gemacht wurde:

DIE FINALE WISSENSBASIS IN ZAHLEN

KENNZAHL	FINALER STAND
Validierte Wissensobjekte	**15.007**
Atomisierte regulatorische Anforderungen	**9.515**
CEE-Fiches	**214**
Kanonische Quellen	**346**

KENNZAHL	FINALER STAND
Nachvollziehbare Source-/Artifact-Lineage-Objekte	**386**
Exception-, Conflict- und Gate-Einträge	**2.025**
Explizite Human-/Counsel-Gates	**16**
Kontrollierte Artefakte im finalen Release	**25**

Das ist der entscheidende Unterschied zwischen einer großen Dokumentensammlung und einer Wissensbasis.

Die Informationen wurden nicht nur gesammelt.

Sie wurden **atomisiert, klassifiziert, mit stabilen Identitäten versehen, mit Quellen verbunden, auf Konflikte geprüft, mit Evidenzstatus versehen und mit expliziten Grenzen für ihre spätere Nutzung ausgestattet.**

Eine solche Menge kann kein Mensch linear lesen, dauerhaft im Kopf behalten und bei jeder Entscheidung vollständig berücksichtigen.

Aber auch eine KI wird nicht dadurch besser, dass man ihr einfach Millionen Token ungeordnet in einen Prompt kippt.

Genau hier beginnt **Context Engineering**.

Context Engineering: Nicht alles wissen – das Richtige verfügbar machen

Viele Diskussionen über generative KI drehen sich noch immer um Prompt Engineering.

Wie formuliere ich eine möglichst gute Frage?

Das ist relevant. Aber bei anspruchsvollen Anwendungen greift es zu kurz.

Viel entscheidender wird die Frage:

Welches Wissen stelle ich der KI für eine konkrete Aufgabe in welchem Kontext zur Verfügung?

Ein einfaches Bild dafür ist eine juristische Bibliothek.

Stellen wir uns vor, sämtliche deutschen Gesetze, Verordnungen, Kommentare und Urteile seit Gründung der Bundesrepublik wären digital in einer einzigen Bibliothek verfügbar.

Dieser Wissensraum wäre gewaltig.

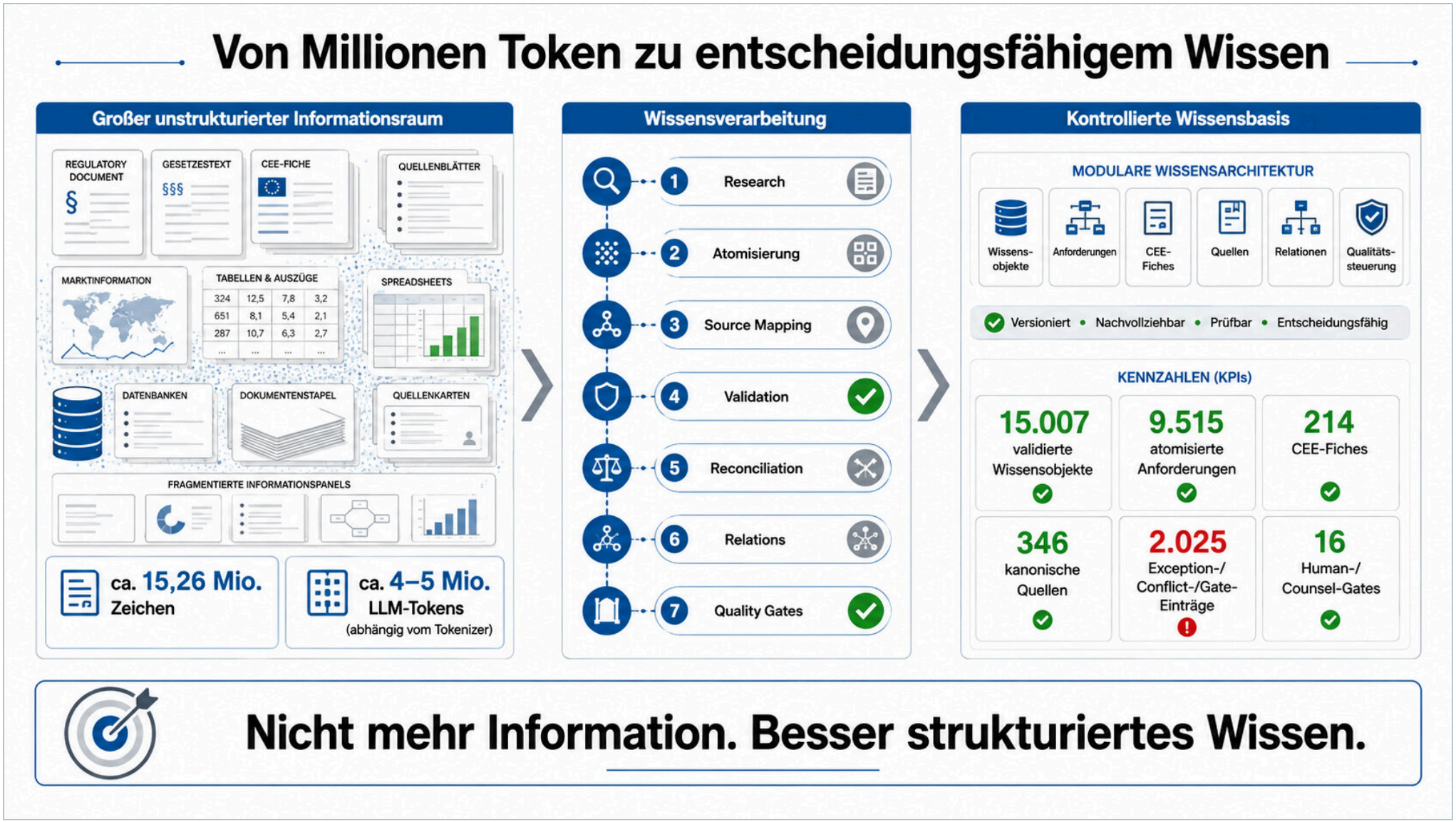


Abbildung 1: Von Millionen Token zu entscheidungsfähigem Wissen - vom unstrukturierten Informationsraum zur kontrollierten Wissensbasis.

Wenn ich aber einen konkreten Fall aus dem Verkehrsrecht lösen möchte, brauche ich nicht die gesamte Bibliothek gleichzeitig.

Zuerst interessiert mich das richtige Rechtsgebiet.

Dann der konkrete Sachverhalt.

Dann bestimmte Normen.

Dann relevante Definitionen.

Dann einschlägige Urteile.

Vielleicht bestimmte Sonderkonstellationen.

Und möglicherweise offene Fragen, bei denen juristische Interpretation erforderlich bleibt.

Genau das ist **Context Engineering** für KI.

Die Aufgabe besteht nicht darin, möglichst viele Millionen Token gleichzeitig in ein Modell zu laden.

Die Aufgabe besteht darin, aus einem sehr großen Wissensraum **für eine konkrete Fragestellung den relevanten, belastbaren und ausreichend vollständigen Kontext zusammenzustellen**.

Dazu gehören nicht nur Texte.

Dazu gehören auch:

- Quellen,
- Relationen,
- Definitionen,
- Gültigkeitsbereiche,
- Evidenzstatus,
- Konflikte,
- Annahmen,
- Aktualität,
- bekannte Grenzen,
- und Freigabe- bzw. Eskalationsregeln.

Damit wird eine Wissensbasis zu weit mehr als einem digitalen Archiv.

Sie wird zur **Navigations- und Kontextarchitektur für Menschen und Maschinen**.

Der entscheidende Schritt: vom Dokument zur prüfbaren Aussage

Dafür musste auch die Einheit verändert werden, in der Wissen bearbeitet wird.

Ein Dokument ist dafür häufig zu groß.

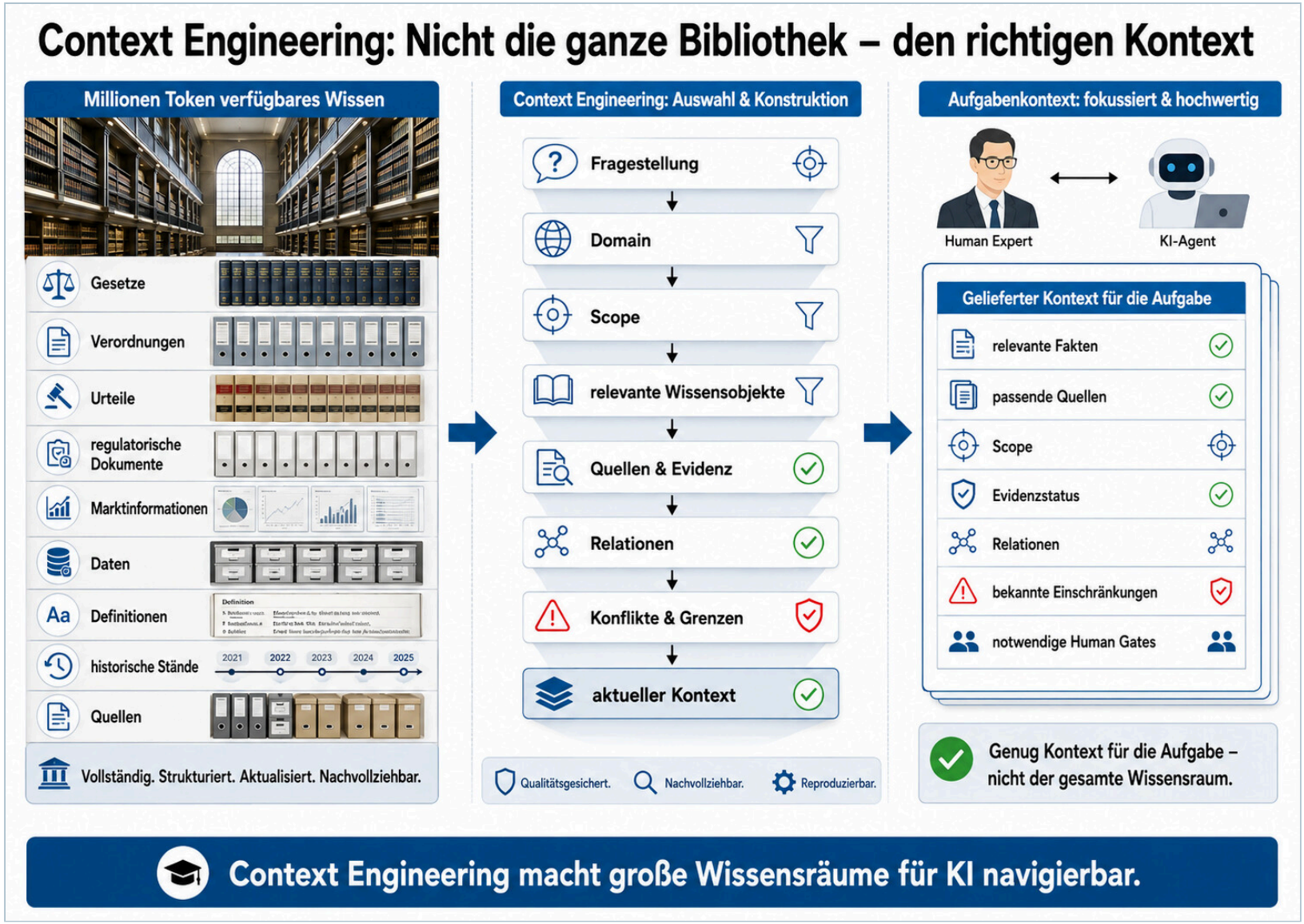


Abbildung 2: Context Engineering - nicht die ganze Bibliothek, sondern der für die konkrete Aufgabe relevante und belastbare Kontext.

Auch ein Absatz ist oftmals noch zu unscharf.

Die zentrale Einheit der Validierung wird deshalb der **Claim beziehungsweise die atomisierte Anforderung** – also eine Aussage, die sich möglichst unabhängig überprüfen lässt.

Nehmen wir einen scheinbar einfachen Satz:

> **„Alternative Anbieter gewinnen im französischen Energiemarkt Marktanteile."**

Was genau bedeutet das?

- Strom oder Gas?
- Welcher Zeitraum?
- Kunden, Verträge, Lieferstellen, Absatzmenge oder Umsatz?
- Welche Definition verwendet diese Quelle?
- Ist die Aussage noch aktuell?
- Privatkunden oder Geschäftskunden?
- Welcher Marktanteil?
- Welche Quelle?
- Gibt es eine zweite unabhängige Bestätigung?

Erst wenn diese Fragen beantwortet sind, wird aus einer plausiblen Aussage ein belastbares Wissensobjekt.

Deshalb erhält eine materielle Information eine eindeutige Identität, Quellenbezug, Scope und Validierungsstatus.

Das Prinzip dahinter ist einfach:

Nicht der schön formulierte Absatz wird validiert. Die einzelne überprüfbare Aussage wird validiert.

Qualität heißt nicht, möglichst viel auf Grün zu setzen

Besonders wichtig war dabei, Unsicherheit nicht aus dem System herauszuoptimieren.

Von den **9.515 atomisierten regulatorischen Anforderungen** wurden final:

VALIDIERUNGSSTATUS	ANZAHL
VERIFIED-A	3.816
VERIFIED-B	2.233
PARTIAL	2.301
UNVERIFIED	559
CONFLICT	416
OUTDATED	185
ASSUMPTION	5

Diese Verteilung ist kein Qualitätsproblem.

Im Gegenteil.

Sie ist eine der wichtigsten Eigenschaften einer professionellen Wissensbasis.

Denn ein gutes Wissenssystem muss nicht behaupten, alles zu wissen.

Es muss erkennen können:

Was wissen wir?

Woher wissen wir es?

Wie belastbar wissen wir es?

Und was wissen wir noch nicht?

Gerade bei Regulatorik ist diese Unterscheidung elementar.

Ein Konflikt darf nicht deshalb verschwinden, weil ein Sprachmodell daraus eine elegante Synthese erzeugen kann.

Ein fehlender Beleg bleibt ein fehlender Beleg.

Eine veraltete Information bleibt als solche gekennzeichnet.

Und eine Frage, die juristische, unternehmensspezifische oder fachliche Beurteilung verlangt, muss beim Menschen bleiben.

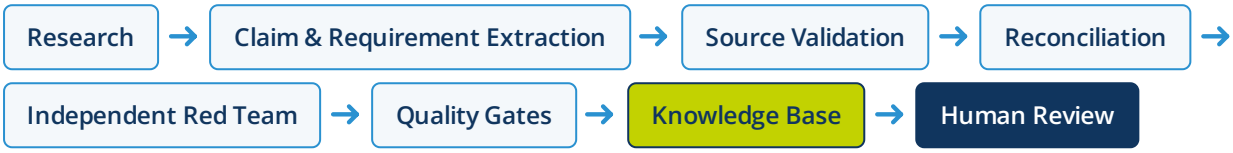
Die Aufgabe von Context Engineering ist nicht, alles gleichzeitig verfügbar zu machen – sondern das Richtige zur richtigen Zeit.

Die Prozesskette: vom Rechercheergebnis zur belastbaren Faktenbasis

Der Kern des Ansatzes ist deshalb kein bestimmtes KI-Produkt.

Es ist eine **Prozess- und Qualitätsarchitektur**.

Vereinfacht folgt sie dieser Kette:



1. Research

Das Themenfeld wird systematisch erschlossen.

Regulatorische Primärquellen, öffentliche Daten, offizielle Dokumente und weitere relevante Quellen werden identifiziert und in eine erste Struktur gebracht.

KI beschleunigt diesen Schritt enorm.

Aber das Ergebnis ist zunächst **Research – noch keine Wahrheit**.

2. Claim & Requirement Extraction

Materielle Aussagen und regulatorische Anforderungen werden aus den Dokumenten extrahiert und atomisiert.

Aus umfangreichen Texten entstehen kleine, adressierbare Wissensobjekte.

3. Source Validation

Jede relevante Aussage wird gegen ihre Evidenz geprüft.

Dabei gilt eine Quellenhierarchie.

Eine regulatorische Primärquelle besitzt für eine regulatorische Aussage einen anderen Stellenwert als ein Presseartikel oder eine Beratungsstudie.

4. Reconciliation

Unterschiedliche Quellen, Definitionen und Datenstände werden miteinander abgeglichen.

Widersprüche werden nicht verdeckt, sondern explizit registriert.

5. Independent Red Team

Anschließend wird die Wissensbasis bewusst angegriffen.

Andere Agenten beziehungsweise Modelle suchen nach:

- falschen Schlussfolgerungen,
- inkonsistenten Definitionen,
- fehlenden Quellen,
- ungerechtfertigten Verallgemeinerungen,
- unerlaubten Status-Upgrades,
- stillschweigend geschlossenen Konflikten,
- und Schwächen in der Nachvollziehbarkeit.

Das System, das eine Information erzeugt, sollte möglichst nicht die einzige Instanz sein, die diese Information anschließend bestätigt.

6. Quality Gates

Definierte Prüfungen entscheiden, was akzeptiert wird, was eingeschränkt verwendet werden darf und was offenbleiben muss.

7. Knowledge Base

Erst jetzt wird aus Research eine kontrollierte Wissensbasis.

Quellen, Objekte, Relationen, Konflikte, Evidenzstatus und Nutzungseinschränkungen bleiben nachvollziehbar erhalten.

8. Human Review

Wo die Grenze maschineller Prüfung erreicht ist, übernimmt der Mensch.

Nicht als nachträgliche Dekoration.

Sondern als bewusster Bestandteil der Architektur.

Agentisch heißt nicht autonom

Eine solche Arbeitsweise lässt sich kaum noch sinnvoll als Folge einzelner Chats organisieren.

Sie benötigt eine Arbeitsumgebung, in der verschiedene Aufgaben, KI-Systeme, Agenten, Dateien, Wissensbestände und Quality Gates miteinander orchestriert werden können.

Ich bezeichne meinen eigenen KI-Arbeitsplatz dafür als **Personal Agentic Operating System**.

Der Name ist dabei nicht entscheidend.

Entscheidend ist das Prinzip.

Ein agentischer Arbeitsplatz ermöglicht es, unterschiedliche Aufgaben an spezialisierte Rollen zu verteilen:

- ein Agent recherchiert,
- ein anderer extrahiert Anforderungen,
- ein weiterer prüft Quellen,
- ein weiterer sucht nach Konflikten,
- ein Red Team versucht das Ergebnis zu widerlegen,
- ein Syntheseprozess konsolidiert die akzeptierten Ergebnisse,
- Quality Gates kontrollieren Übergänge,
- und der Mensch entscheidet an definierten Punkten.

Das unterscheidet sich fundamental von der Nutzung eines einzelnen Chatbots.

Der Mensch bearbeitet nicht mehr jede Information selbst.

Er **designt und orchestriert ein Arbeitssystem**, in dem Menschen und Maschinen unterschiedliche Aufgaben übernehmen.

Das ist eine notwendige Grundlage, wenn KI nicht nur einzelne Texte schreiben, sondern komplexe Wissensarbeit unterstützen soll.

Agentisch heißt nicht autonom: Die KI übernimmt Arbeit. Der Mensch behält Urteil und Verantwortung.

Human in the Lead – Human in the Loop

Agentisch bedeutet dabei gerade **nicht**, möglichst viel Verantwortung an Maschinen abzugeben.

Je leistungsfähiger KI-Systeme werden, desto wichtiger wird eine klare Trennung zwischen maschineller Arbeit und menschlicher Verantwortung.

Für mich sind dabei zwei Prinzipien entscheidend.

Human in the Lead

Der Mensch bestimmt:

- Ziel und Scope,
- Qualitätsmaßstab,
- zulässige Evidenz,
- Risikogrenzen,
- Prozessarchitektur,

- Freigaberegeln,
- und letztlich die Entscheidung.

Die KI arbeitet innerhalb dieses Rahmens.

Sie definiert ihn nicht selbst.

Human in the Loop

An bestimmten Stellen wird der Mensch bewusst in den Prozess zurückgeholt.

Beispielsweise bei:

- widersprüchlicher Evidenz,
- juristischer Interpretation,
- materiellen Unsicherheiten,
- unternehmensspezifischen Entscheidungen,
- fehlenden internen Informationen,
- oder Sachverhalten, bei denen fachliches Urteil nicht substituiert werden kann.

In der finalen Wissensbasis blieben deshalb **16 Human- beziehungsweise Counsel-Gates ausdrücklich offen**.

Diese Punkte wurden nicht mit zusätzlichen KI-Läufen künstlich „geschlossen".

Sie markieren bewusst die Grenze dessen, was mit der vorhandenen Evidenz verantwortbar entschieden werden kann.

Die KI übernimmt Arbeit. Der Mensch behält Urteil und Verantwortung.

Wie viel Qualitätssicherung steckt dahinter?

Die Qualitätssicherung war kein einzelner finaler Check.

Sie wurde über mehrere Ebenen aufgebaut.

Die Fiche-Schicht wurde über **20 dokumentierte Validierungsbatches** aufgebaut und geprüft.

Darauf folgten weitere Reconciliation-, Release- und Kontrollstufen.

Das unabhängige Red Checking umfasste **drei dokumentierte Review-Runden**.

Am Ende standen unter anderem:

QUALITÄTSKONTROLLE	ERGEBNIS
Finale Master-Control-Room-QA	**111 / 111 PASS**

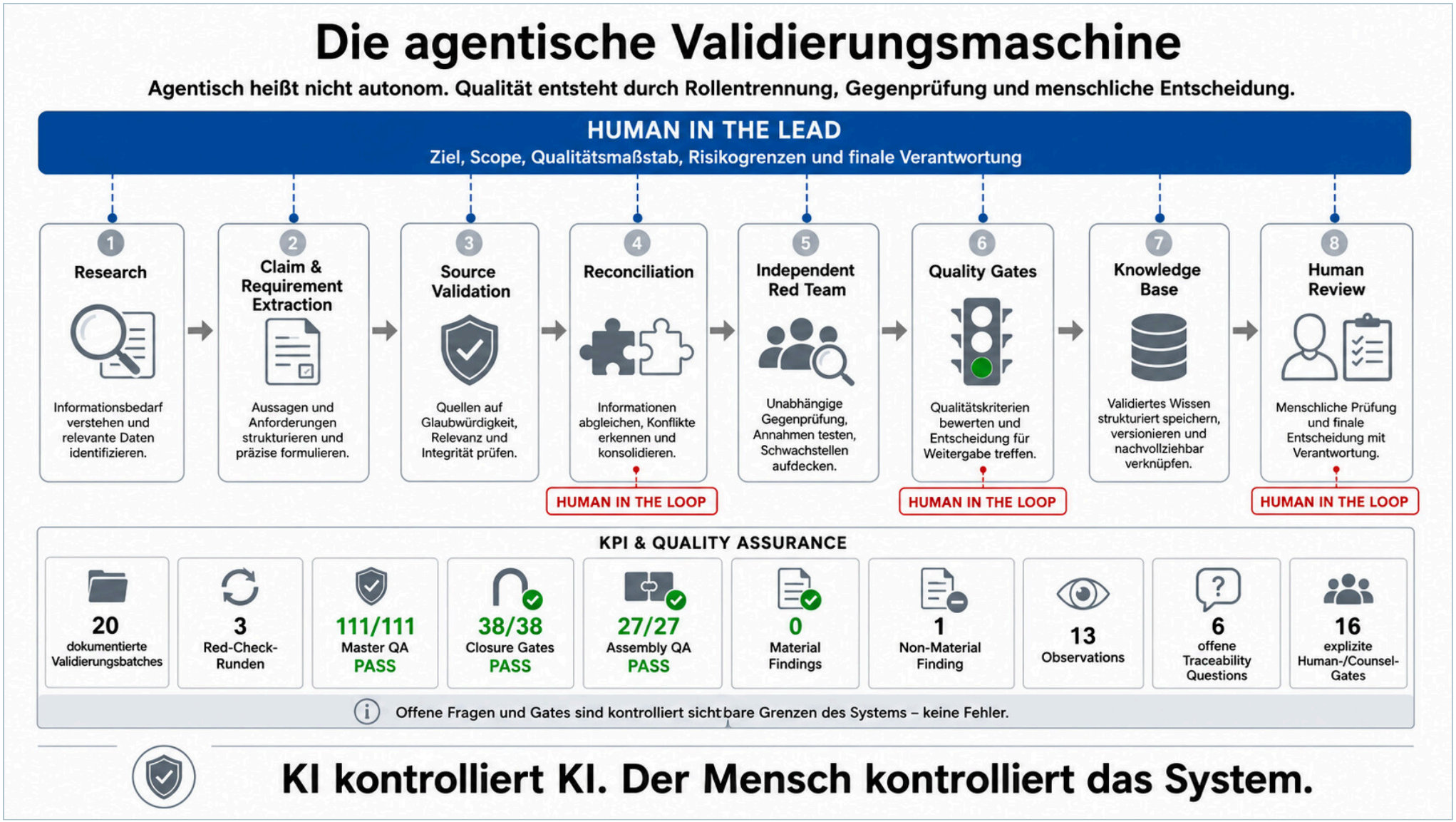


Abbildung 3: Die agentische Validierungsmaschine - Rollen- und Qualitätsarchitektur mit Human in the Lead und definierten Human-in-the-Loop-Gates.

QUALITÄTSKONTROLLE	ERGEBNIS
Closure-Gate-Prüfungen	38 / 38 PASS
Final Assembly QA	27 / 27 PASS
Material Findings im unabhängigen Red Check	0
Non-Material Findings	1
Bewusst erhaltene Observations	13
Offene, begrenzte Traceability Questions	6

Auch hier ist der letzte Punkt wichtig.

Das Ziel eines guten Validierungssystems besteht nicht darin, am Ende überall **PASS** zu schreiben.

Es besteht darin, sicherzustellen, dass bekannte Grenzen nicht auf dem Weg zur schönen Endpräsentation verschwinden.

Die finale Wissensbasis enthält deshalb weiterhin **2.025 Exception-, Conflict- und Gate-Einträge**.

Unsicherheit wird nicht versteckt.

Sie wird sichtbar und damit steuerbar.

Was dadurch möglich wird

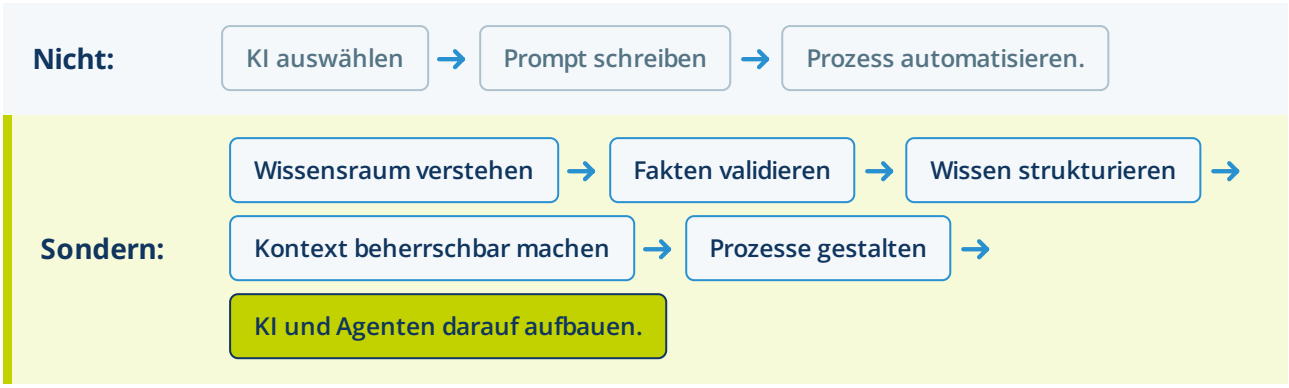
Der Aufbau einer solchen Faktenbasis ist kein Selbstzweck.

Er schafft das Fundament für die eigentliche Transformation.

Auf dieser Basis lässt sich beispielsweise:

- ein Target Operating Model entwickeln,
- Prozessdesign auf regulatorische Anforderungen zurückführen,
- Rollen und Verantwortlichkeiten sauber ableiten,
- Compliance Controls mit konkreten Anforderungen verbinden,
- neue Mitarbeiter schneller einarbeiten,
- Entscheidungen systematisch vorbereiten,
- eine digitale Plattform aufbauen,
- und später spezialisierten KI-Agenten kontrollierter Kontext bereitstellen.

Das verändert die Reihenfolge, in der man über KI nachdenken sollte.



Warum das ein Thema digitaler Transformation ist

Das Beispiel zeigt aus meiner Sicht auch, warum KI nicht isoliert als Technologie betrachtet werden sollte.

Das Problem ist nicht:

„Welches Modell benutzen wir?“

Die relevanteren Fragen lauten:

- Wie organisieren wir Wissen?
- Wie sichern wir Evidenz?
- Wie behandeln wir Unsicherheit?
- Wie verteilen wir Arbeit zwischen Menschen und Maschinen?
- Welche Entscheidungen dürfen automatisiert vorbereitet werden?
- Wo bleibt menschliches Urteil zwingend erforderlich?
- Wie integrieren wir das Ganze in Prozesse, Governance und Operating Model?

Damit wird aus einem KI-Anwendungsfall ein Thema von **digitaler Transformation**.

Denn am Ende verändern wir nicht nur ein Werkzeug.

Wir verändern, **wie eine Organisation Wissen erzeugt, prüft, verfügbar macht und in Entscheidungen übersetzt.**

Die eigentliche Erkenntnis

Wir diskutieren aktuell sehr viel darüber, welches Large Language Model besser ist, welches den größten Kontext verarbeiten kann oder welcher Anbieter gerade technologisch vorne liegt.

Das sind relevante technische Fragen.

Aber sie sind nicht der Kern.

Modelle werden leistungsfähiger.

Information wird immer einfacher verfügbar.

Recherche wird immer schneller.

Der nachhaltigere Vorteil entsteht an einer anderen Stelle:

Kann eine Organisation Wissen so sammeln, prüfen, strukturieren und verknüpfen, dass Menschen und KI damit verlässlich arbeiten können?

Genau dort treffen sich für mich KI und digitale Transformation.

Denn KI kann ein enormer Hebel bei der Beherrschung von Komplexität sein.

Aber nicht, weil wir ihr einfach immer mehr Informationen geben.

Sondern weil wir lernen, **belastbares Wissen zu erzeugen und daraus für jede Aufgabe den richtigen Kontext bereitzustellen.**

Oder auf einen Satz reduziert:

> **„ Ein guter Prompt kompensiert keine schlechte Wissensbasis.**

Und auch eine riesige Wissensbasis allein schafft noch keine Handlungsfähigkeit.

Erst aus **validiertem Wissen, sauberem Context Engineering, agentischen Arbeitsprozessen und menschlichem Urteil** entsteht ein System, das aus Komplexität belastbare Entscheidungen machen kann.

DDIM.

Der Praxisbericht stammt aus einem realen Kundenprojekt. Marc Kresin veröffentlicht ihn aus seiner Rolle als Leiter der DDIM-Fachgruppe Digitale Transformation & KI.