DaVinciA+

Ein Referenzrahmen für gesteuerte, validierte und transparente KI-Systeme

Published by A.Ward Publications
In collaboration with Brehon AI Solutions

ISBN:978-1-918501-02-5

Kapitel 1 – Zusammenfassung für die Leitungsebene
Kapitel 2 – Geltungsbereich, Zielgruppe und Zweck
Kapitel 3 – Rahmenüberblick
Kapitel 4 – Architektur
Kapitel 5 – Validierungslebenszyklus
Kapitel 6 – Governance und Aufsicht
Kapitel 7 – Compliance-Ausrichtung
Kapitel 8 – Bereitstellungs- und Einführungsmodelle
Kapitel 9 – Fallstudien
Kapitel 10 – Technischer Anhang
Kapitel 11 – Zusammenfassung und Glossar
Anhang A — Mindest-Evidenzpaket für Governance-Review

Kapitel 1 – Zusammenfassung für die Leitungsebene

DaVinciA⁺ ist ein Referenzrahmen für die Governance, Validierung und operative Aufsicht von Systemen der künstlichen Intelligenz.
Er stellt strukturierte Prinzipien, Leitlinien über den Lebenszyklus hinweg sowie Governance-Mechanismen bereit, um die verantwortungsvolle Konzeption, Einführung und den Betrieb von KI in regulierten und unternehmensweiten Umgebungen zu unterstützen.
DaVinciA⁺ ist herstellerunabhängig, technologieneutral und implementierungsunabhängig und ist dafür vorgesehen, ergänzend zu bestehenden regulatorischen, qualitätsbezogenen und risikomanagementbezogenen Standards angewendet zu werden.
Der Rahmen unterstützt Compliance-Aktivitäten, indem er übergeordnete Verpflichtungen – wie Risikomanagement, menschliche Aufsicht, Transparenz und Validierung – in wiederholbare operative Praktiken und Governance-Artefakte übersetzt.
DaVinciA⁺ ist kein Zertifizierungsschema, ersetzt keine anwendbaren Gesetze oder Standards und stellt keine regulatorische Genehmigung oder Rechtsberatung dar.

Systeme der künstlichen Intelligenz sind inzwischen in Entscheidungsprozesse eingebettet, die Sicherheit, Compliance und regulatorische Ergebnisse beeinflussen. In vielen Organisationen werden diese Systeme ohne eine einheitliche Struktur eingesetzt, die Transparenz, Nachvollziehbarkeit und Auditierbarkeit über ihren gesamten Lebenszyklus hinweg sicherstellen kann. DaVinciA⁺ definiert einen Governance- und Validierungsrahmen, der darauf ausgelegt ist, Struktur, Verantwortlichkeit und Evidenzgenerierung für KI-Systeme in solchen Umgebungen durchzusetzen. DaVinciA⁺ wurde entwickelt,

um diese Lücke direkt zu adressieren. DaVinciA⁺ ist ein strukturierter Governance- und Validierungsrahmen, der KI-Systeme über ihren gesamten Lebenszyklus hinweg transparent, nachvollziehbar und auditierbar macht. Er stellt einen Governance- und Validierungsrahmen dar, der nicht als theoretisches Modell, sondern als praktisches Mittel konzipiert ist, um KI so zu strukturieren, dass ihr Verhalten über den gesamten Lebenszyklus hinweg verstanden, überwacht und begründet werden kann. Die Governance von KI-Systemen muss bereits zum Zeitpunkt der Konzeption etabliert werden. Aufsichtsmechanismen, die erst nach der Einführung implementiert werden, sind zwangsläufig unvollständig, da das Systemverhalten bereits Annahmen, Beschränkungen und Entwurfsentscheidungen widerspiegelt, die zuvor verankert wurden. DaVinciA⁺ betont daher, dass Governance explizit innerhalb der Systemarchitektur ausgedrückt wird, vor der Einführung und während des gesamten Betriebs. Nach der Einführung wird das Verhalten bereits Annahmen und Entwurfsentscheidungen widerspiegeln, die lange vor der Berücksichtigung von Aufsichtsmechanismen getroffen wurden. DaVinciA⁺ greift daher an den Grundlagen an. Es betont, dass KI-Systeme durch drei miteinander verbundene Ebenen beschrieben werden – Identität und Zweck, Wissen und Logik sowie Aufsicht und Audit – wobei jede Ebene eine eigene Form von Beschränkung und Verantwortlichkeit bereitstellt. Diese Ebenen schaffen eine stabile Struktur, um die sich der Rest des Systems entwickeln kann, und ermöglichen es Organisationen, KI zu skalieren, ohne die Sichtbarkeit oder Kontrolle darüber zu verlieren, was das System tut oder warum. Für Führungskräfte reduziert diese Struktur operative Unsicherheit, beschleunigt die Vorbereitung auf regulatorische Prüfungen und senkt die langfristigen Kosten von Nacharbeiten, indem disziplinierte Governance frühzeitig eingeführt wird.

Neben dieser architektonischen Struktur unterstützt DaVinciA⁺ einen Lebenszyklusansatz, der auf Qualifizierungspraktiken beruht, die historisch Hochzuverlässigkeitsbranchen vorbehalten waren. Installationsprüfungen, operative Verifikation und Leistungsvalidierung bilden eine progressive Abfolge, die sicherstellt, dass das System in der Konfiguration korrekt ist, sich korrekt verhält und im realen Einsatz korrekt funktioniert. Nach der Einführung erstreckt sich dieselbe Disziplin auf die kontinuierliche Überwachung. Drift wird als erwartetes Phänomen behandelt, nicht als Überraschung; Evidenz wird kontinuierlich statt episodisch gesammelt; und Änderungen werden unter kontrollierter, dokumentierter Überprüfung gesteuert. Der Lebenszyklus validiert nicht lediglich ein Modell – er validiert die gesamte operative Umgebung, in der die KI funktioniert.

Der Bedarf an einer solchen Disziplin ist besonders deutlich geworden, da Organisationen von Einzelmodell-Anwendungsfällen zu Multi-Agenten-Ökosystemen übergehen. Moderne KI agiert typischerweise nicht als einzelnes Modell, das isolierte Prompts beantwortet. Sie ist zunehmend ein Netzwerk spezialisierter Agenten, von denen jeder einen Abschnitt eines Workflows ausführt, jeder auf die Ausgaben anderer angewiesen ist und jeder in der Lage ist, compliance-relevante Ergebnisse zu beeinflussen. In unstrukturierten Umgebungen können diese Agenten driften, unvorhersehbar delegieren oder gegensätzlich agieren. DaVinciA⁺ führt formale Governance in diese Interaktionen ein, durch explizite Grenzen, kontrollierte Delegationspfade und Audit-Mechanismen, die jeden Austausch aufzeichnen. Das Ergebnis ist ein System, in dem Multi-Agenten-Verhalten rekonstruierbar wird, statt emergent oder intransparent zu sein.

DaVinciA⁺ ist bewusst technologieneutral. Es schreibt nicht vor, wie eine Organisation Modelle trainieren muss oder welche Orchestrierungswerkzeuge sie verwenden soll. Stattdessen definiert es die Governance-Erwartungen, die unabhängig von Stack, Sektor oder Anwendungsfall gelten sollen. Diese Neutralität ermöglicht die Integration in bestehende Infrastrukturen – cloudbasierte ML-Pipelines, lokale Rechencluster, Agent-Builder, Workflow-Engines – ohne technologischen Lock-in zu erzwingen. Der Rahmen liegt oberhalb der technischen Substratschicht und schafft Kohärenz über heterogene Systeme hinweg.

Ebenso wichtig ist seine Ausgestaltung unter Bezugnahme auf globale regulatorische Erwartungen. Anstatt Compliance zu beanspruchen, spiegelt DaVinciA⁺ die strukturellen Prioritäten wider, die in maßgeblichen regulatorischen und normativen Rahmenwerken zu finden sind – dem EU-KI-Gesetz, ISO 42001, GAMP 5, MDR/IVDR, ISO 13485 und 14971, IEC 62304 sowie den FDA GMLP – und überführt sie in operative Praktiken. Es stellt die Arten von Prozessen, Artefakten und Nachvollziehbarkeit bereit, die in diesen Rahmenwerken üblicherweise erwartet werden, vermeidet jedoch jede Andeutung, sie zu ersetzen. Es ist eine Governance-Überlagerung, kein Zertifizierungsregime. Organisationen, die es anwenden, unterziehen sich weiterhin allen erforderlichen regulatorischen Bewertungen; DaVinciA⁺ bereitet sie lediglich mit der Evidenz, Disziplin und Dokumentation vor, die diese Bewertungen verlangen.
Über all diese Elemente hinweg – Architektur, Validierung, Aufsicht, Protokollierung und regulatorische Ausrichtung – zieht sich ein zentrales Prinzip durch den Rahmen: KI muss gegenüber der Organisation, die sie einsetzt, rechenschaftspflichtig bleiben. Rechenschaftspflicht ist in diesem Kontext nicht abstrakt. Sie ist die Fähigkeit, mit Evidenz darzulegen, wie das System konzipiert wurde, wie es sich verhält, wie es überwacht wird und wie Risiken gesteuert werden. DaVinciA⁺ ermöglicht diesen Nachweis. Es gibt Organisationen eine Möglichkeit, Ordnung vor Skalierung, Klarheit vor Komplexität und Nachvollziehbarkeit vor der Einführung zu etablieren.
Auf diese Weise rahmt es KI nicht als volatile Fähigkeit, die defensiv zu managen ist, sondern als operatives Asset, das mit derselben Disziplin gesteuert werden kann wie andere kritische Systeme. DaVinciA⁺ ermöglicht es Unternehmen, KI-Initiativen mit Zuversicht voranzutreiben, in dem Wissen, dass Leistung, Sicherheit und Compliance kontinuierlich beobachtbar bleiben und unter menschlicher Autorität stehen. Es verlangsamt Innovation nicht; es liefert die Struktur, die Innovation nachhaltig macht.

Kapitel 2 – Geltungsbereich, Zielgruppe und Zweck

2.1 Geltungsbereich
DaVinciA⁺ gilt für Systeme der künstlichen Intelligenz, deren Ausgaben operative, compliance-relevante, sicherheitskritische oder entscheidungskritische Prozesse beeinflussen. Der Rahmen ist für den Einsatz sowohl in regulierten als auch in nicht regulierten Bereichen vorgesehen, in denen Nachvollziehbarkeit, Verantwortlichkeit und Aufsicht erforderlich sind. „Universell“ bezieht sich auf Governance-Prinzipien, die sektor- und rechtsraumübergreifend anwendbar sind, nicht auf eine einheitliche regulatorische Behandlung oder Risikoklassifizierung.
Der Rahmen ist anwendbar auf:
– Einzelmodell-KI-Systeme, die in operativen Workflows eingesetzt werden
– Multi-Agenten-KI-Systeme, die verteiltes oder delegiertes Schlussfolgern durchführen
– KI-Systeme, die in regulierte Umgebungen integriert sind (einschließlich Gesundheitswesen, MedTech, pharmazeutische, finanzielle und Infrastrukturbereiche)
– Unternehmensweite KI-Implementierungen, die Auditierbarkeit und Governance über den gesamten Lebenszyklus erfordern
DaVinciA⁺ ist bewusst technologieneutral. Es definiert keine spezifischen Modelle, Plattformen, Orchestrierungswerkzeuge oder Infrastrukturarchitekturen vor. Die Governance-Erwartungen bleiben unabhängig von der technischen Implementierung konsistent.

2.2 Expliziter Nicht-Geltungsbereich

DaVinciA⁺
– definiert oder benchmarkt keine Modellleistung oder Genauigkeit
– definiert keine Modelltrainingsmethoden oder den Aufbau von Datensätzen vor
– gibt keine klinischen, rechtlichen oder sicherheitsbezogenen Aussagen zu Systemergebnissen
– ersetzt oder überlagert keine regulatorischen Bewertungen, Zertifizierungen oder Genehmigungen
– fungiert nicht als Konformitätsbewertungs- oder Zertifizierungsschema
Diese Ausschlüsse sind bewusst gewählt. DaVinciA⁺ regelt Struktur, Aufsicht und Evidenz, nicht die Modellfähigkeit oder die Ergebnisleistung.

2.3 Zielgruppe

Dieses Dokument richtet sich an:
– Führungskräfte mit Verantwortung für KI-Risiken, Verantwortlichkeit und Governance
– Fachkräfte aus den Bereichen Regulierung, Qualität und Compliance
– Technische Führungskräfte, die KI-Systeme entwerfen, implementieren oder überwachen
– Auditoren und Governance-Prüfer, die KI-Implementierungen bewerten

2.4 Zweck

DaVinciA⁺ wird als Referenz-Governance-Rahmen veröffentlicht. Sein Zweck ist es, eine gemeinsame Struktur zu etablieren, anhand derer KI-Systeme konsistent entworfen, geprüft und gesteuert werden können, unabhängig von Domäne, Sektor oder technischer Implementierung.

Nicht-Standard-Erklärung

DaVinciA⁺ ist kein Standard, keine Spezifikation und kein Konformitätsbewertungsschema. Es ist ein Referenz-Governance-Rahmen, der dazu dient, formale regulatorische und standardbasierte Prozesse zu unterstützen, nicht zu ersetzen.

Was DaVinciA⁺ ist

Ein Referenz-Governance-Rahmen für KI-Systeme

Ein strukturierter Ansatz für Validierung und Aufsicht

Eine Methode zur Operationalisierung regulatorischer Verpflichtungen

Eine neutrale Überlagerung bestehender Standards und Gesetze

Eine Grundlage für auditbereite KI-Betriebsabläufe

Was DaVinciA⁺ nicht ist

Ein Zertifizierungs- oder Akkreditierungsschema

Eine Regulierungsbehörde

Eine proprietäre Softwareplattform

Ein Ersatz für ISO-, IEC-, MDR-, FDA- oder rechtliche Verpflichtungen

Keine Garantie für regulatorische Genehmigung

Kapitel 3 – Rahmenüberblick

KI-Systeme fungieren als operative Entscheidungsinstrumente und nicht als isolierte technische Komponenten. Ihre Ausgaben beeinflussen regulierte Prozesse, sicherheitskritische Tätigkeiten und die organisatorische Verantwortlichkeit. Mit zunehmendem Einfluss dieser Systeme ist ein kohärenter Rahmen erforderlich, um sicherzustellen, dass sie über ihren gesamten Lebenszyklus hinweg strukturiert, begrenzt und steuerbar bleiben. DaVinciA⁺ etabliert diese Struktur, indem definiert wird, wie KI-Systeme beschrieben, gesteuert und validiert werden, unabhängig von der technischen Implementierung. Mit wachsendem Einfluss wächst auch der Bedarf an einem kohärenten Rahmen, der Struktur, Disziplin und Transparenz in die Art und Weise bringt, wie diese Systeme aufgebaut und betrieben werden. DaVinciA⁺ wurde entwickelt, um diesem Bedarf zu entsprechen, indem es ein einheitliches Modell zur Beschreibung, Governance und Validierung von KI über ihren gesamten Lebenszyklus hinweg bereitstellt.
Der Rahmen beginnt mit einer grundlegenden Annahme: Ein KI-System muss nicht nur anhand der Aufgaben verstanden werden, die es ausführt, sondern anhand der Bedingungen, unter denen es diese ausführt. Die traditionelle Softwareentwicklung erkennt seit Langem die Bedeutung von Zweck, Vorbedingungen, Beschränkungen und Verantwortlichkeit an. KI erfordert eine analoge Struktur, angepasst an Systeme, deren Verhalten aus erlernten Mustern und nicht aus deterministischem Code hervorgeht. DaVinciA⁺ bringt diese Struktur durch drei voneinander abhängige Ebenen zum Ausdruck, die definieren, was das System ist, wie es schlussfolgert und wie es innerhalb seiner autorisierten Grenzen bleibt.
Die erste Ebene etabliert die Identität und den Zweck des Systems. Die Identität definiert, was das System ist; der Zweck definiert, was das System tun darf. Die Trennung dieser Konzepte beseitigt Unklarheiten und verhindert eine unkontrollierte Ausweitung der Systemverantwortlichkeiten. Sie klärt den Bereich, in dem die KI operieren soll, die ihr zugewiesenen spezifischen Verantwortlichkeiten sowie die Grenzen, innerhalb derer sie verbleiben muss. Dadurch wird Unklarheit an der Quelle beseitigt. Durch die Formalisierung von Mission und Beschränkungen des Systems verhindert DaVinciA⁺ eine Ausweitung des Anwendungsbereichs und ermöglicht, dass jede nachfolgende Entwurfsentscheidung an diesen anfänglichen Verpflichtungen gemessen werden kann. Diese Ebene identifiziert zudem die menschlichen Stakeholder, die für die Ausgaben des Systems verantwortlich sind, und verankert die Governance in persönlicher und organisatorischer Verantwortung.
Die zweite Ebene betrifft Wissen und Logik – die interne Mechanik, durch die die KI Eingaben interpretiert, Informationen bewertet und Ausgaben erzeugt. In konventionellen Implementierungen werden diese Mechanismen häufig durch Abstraktion verschleiert. DaVinciA⁺ betont, dass sie artikuliert und, wo möglich, begrenzt werden. Es definiert die Datenquellen, auf die das System zugreifen darf, die Formen des Schlussfolgerns, die es anwenden darf, die Werkzeuge, die es aufrufen darf, sowie die Leitplanken, die seine Entscheidungen formen. Durch die explizite Erfassung dieser Elemente und deren

Verwaltung unter Versionskontrolle stellt DaVinciA⁺ die Nachvollziehbarkeit bereit, die für Untersuchungen, Überwachung und regulatorische Prüfungen erforderlich ist. Das Schlussfolgern des Systems wird zu einem gesteuerten Raum statt zu einer Blackbox.

Die dritte Ebene befasst sich mit Aufsicht und Audit. Kein KI-System sollte ohne einen klaren Mechanismus für Überwachung, Eskalation und kontinuierliche Evidenzgenerierung betrieben werden. Diese Ebene führt strukturierte Kontrollpunkte ein, die identifizieren, wann menschliches Eingreifen erforderlich ist, wann Entscheidungen autorisierte Grenzen überschreiten und wann Ausgaben eine Verifikation verlangen. Sie schreibt zudem ein umfassendes Auditprotokoll bereit, das Handlungen, Kontext und Begründung nachverfolgt. Dieses Protokoll existiert nicht um seiner selbst willen; es schafft die Voraussetzungen dafür, dass Organisationen Verantwortlichkeit nachweisen, Anomalien untersuchen und externe Prüfungen erfüllen können.

Diese drei Ebenen bilden den Kern der DaVinciA⁺-Architektur, doch der Rahmen geht über die strukturelle Beschreibung hinaus und erstreckt sich auf operative Methodik. Die DaVinciA-Technik liefert die philosophische Grundlage dafür, wie KI innerhalb dieser Architektur aufgebaut werden soll. Sie betont Zweckklarheit, ökonomisches Design, kontrolliertes Schlussfolgern und den Respekt vor menschlicher Aufsicht. Sie entmutigt unnötige Komplexität, unkontrollierte Delegation und mehrdeutiges Verhalten. In der Wirkung versucht sie, Intentionalität in einen Technologiebereich zurückzubringen, der sich häufig schneller entwickelt, als Governance reagieren kann.

Die Relevanz des Rahmens wird besonders deutlich in Umgebungen, in denen mehrere KI-Agenten zusammenarbeiten sollen. Ohne Struktur können Multi-Agenten-Systeme Delegationsketten erzeugen, die schwer zu beobachten oder zu rekonstruieren sind. DaVinciA⁺ führt explizite Schnittstellen zwischen Agenten ein, definiert zulässige Kommunikationswege und ermöglicht, dass jeder Austausch in einem einheitlichen Audit-Trail erfasst wird. Dies transformiert ansonsten dynamisches, lose begrenztes Verhalten in eine Abfolge kontrollierter Interaktionen, die überprüft, getestet und begründet werden können.

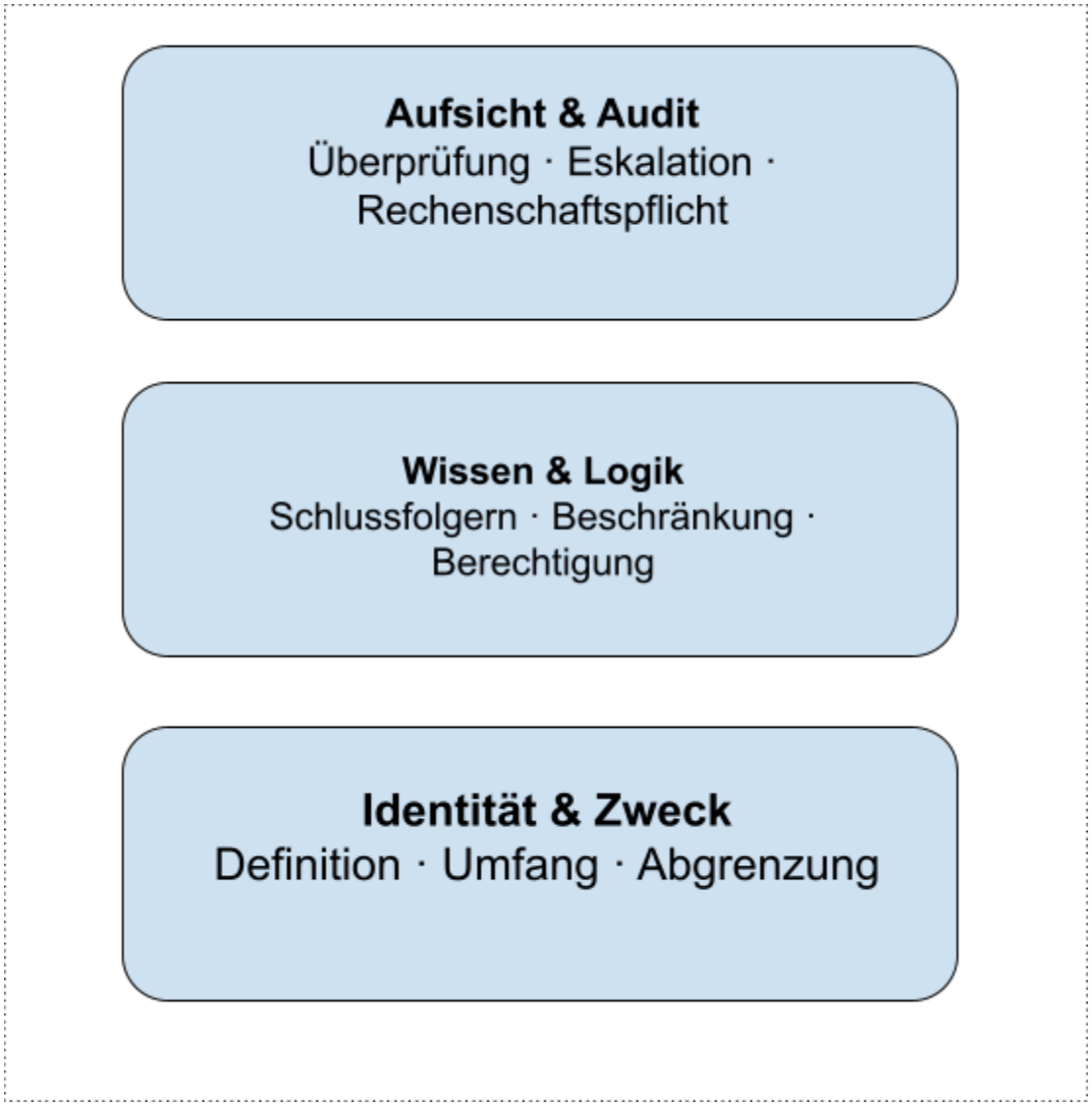

Abbildung 1 — Kernarchitektur der Governance

Konzeptionelle Governance-Ebenen, die gleichzeitige Beschränkungen für KI-Systeme darstellen.

Diese Abbildung zeigt keinen Systemfluss, keine Ausführungsreihenfolge und keine technische Implementierung.

Ein wesentliches Merkmal von DaVinciA⁺ ist seine Neutralität. Es betont nicht, dass Organisationen bestimmte Modelle, Plattformen oder Orchestrierungssysteme einsetzen. Stattdessen stellt es eine Governance-Ebene bereit, die über Cloud-Umgebungen, lokale HPC-Installationen und agentische Orchestrierungswerkzeuge hinweg anwendbar ist. Die technische Implementierung kann variieren; die Governance-Prinzipien nicht. Diese Designentscheidung ermöglicht es DaVinciA⁺, als integrierender Referenzrahmen innerhalb komplexer Unternehmensarchitekturen zu fungieren und Konsistenz zu schaffen, selbst wenn sich die zugrunde liegenden Werkzeuge unterscheiden.

„Konzeptionelle Governance-Architektur. Pfeile stellen Informations- und Beschränkungsbeziehungen dar, keine technischen Datenpipelines.“

Der Rahmen spiegelt die strukturelle Entwicklung aufkommender KI-Regulierungen wider, indem er Prinzipien operationalisiert, die konsistent im EU-KI-Gesetz, in ISO 42001, GAMP 5, MDR/IVDR und in FDA-Leitlinien erscheinen – ohne Konformität zu implizieren. Auch wenn er keine Compliance behauptet, sind seine Konzepte von den Erwartungen geprägt, die im EU-KI-Gesetz, in ISO 42001, GAMP 5, MDR/IVDR, ISO 13485, ISO 14971 und IEC 62304 formuliert sind. Diese Regelwerke teilen einen gemeinsamen Schwerpunkt auf dokumentiertem Risikomanagement, Transparenz, verantwortungsvoller Aufsicht und Disziplin über den gesamten Lebenszyklus hinweg. DaVinciA⁺ überträgt diese Erwartungen operativ in praktische Governance-Mechanismen, die Organisationen frühzeitig übernehmen können, lange bevor formale regulatorische Verpflichtungen greifen.
In ihrer Gesamtheit schaffen diese Elemente ein kohärentes Governance-Modell für KI-Systeme. DaVinciA⁺ bietet eine Möglichkeit, darzustellen, was ein KI-System ist, wie es Entscheidungen trifft, wie es begrenzt ist und wie sein Verhalten über die Zeit hinweg überwacht wird. Es ermöglicht Organisationen, KI zu entwickeln, die nicht nur funktional, sondern rechenschaftspflichtig ist, nicht nur leistungsfähig, sondern bewusst gestaltet. Es ersetzt reaktive Compliance durch proaktive Struktur und bietet eine Grundlage, auf der sichere, skalierbare und vertrauenswürdige KI entwickelt werden kann.

Kapitel 4 – Architektur

Die Systemarchitektur definiert die Bedingungen, unter denen Intelligenz operieren darf. In KI-Systemen steuert die Architektur das Verhalten unter Unsicherheit, bestimmt, wie sich Risiken manifestieren, und legt fest, ob Verantwortlichkeit mit Evidenz belegt werden kann. DaVinciA⁺ behandelt Architektur als ein Governance-Konstrukt und nicht als einen Software-Blueprint. Sie definiert die Bedingungen, unter denen Intelligenz agieren darf, die Grenzen, innerhalb derer Entscheidungen gebildet werden, sowie die Mechanismen, durch die Aufsicht aufrechterhalten wird. Diese Perspektive verlagert den Fokus von einzelnen Modellfähigkeiten auf das übergeordnete System, das diese enthält und begrenzt.

Bedrohungsmodellierung und Überlegungen zu Fehlermodi
Mit zunehmender Komplexität und Autonomie von KI-Systemen können nicht offensichtliche Fehlermodi und Bedrohungsvektoren entstehen, die durch funktionale Tests allein nicht unmittelbar erkennbar sind. DaVinciA⁺ erkennt Bedrohungsmodellierung als eine wichtige ergänzende Praxis zur Antizipation und Analyse solcher Risiken an.
Organisationen können etablierte Rahmenwerke wie STRIDE für Cybersicherheitsbedrohungen und

LINDDUN für Datenschutz-Folgenabschätzungen anwenden, sofern dies angemessen ist. Für KI-spezifische Risiken – einschließlich unkontrollierter Delegationsschleifen, Missbrauch von Werkzeugen, unterlassener Eskalation oder Zusammenbruch von Schlussfolgerungspfaden – stellt DaVinciA⁺ ein nachvollziehbares, lauf- und schrittbasiertes Auditmodell bereit, das sowohl eine nachträgliche Analyse als auch Vorabtests vor der Einführung im Rahmen der Operational Qualification (OQ) ermöglicht.
Zukünftige Anhänge werden repräsentative Bedrohungsbäume und Fehlermuster formalisieren, die aus diesem Nachvollziehbarkeitsmodell abgeleitet sind. Diese Materialien werden beratender Natur sein und keine spezifischen Minderungsmaßnahmen oder Implementierungsentscheidungen vorschreiben.

Die Architektur beginnt mit der Prämisse, dass ein KI-System über die Rollen, die es erfüllt, die Informationen, die es nutzt, und die Kontrollen, die sein Verhalten formen, verstanden werden muss. In praktischer Hinsicht erfordert dies Klarheit über den Zweck des Systems und die Umgebung, in der es funktioniert. DaVinciA⁺ führt daher eine formale Beschreibung von Identität und Zweck als grundlegendes architektonisches Element ein. Diese Beschreibung legt fest, was das System erreichen soll, welche Grenzen es einhalten muss und welche Verantwortlichkeiten ausschließlich beim Menschen verbleiben. Sie ermöglicht, dass Entwurfsentscheidungen anhand einer expliziten Zweckbestimmung bewertet werden können, wodurch das Risiko von Scope Drift oder unbeabsichtigter Funktionsausweitung reduziert wird.
Auf dieser Grundlage aufbauend beschreibt die Architektur das Wissen und die Logik, die die Entscheidungen des Systems leiten. KI-Modelle operieren häufig in einem breiten und lose abgegrenzten Informationsraum und greifen auf Daten und Werkzeuge zurück, die sich im Laufe der Zeit verändern können. Um dieser Tendenz zur Intransparenz entgegenzuwirken, betont DaVinciA⁺ einen definierten Satz von Wissensquellen, Schlussfolgerungsprozessen und zulässigen Handlungen. Es definiert, dass die Mechanismen, durch die das System Daten interpretiert, Werkzeuge aufruft und den Kontext bewertet, dokumentiert und versioniert werden. Diese Struktur ermöglicht es Organisationen, nachzuvollziehen, wie Entscheidungen gebildet werden, zu bewerten, ob diese Entscheidungen innerhalb politischer und regulatorischer Grenzen bleiben, und Abweichungen oder unerwartete Ergebnisse zu untersuchen.
Aufsicht und Audit bilden die dritte strukturelle Komponente der Architektur. Kein KI-System, gleich wie gut es entworfen ist, sollte außerhalb der Reichweite von Aufsicht betrieben werden. DaVinciA⁺ verankert Aufsicht daher direkt in der Architektur, anstatt sie als externe oder optionale Ebene zu behandeln. Es definiert, wann eine menschliche Überprüfung erforderlich ist, wie Unsicherheit oder Konflikte zu eskalieren sind und welche Evidenz in jeder Betriebsphase zu erzeugen ist. Dazu gehört die Aufzeichnung der Schlussfolgerungen des Systems, die Dokumentation der Werkzeugnutzung sowie die Erfassung kontextueller Details, die eine nachträgliche Analyse ermöglichen. Durch die Integration der Aufsicht in die Architektur selbst stellt DaVinciA⁺ sicher, dass Verantwortlichkeit nicht auf retrospektiver Rekonstruktion beruht, sondern kontinuierlich während des Systembetriebs erzeugt wird.
Diese architektonischen Elemente werden in Multi-Agenten-Umgebungen besonders wichtig. Mit der Einführung agentischer Workflows verlassen sich einzelne Komponenten zunehmend auf die Ausgaben anderer. Ohne Struktur können diese Interaktionen Verhaltensweisen erzeugen, die schwer vorhersehbar oder überprüfbar sind. DaVinciA⁺ begegnet dem, indem es kontrollierte Kommunikationspfade zwischen Agenten definiert und betont, dass jede Interaktion im Audit-Trail erfasst wird. Es ermöglicht, dass Delegation innerhalb autorisierter Grenzen erfolgt, dass kein Agent seinen Umfang eigenständig erweitern kann und dass menschliche Aufsicht ausgelöst wird, wenn Interaktionen Unsicherheit oder Risiko erzeugen. Dadurch wird potenziell emergentes oder intransparentes Verhalten in eine Abfolge

rechenschaftspflichtiger Schritte überführt. In der Praxis bedeutet dies, dass ein Agent nur über autorisierte Pfade delegieren darf, seinen Umfang nicht autonom ändern kann und für jede Interaktion Auditdaten erzeugen muss. Diese Kontrollen wandeln dynamisches Agentenverhalten in rekonstruierbare, überprüfbare Sequenzen um.

Die Architektur berücksichtigt auch die praktischen Realitäten des Unternehmenseinsatzes. KI-Systeme sind selten statisch; Modelle werden ersetzt, Datensätze entwickeln sich weiter, Werkzeuge werden hinzugefügt und Workflows ändern sich. DaVinciA⁺ antizipiert diese Dynamik, indem Mechanismen für kontrollierte Änderungen in das architektonische Design eingebettet werden. Konfiguration und Logik verbleiben unter Versionskontrolle, sodass Organisationen nachvollziehen können, wie Aktualisierungen das Verhalten beeinflussen. Validierungskontrollpunkte stellen sicher, dass modifizierte Systeme weiterhin innerhalb ihres vorgesehenen Umfangs operieren. Auditprotokolle liefern die Evidenz, die erforderlich ist, um nachzuweisen, dass Änderungen verantwortungsvoll und unter angemessener Aufsicht umgesetzt wurden.

Eine der Stärken der DaVinciA⁺-Architektur ist ihre Unabhängigkeit von einem bestimmten Technologie-Stack. Die Governance-Prinzipien gelten gleichermaßen für cloudbasierte Dienste, On-Premise-Plattformen, Workflow-Orchestratoren und Agenten-Building-Toolkits. Diese Neutralität ermöglicht es der Architektur, als vereinheitlichende Ebene über diverse Systeme hinweg zu fungieren und Organisationen einen konsistenten Rahmen zu bieten, selbst wenn sich technische Komponenten zwischen Abteilungen oder Projekten unterscheiden. Der Schwerpunkt liegt auf Struktur, Nachvollziehbarkeit und Kontrolle und nicht auf den technischen Details der Modellentwicklung.

In ihrer Gesamtheit ist die Architektur darauf ausgelegt, sicherzustellen, dass KI-Systeme über ihren gesamten Lebenszyklus hinweg verständlich, kontrollierbar und rechenschaftspflichtig bleiben. Sie bietet die notwendigen Beschränkungen für einen sicheren Betrieb, ohne Innovation zu behindern oder die Modellauswahl einzuschränken. Durch die Definition, wie Zweck, Schlussfolgerung und Aufsicht auszudrücken sind, bietet DaVinciA⁺ einen praktikablen Weg zu einer verantwortungsvollen Einführung im großen Maßstab. Sie schafft eine stabile Grundlage, auf der komplexe KI-Fähigkeiten aufgebaut, integriert und mit Vertrauen gesteuert werden können.

Kapitel 5 – Validierungslebenszyklus

Die Validierung von KI-Systemen ist eine kontinuierliche Lebenszyklusaktivität und keine punktuelle Bewertung. DaVinciA⁺ übernimmt ein Qualifizierungsmodell, das aus Hochzuverlässigkeitsbranchen abgeleitet ist, und betont den Nachweis, dass Systeme korrekt konfiguriert sind, innerhalb definierter Beschränkungen operieren und unter realen Einsatzbedingungen dauerhaft zweckgeeignet bleiben. Die traditionelle Softwarevalidierung geht von deterministischem Verhalten aus: Sobald ein System installiert und getestet ist, bleiben seine Ausgaben vorhersehbar, sofern keine expliziten Änderungen vorgenommen werden. KI-Systeme stellen diese Annahme in Frage. Ihr Verhalten hängt nicht nur vom Code ab, sondern von Modellparametern, Datenverteilungen, Werkzeuginteraktionen und der weiteren Betriebsumgebung. Aus diesem Grund verfolgt DaVinciA⁺ einen lebenszyklusbasierten Validierungsansatz, der sicherstellt, dass das System sowohl zum Zeitpunkt der Einführung als auch über die Zeit hinweg nachweislich zweckgeeignet bleibt.

Der Lebenszyklus beginnt mit der Feststellung, dass das System korrekt konfiguriert wurde. Die Installationsqualifizierung (Installation Qualification, IQ) verifiziert, dass alle Komponenten – Modelle, Werkzeuge, Orchestratoren, Datenquellen und Leitplanken – wie vorgesehen implementiert sind und den dokumentierten Spezifikationen entsprechen. In konventionellen Systemen ist dieser Schritt unkompliziert; in KI-Implementierungen erfordert er zusätzliche Sorgfalt, da Änderungen an Modellversionen, Umgebungseinstellungen oder Werkzeugberechtigungen das Systemverhalten wesentlich verändern können. DaVinciA⁺ behandelt die Konfiguration als kontrolliertes Artefakt, um sicherzustellen, dass die strukturelle Integrität des Systems von Beginn an gewahrt bleibt.

Die Betriebsqualifizierung (Operational Qualification, OQ) untersucht, wie sich das System unter erwarteten Bedingungen verhält. Ziel ist nicht lediglich die Funktionsprüfung, sondern das Verständnis der Konturen des systeminternen Schlussfolgerns sowie die Bestätigung, dass Leitplanken, Eskalationspfade und Aufsichtsmechanismen wie vorgesehen reagieren. Diese Phase ermöglicht, dass das System seinen definierten Umfang respektiert, mit Unsicherheit angemessen umgeht und Ausgaben erzeugt, die innerhalb politischer und regulatorischer Beschränkungen bleiben. Hier wird die Unterscheidung zwischen korrektem Betrieb und korrektem Ergebnis entscheidend: Ein KI-System kann Ausgaben erzeugen, die plausibel erscheinen, und dennoch interne Regeln verletzen oder Aufsicht umgehen. DaVinciA⁺ betont die Notwendigkeit, das Verhalten zu validieren, nicht nur die Ergebnisse. Diese Unterscheidung ist für KI zentral: Ein System kann akzeptable Ausgaben liefern, während es interne Regeln verletzt, Aufsichtsschritte überspringt oder Eskalationsauslöser umgeht. Die Verhaltensvalidierung stellt strukturelle Compliance sicher, nicht nur Ergebnisplausibilität.

Die Leistungsqualifizierung (Performance Qualification, PQ) konzentriert sich darauf, ob das System in seinem realen Einsatzkontext zuverlässig arbeitet. Im Gegensatz zu den früheren Phasen, die in kontrollierten Umgebungen durchgeführt werden, betrachtet diese Phase das System innerhalb laufender Workflows, im Zusammenspiel mit tatsächlichen Nutzern, Daten und operativen Anforderungen. Zweck der PQ ist nicht die Feststellung von Perfektion, sondern der Nachweis, dass das System Stabilität wahrt, dass Aufsicht wirksam bleibt und dass Abweichungen erkannt und adressiert werden. KI-Systeme, die in regulierten oder sicherheitskritischen Umgebungen betrieben werden, erfordern in dieser Phase besondere Aufmerksamkeit, da kontextuelle Faktoren das Verhalten auf subtile Weise beeinflussen können, die durch kontrollierte Tests nicht vollständig vorhersehbar sind.

Die Validierung endet nicht mit der Einführung. KI-Systeme entwickeln sich weiter, wenn sich

Umgebungen ändern, Werkzeuge aktualisiert werden und Modelle neu trainiert oder ersetzt werden. DaVinciA⁺ verankert daher die kontinuierliche Überwachung als zentrales Element des Lebenszyklus. Drift – sei sie statistischer, verhaltensbezogener oder kontextueller Art – wird vorausgesetzt und nicht als Anomalie behandelt. Die Überwachung erfasst Evidenz über Systemläufe hinweg und ermöglicht es Organisationen, aufkommende Risiken zu identifizieren, zu bewerten, ob das Verhalten innerhalb der Spezifikation bleibt, und festzustellen, wann eine Revalidierung erforderlich ist. Diese kontinuierliche Evidenzakkumulation ermöglicht, dass Compliance nicht stillschweigend im Laufe der Zeit erodiert.
Die Änderungssteuerung (Change Control) stellt die Governance-Struktur für Aktualisierungen bereit. DaVinciA⁺ behandelt jede Änderung an Modellen, Prompts, Logikflüssen, Werkzeugberechtigungen oder Wissensquellen als eine Änderung, die einer dokumentierten Überprüfung bedarf. Ziel ist nicht, Iteration zu behindern, sondern sicherzustellen, dass Aktualisierungen mit klarer Absicht umgesetzt und durch Evidenz gestützt werden. Änderungen sollen im Hinblick auf ihre potenziellen Auswirkungen auf Verhalten, Sicherheit und Compliance-Verpflichtungen bewertet werden. Wo erforderlich, wird das System in frühere Phasen des Validierungslebenszyklus zurückgeführt, um zu bestätigen, dass es weiterhin innerhalb der ursprünglich definierten Grenzen operiert.
Während dieses gesamten Lebenszyklus spielt die Dokumentation eine zentrale Rolle. Validierungsartefakte sind kein administratives Nebenprodukt; sie sind das Mittel, durch das eine Organisation nachweist, dass sie das System verstanden, gesteuert und verantwortungsvoll kontrolliert hat. Installationsnachweise, Testergebnisse, Leistungsbeobachtungen, Überwachungsprotokolle und Änderungshistorien bilden einen kohärenten Evidenzkörper, der regulatorische Anfragen, interne Untersuchungen und laufende Verantwortlichkeit unterstützt. DaVinciA⁺ stellt die Struktur bereit, die erforderlich ist, um diese Evidenz konsistent und in einer Form zu erzeugen, die an den Erwartungen von Regulierungsbehörden und Normungsgremien ausgerichtet ist.
Das Lebenszyklusmodell ermöglicht, dass KI-Systeme über ihren gesamten operativen Horizont hinweg rechenschaftspflichtig bleiben. Es erkennt an, dass Validierung keine statische Zertifizierung ist, sondern ein lebendiger Prozess, der sich an die sich entwickelnde Natur von KI anpassen muss. Durch die Verankerung strukturierter Kontrollpunkte, kontinuierlicher Überwachung und disziplinierter Änderungssteuerung bietet DaVinciA⁺ Organisationen die Mittel, das Vertrauen in ihre Systeme auch bei veränderten Rahmenbedingungen aufrechtzuerhalten. Es stellt einen praxisnahen, rigorosen Ansatz bereit, um sicherzustellen, dass KI in allen Phasen ihres Einsatzes sicher, vorhersehbar und mit ihrem beabsichtigten Zweck in Einklang bleibt.

„Lebenszyklusdarstellung (IQ → OQ → PQ → Überwachung). Die Abfolge ist konzeptionell und kann im Rahmen der Änderungssteuerung oder der Behebung von Drift rückgekoppelt werden.“

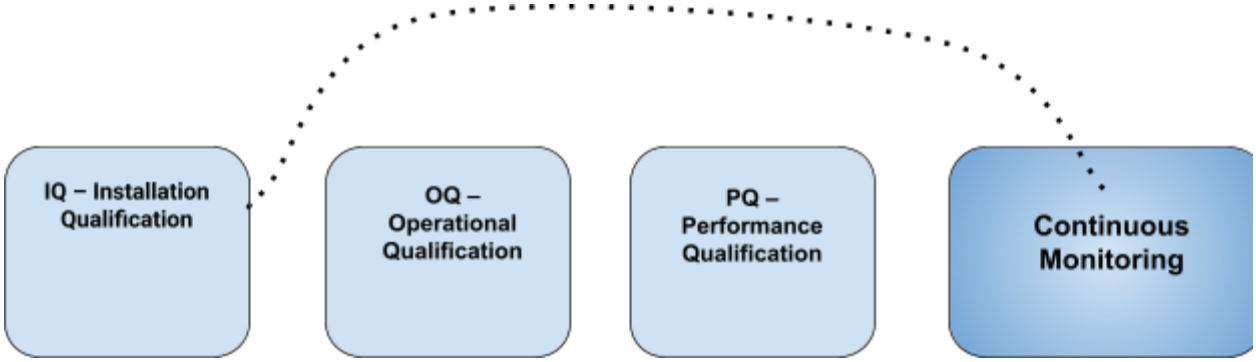

Abbildung 2 — Validierungslebenszyklus-Zustände
Konzeptionelle Validierungszustände zur Veranschaulichung kontinuierlicher und bedingter Qualifizierung.
Die Abfolge ist illustrativ und schreibt weder Prozessreihenfolge, Zeitpunkte noch Implementierung vor.

Kapitel 6 – Governance und Aufsicht

Governance ist der Mechanismus, durch den eine Organisation Autorität über ihre KI-Systeme ausübt. Sie ermöglicht, dass Verantwortung eindeutig zugewiesen ist, dass Entscheidungsprozesse transparent bleiben und dass Systeme innerhalb definierter ethischer, operativer und regulatorischer Grenzen betrieben werden. DaVinciA⁺ betrachtet Governance nicht als periphere Aktivität, sondern als integralen Bestandteil des Systemdesigns. Aufsicht muss in die Architektur eingebettet, durch operative Praktiken ausgedrückt und durch Evidenz gestützt sein, die interner wie externer Prüfung standhält.
Ein zentrales Prinzip des Rahmens ist, dass KI-Systemen keine implizite Autonomie eingeräumt werden darf. Selbst hochleistungsfähige Modelle sollen innerhalb definierter Beschränkungen und unter der Aufsicht identifizierbarer menschlicher Rollen operieren. Governance beginnt daher mit der Festlegung, wer für das Verhalten des Systems verantwortlich ist. Dazu gehören Personen, die für die Definition des Umfangs, die Aufrechterhaltung der Konfiguration, die Überwachung des Betriebs und die Genehmigung von Änderungen zuständig sind. In regulierten Umgebungen entsprechen diese Verantwortlichkeiten in natürlicher Weise bestehenden Funktionen in den Bereichen Qualität, Klinik, Regulierung und Technik. DaVinciA⁺ stellt eine Struktur bereit, durch die diese Verantwortlichkeiten klar und konsistent ausgedrückt werden können. Typische Aufsichtsrollen umfassen:
– System Owner – verantwortlich für die Definition des Zwecks und die Genehmigung von Grenzen
– Qualitäts-/Regulierungsleitung – ermöglicht, dass Prozesse organisatorischen und regulatorischen Erwartungen entsprechen
– Operativer Prüfer – führt Human-in-the-Loop-Bewertungen bei Unsicherheit oder Eskalation durch

Diese Rollen erhalten menschliche Autorität über den gesamten Lebenszyklus hinweg, ohne Innovation einzuschränken.

Aufsicht wird anschließend durch eine Kombination aus prozeduralen und technischen Kontrollen umgesetzt. Prozedural sollen Organisationen festlegen, wann eine menschliche Überprüfung erforderlich ist, wie Unsicherheit oder Risiko eskaliert wird und welche Dokumentation automatisierte Entscheidungen begleiten muss. Technisch wird Aufsicht durch Leitplanken durchgesetzt, die das Systemverhalten begrenzen, sowie durch Mechanismen, die jede Handlung in einer für Analysen geeigneten Form aufzeichnen. Diese duale Struktur ermöglicht, dass Aufsicht sowohl im operativen Betrieb als auch im Auditkontext wirksam ist, in dem Evidenz erforderlich sein kann, um nachzuvollziehen, wie eine Entscheidung zustande kam.

KI-Systeme, die auf mehreren Agenten beruhen, bringen zusätzliche Governance-Herausforderungen mit sich. Wenn Agenten zusammenarbeiten oder Aufgaben aneinander delegieren, kann sich der Entscheidungsprozess über mehrere Komponenten verteilen. Ohne Struktur kann diese Verteilung Verantwortlichkeit verschleiern oder Verhalten erzeugen, das schwer interpretierbar ist. DaVinciA⁺ adressiert diese Herausforderung, indem es kontrollierte Pfade zwischen Agenten definiert und betont, dass jede Interaktion im Auditprotokoll erfasst wird. Delegation darf nicht außerhalb autorisierter Routen erfolgen, und Agenten dürfen ihren Verantwortungsbereich nicht autonom erweitern oder ihre operativen Grenzen verändern. Auf diese Weise bewahrt der Rahmen Klarheit, selbst wenn Workflows komplex werden.

Ein weiterer wesentlicher Aspekt der Aufsicht ist der Umgang mit Unsicherheit. KI-Systeme operieren häufig in Kontexten, in denen Eingabedaten unvollständig, mehrdeutig oder inkonsistent sind. In solchen Situationen darf sich die Entscheidungsfindung nicht ausschließlich auf algorithmische Inferenz stützen. DaVinciA⁺ betont, dass Systeme Unsicherheit erkennen und angemessen eskalieren. Menschliche Aufsicht ist kein theoretischer Schutzmechanismus, sondern eine operative Komponente, die in das Verhalten des Systems eingewoben ist. Eskalationskriterien sollen explizit, dokumentiert und getestet sein, sodass menschliche Prüfer eingreifen, wenn ihr Urteil erforderlich ist.

Governance erstreckt sich auch auf die Dokumentation und Evidenz, die den Systembetrieb begleiten. Organisationen sollen nicht nur nachweisen können, dass ein System eine Aufgabe akzeptabel ausgeführt hat, sondern dass es innerhalb autorisierter Prozesse, unter Verwendung genehmigter Schlussfolgerungen und Datenquellen sowie unter wirksamer Aufsicht operiert hat. DaVinciA⁺ definiert, dass Auditprotokolle Kontext, Schlussfolgerung, Werkzeugnutzung und Ergebnisse jedes Systemlaufs erfassen. Diese Aufzeichnung ermöglicht die Untersuchung von Anomalien, unterstützt regulatorische Anfragen und bildet die Evidenzbasis für kontinuierliche Verbesserung. Sie verwandelt Aufsicht von einer abstrakten Erwartung in einen praktischen, überprüfbaren Prozess.

Wesentlich ist, dass Governance anpassungsfähig bleibt. Mit der Weiterentwicklung regulatorischer Rahmenwerke, mit Modelländerungen und mit der Ausweitung des KI-Einsatzes sollen sich Aufsichtsmechanismen mitentwickeln. DaVinciA⁺ stellt eine Struktur bereit, die flexibel genug ist, um neue Anforderungen aufzunehmen, ohne die Stabilität des Systems zu untergraben. Der Schwerpunkt auf Dokumentation, Auditierbarkeit und kontrollierter Entscheidungsfindung ermöglicht, dass Aktualisierungen methodisch integriert werden können, mit klarem Verständnis ihrer Auswirkungen auf Verantwortlichkeiten und Risiken.

In ihrer Gesamtheit schaffen diese Praktiken ein Governance-Umfeld, in dem KI unter bewusster menschlicher Kontrolle bleibt. Verantwortlichkeit ist explizit statt implizit, Aufsicht ist kontinuierlich statt episodisch, und Evidenz entsteht organisch während des Systembetriebs. DaVinciA⁺ hilft Organisationen, sich von informeller Überwachung hin zu einem strukturierten, transparenten und verteidigungsfähigen Governance-Modell zu entwickeln, das sowohl operative Anforderungen als auch regulatorische

Erwartungen unterstützt. Es transformiert Aufsicht von einer reaktiven Tätigkeit in einen grundlegenden Bestandteil verantwortungsvoller KI-Einführung.

Eskalationsschwellen-Matrix (Auszug aus der Vorlage DMS-GOV-011)

„Illustrativer Governance-Entscheidungsfluss. Eskalationsschwellen und Unsicherheitsauslöser werden von der einsetzenden Organisation definiert. Kein vorschreibender Algorithmus."

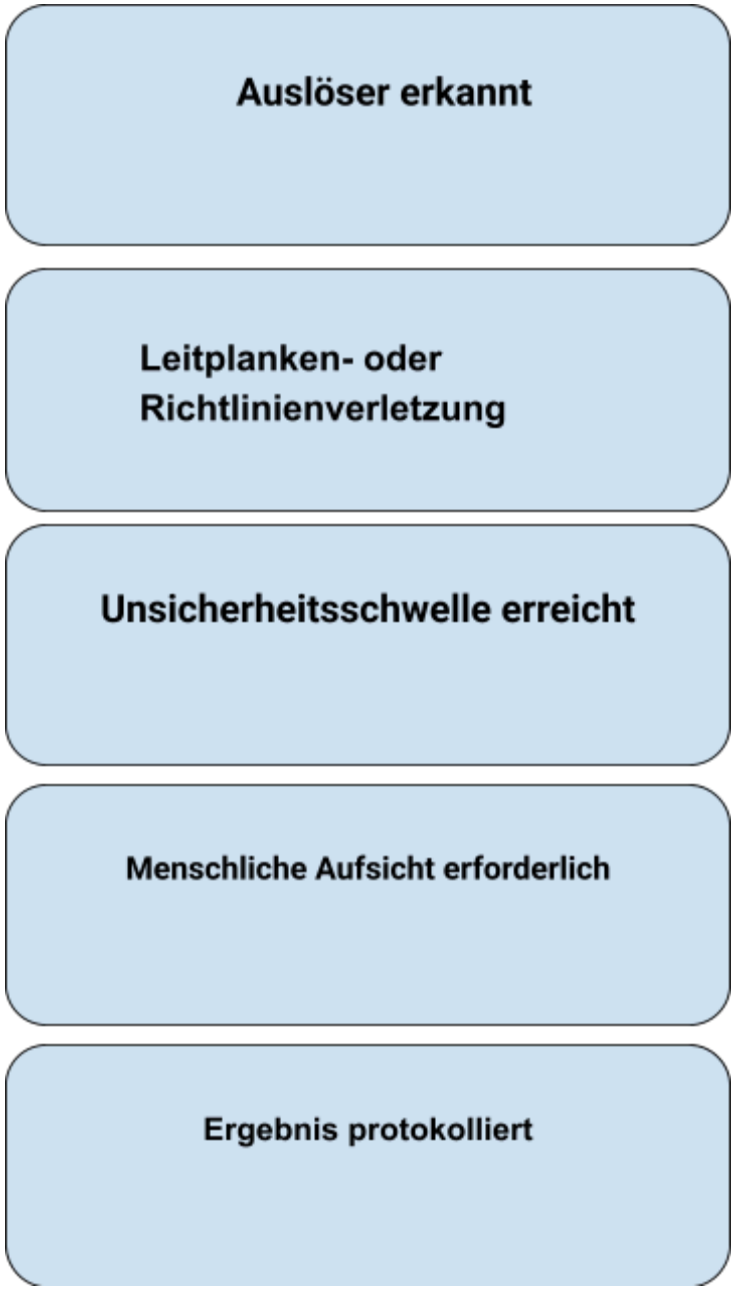

Abbildung 3 — Eskalationsschwellenstruktur
Illustrative Darstellung von Eskalationsbedingungen und Governance-Reaktionen.
Diese Abbildung stellt keine ausführbare Logik, keine automatisierte Entscheidungsfindung und keinen System-Kontrollfluss dar.

DaVinciA⁺ definiert Eskalation nicht als ad-hoc-Reaktion, sondern als strukturierten Governance-Mechanismus, der durch klar definierte Bedingungen ausgelöst wird. Die nachstehende Tabelle illustriert repräsentative Eskalationsschwellen und entsprechende Aufsichtsmaßnahmen. Diese Schwellen sind durch die einsetzende Organisation konfigurierbar und sollen ermöglicht, dass Unsicherheit, Grenzverletzungen oder entstehende Risikosituationen unter dokumentierter menschlicher Autorität adressiert werden.

Trigger-Bedingung	Eskalationsziel	Aufsichtsrolle	Erforderliche Maßnahme	Zeitkritikalität
Leitplankenverletzung (z. B. Überschreitung von Bias- oder Richtlinienschwellen)	Human-in-the-Loop-Prüfer	Operativer Prüfer	Genehmigen, blockieren oder umleiten der Systemausgabe	Unmittelbar
Vertrauensniveau unter definierter Schwelle	QA-Leitung	Qualität / Regulierung	Begründung verlangen, überprüfen oder erneut testen	24–48 Stunden
Delegation außerhalb autorisierter Pfade	System Owner	Risikoverantwortlicher	Systemstopp, Protokollierung und Revalidierung	Unmittelbar
Unbekannte Eingabeklasse oder Verschiebung der Datenverteilung	Eskalationsgremium	QA & RA	Überprüfung der Datenherkunft und Kennzeichnung von Drift	≤ 72 Stunden

Diese Struktur spiegelt die in Vorlage 2.2 des DaVinciA⁺ Governance-&-Oversight-Moduls definierte Eskalationslogik wider und zeigt, wie Governance-Intention in auditierbares Systemverhalten operationalisiert wird.

Risikoklassifikationsmodell

Risikostufe
Beschreibung
Aufsichtserfordernis

Stufe 1: Minimal
Nicht-kritische, reversible Ausgaben
Periodische Überprüfung

Stufe 2: Moderat
Indirekter Einfluss auf Sicherheit/Compliance
HITL-Eskalation bei Drift

Stufe 3: Kritisch
Patientensicherheit, finanzielles Risiko, rechtliche Exposition
HITL immer + Revalidierung

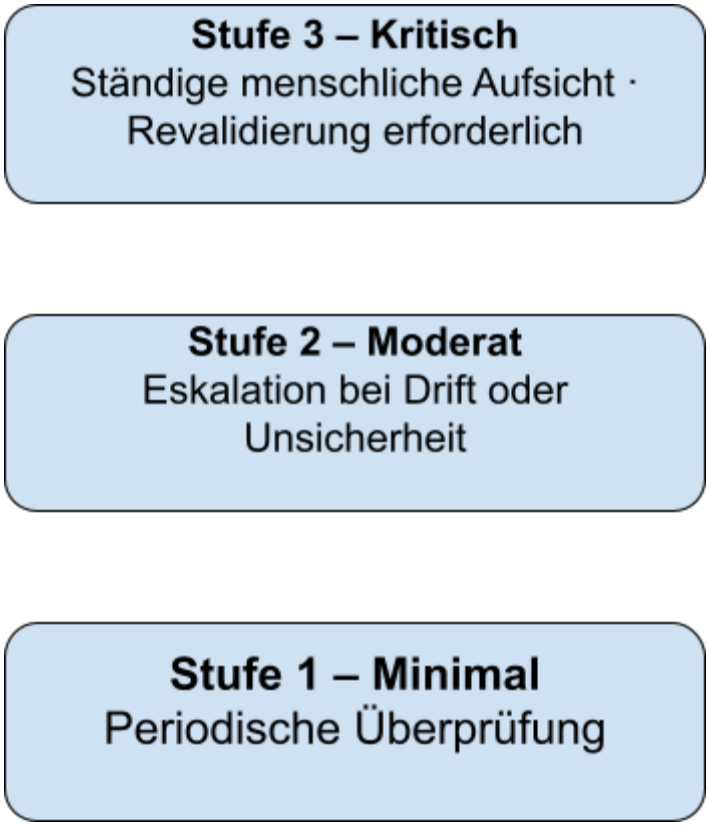

Abbildung 4 — Risikostufenklassifikation (illustrativ)

Indikative Aufsichtsstufen basierend auf potenziellen Auswirkungen.
Die Risikostufen spiegeln nicht die Leistungsfähigkeit des Systems wider und implizieren keine Zertifizierung, Genehmigung oder regulatorische Klassifizierung.

DaVinciA⁺ Governance RACI-Matrix

Diese Matrix definiert die Rollen, die erforderlich sind, um die Governance von KI-Systemen unter DaVinciA⁺ umzusetzen, zu überwachen und aufrechtzuerhalten. Sie unterscheidet zwischen Rollen, die für die Ausführung verantwortlich sind, für Ergebnisse rechenschaftspflichtig sind, bei Entscheidungen konsultiert werden oder fortlaufend informiert werden.

Rolle \ Verantwortung
Zweck definieren
Aufsichtslogik genehmigen
Eskalationsmaßnahmen

Rolle	Zweck definieren	Aufsichtslogik genehmigen	Eskalationsmaßnahmen	Leistung überwachen	Interventionen ausführen
System Owner	A	C	C	I	A
QA-/Regulierungsleitung	C	A	C	A	C
Operativer Prüfer	I	I	R	R	C
Risikoverantwortlicher	C	C	A	C	I
KI-Architekt	R	R	I	C	C
Entwickler	I	I	I	C	R

Legende (RACI)

R – Verantwortlich
Die primär ausführende Rolle. Diese Rolle führt die Aufgabe oder Aktivität aus.

A – Rechenschaftspflichtig
Der finale Eigentümer. Diese Rolle gibt frei und ist für das Ergebnis verantwortlich.

C – Konsultiert
Muss vor Handlung oder Entscheidung konsultiert werden; liefert Input oder fachliche Bewertung.

I – Informiert
Muss informiert werden, nimmt jedoch nicht an Entscheidung oder Ausführung teil.

Governance-Logik hinter der Matrix

Der System Owner trägt die Rechenschaftspflicht sowohl für die Definition des Zwecks als auch für die Durchführung von Interventionen auf Geschäftsebene (letztlich verantwortlich).

Die QA-/Regulierungsleitung ist rechenschaftspflichtig für die Aufsichtslogik und die Angemessenheit der Überwachung, was regulatorischen Erwartungen und Auditprüfungen entspricht.

Der operative Prüfer ist verantwortlich für Eskalationsmaßnahmen in Echtzeit und für die Überwachung der Leistung – er bildet die erste Linie des gesteuerten Betriebs.

Der Risikoverantwortliche ist aus Risikosicht für Eskalationsentscheidungen rechenschaftspflichtig (Stoppen, Akzeptieren oder Mindern), jedoch nicht für die tägliche Überwachung.

Der KI-Architekt ist verantwortlich für die Übersetzung des Zwecks in technisches Design (Zweck + Aufsichtslogik) und wird bei Leistung und Interventionen konsultiert.

Der Entwickler ist verantwortlich für die tatsächliche Umsetzung von Interventionen (Bereitstellung von Änderungen, Hotfixes, Rollbacks), sobald Entscheidungen getroffen wurden.

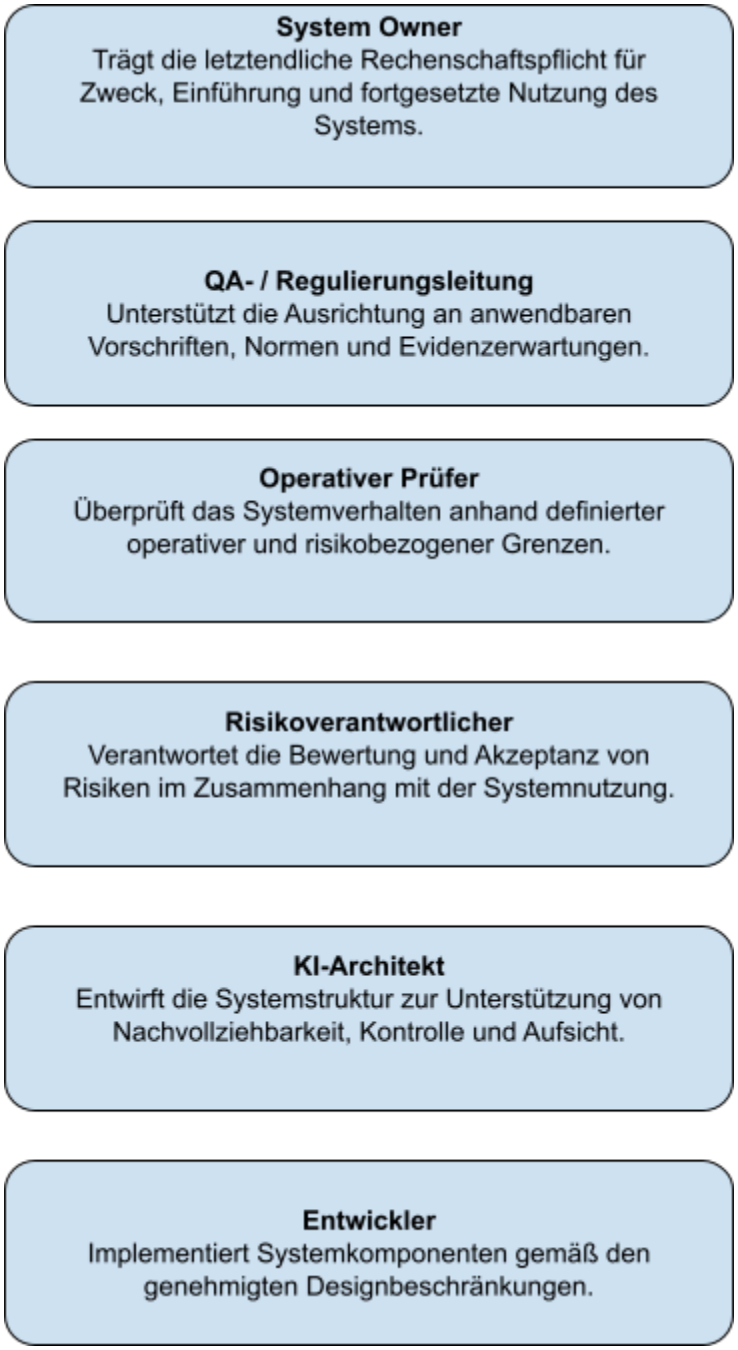

Abbildung 5 — Governance-Rollenstruktur
Illustrative Zuordnung menschlicher Autorität und Rechenschaftspflicht unter DaVinciA⁺.
Die Rollen stellen Governance-Verantwortung dar, nicht Aufgabenabfolge oder operativen Workflow.

Kapitel 7 – Compliance-Ausrichtung

Regulatorische Compliance wird durch Evidenz nachgewiesen, nicht durch Behauptung. DaVinciA⁺ beansprucht keine Konformität mit rechtlichen oder normativen Rahmenwerken. Es definiert die operativen Strukturen, durch die eine Ausrichtung an regulatorischen Erwartungen geprüft, bewertet und über die Zeit aufrechterhalten werden kann. Da sich globale Rahmenwerke weiterentwickeln – insbesondere das EU-KI-Gesetz, ISO 42001 sowie etablierte Standards für medizinische, pharmazeutische und sicherheitskritische Technologien – benötigen Organisationen eine Möglichkeit, diese Verpflichtungen in operative Begriffe zu übersetzen. DaVinciA⁺ ersetzt diese Anforderungen nicht und fungiert auch nicht als Zertifizierungsschema. Stattdessen stellt es die strukturelle Disziplin bereit, durch die Compliance unterstützt, überprüft und über den Lebenszyklus des Systems hinweg aufrechterhalten werden kann.

Die zentrale Herausforderung für Organisationen besteht darin, dass regulatorische Anforderungen in der Regel prinzipienbasiert und nicht vorschreibend sind. Sie definieren Ergebnisse – wie Transparenz, Risikomanagement, Daten-Governance und menschliche Aufsicht – ohne festzulegen, wie diese Ergebnisse technisch umzusetzen sind. DaVinciA⁺ begegnet dieser Herausforderung, indem es diese Erwartungen in die zuvor beschriebenen architektonischen und lebenszyklusbezogenen Praktiken einbettet. Der Schwerpunkt des Rahmens auf definiertem Zweck, kontrolliertem Schlussfolgern, strukturierter Aufsicht und umfassender Auditprotokollierung ermöglicht es Organisationen, die Arten von Artefakten und Evidenz zu erzeugen, die Regulierungsbehörden routinemäßig erwarten. Regulatorische Bereitschaft wird damit zu einem natürlichen Nebenprodukt disziplinierter Systemgestaltung und nicht zu einem nachträglichen Versuch, Entscheidungen ex post zu rechtfertigen.

Viele der im EU-KI-Gesetz enthaltenen Themen spiegeln strukturelle Prioritäten wider, die auch durch DaVinciA⁺ adressiert werden.

Beispielhafte strukturelle Entsprechungen sind:

Identität & Zweck → spiegelt Governance-Definitionen wider, wie sie in ISO 42001 formuliert sind

Auditprotokollierung → entspricht operativ den im EU-KI-Gesetz formulierten Erwartungen an technische Dokumentation

Drift-Überwachung → spiegelt die in regulatorischen Leitlinien beschriebenen Anforderungen an die Überwachung nach dem Inverkehrbringen wider

Änderungssteuerung → spiegelt Lebenszyklusmanagement-Prinzipien wider, wie sie in GAMP 5 und in den Lebenszyklusanforderungen von ISO 13485 beschrieben sind

Der Fokus des Gesetzes auf Datenqualität, technische Dokumentation, Risikobeobachtung, menschliche Kontrolle, Transparenz und Überwachung nach dem Inverkehrbringen entspricht den im Rahmen verankerten Lebenszykluspraktiken. Ebenso legt ISO 42001 den Schwerpunkt auf Governance-Strukturen, Verantwortlichkeiten und Managementsysteme, die ermöglicht, dass KI sicher und verantwortungsvoll betrieben wird. DaVinciA⁺ unterstützt diese Erwartungen, indem es rechenschaftspflichtige Rollen definiert, Systemgrenzen dokumentiert und eine kontinuierliche Überwachung von Verhalten und Leistung betont. Auch wenn keine Konformität behauptet wird, stellt der Rahmen eine strukturierte Grundlage bereit, anhand derer regulatorische Erwartungen geprüft werden können.

Ähnliche Parallelen lassen sich in stark regulierten Branchen beobachten. Medizinproduktestandards wie ISO 13485, ISO 14971 und IEC 62304 fordern kontrollierte Entwicklungsprozesse, risikobasierte Entscheidungsfindung und dokumentierte Validierungsevidenz. GAMP 5, seit Langem auf Software in regulierten Umgebungen angewendet, betont Lebenszyklusmanagement, Nachvollziehbarkeit und dokumentierte Begründung. DaVinciA⁺ verstärkt diese Prinzipien, ohne zu versuchen, sie zu replizieren oder zu ersetzen. Durch die Strukturierung von KI-Systemen in definierte Ebenen, die Einbettung von Validierungskontrollpunkten und die Sicherstellung der Auditierbarkeit von Entscheidungen stellt der Rahmen die operative Disziplin bereit, die Organisationen bei der Vorbereitung technischer Dokumentation, regulatorischer Einreichungen oder Qualitätsaudits benötigen.

Die Ausrichtung geht über formale Gesetzgebung und Standards hinaus. Interne Governance-Gremien, unternehmensweite Risikofunktionen, klinische Aufsichtsorgane und Audit-Teams benötigen gleichermaßen Transparenz darüber, wie KI-Systeme sich verhalten. Sie sollen die Begründung hinter Entscheidungen nachvollziehen können, bewerten, ob das System innerhalb von Richtliniengrenzen operiert hat, und feststellen, ob Risiken erkannt und angemessen eskaliert wurden. DaVinciA⁺ erleichtert dies, indem es einen transparenten operativen Nachweis erzeugt. Auditprotokolle, Konfigurationshistorien, Leistungsberichte und Änderungsnachweise werden im normalen Betrieb generiert und liefern internen Stakeholdern die Evidenz, die sie für fundierte Entscheidungen benötigen.

Compliance ist in der Praxis selten statisch. Mit der Weiterentwicklung regulatorischer Erwartungen sollen sich auch Systeme und ihre unterstützenden Prozesse weiterentwickeln. DaVinciA⁺ ist so konzipiert, dass es sich anpassen lässt, ohne Stabilität zu untergraben oder Risiken zu erhöhen. Seine Struktur ermöglicht es Organisationen, neue Anforderungen methodisch zu integrieren, indem Aufsichtsregeln aktualisiert, neue Verhaltensweisen validiert, zusätzliche Dokumentation eingeführt oder Eskalationskriterien angepasst werden. Architektur und Lebenszyklus sind hinreichend flexibel, um regulatorische Veränderungen aufzunehmen und gleichzeitig Vorhersehbarkeit und Kontrolle zu wahren. Diese Anpassungsfähigkeit ist besonders wichtig in Rechtsräumen, in denen KI-spezifische Regulierung rasch entsteht und sich Durchsetzungserwartungen im Zeitverlauf entwickeln können.

Indem Compliance als operative Eigenschaft und nicht als deklarative Aussage behandelt wird, unterstützt DaVinciA⁺ Organisationen bei der Vorbereitung auf eine Zukunft, in der Transparenz und Verantwortlichkeit grundlegende Anforderungen an den KI-Einsatz sein werden. Der Rahmen unterstützt nicht nur die Ausrichtung an heutigen Standards, sondern antizipiert die Governance-Erwartungen des kommenden Jahrzehnts. Er positioniert Organisationen so, dass sie Audits, Anfragen und Bewertungen mit Zuversicht begegnen können, und bietet eine disziplinierte Grundlage, von der aus sichere, verantwortungsvolle und nachvollziehbare KI skaliert werden kann.

Kapitel 8 – Bereitstellungs- und Einführungsmodelle

Die Einführung eines KI-Governance-Rahmens in einer Organisation erfordert mehr als eine technische Integration. Sie betont einen pragmatischen Ansatz, der bestehende Prozesse respektiert, operative Rahmenbedingungen berücksichtigt und die schrittweise Reifung von Fähigkeiten unterstützt. DaVinciA⁺ stellt eine Struktur bereit, die flexibel genug ist, um Organisationen in unterschiedlichen Phasen ihrer

KI-Reise zu unterstützen – von frühen Experimenten bis hin zu großskaligen, regulierten Implementierungen. Die Einführungsmodelle sind darauf ausgelegt, sich in etablierte Praktiken zu integrieren, anstatt diese zu ersetzen, und ermöglichen es Organisationen, Governance zu stärken, ohne laufende Arbeit zu beeinträchtigen.

Auf der einführenden Ebene beginnen Organisationen häufig mit fokussierten Anwendungsfällen, bei denen die Risiken begrenzt sind und das operative Umfeld gut verstanden wird. In diesen Kontexten bietet DaVinciA⁺ Light einen vereinfachten Pfad auf der Grundlage der Kernprinzipien Identität, Nachvollziehbarkeit, Aufsicht und kontrollierte Änderung. Es stellt eine strukturierte Methode bereit, um sicherzustellen, dass selbst frühe Prototypen oder Pilotimplementierungen die für interne Prüfungen erforderliche Dokumentation und Evidenz erzeugen. Dieses leichtgewichtige Modell ist bewusst konservativ: Sein Zweck ist es, Disziplin vor Skalierung zu etablieren und zu zeigen, dass Governance angewendet werden kann, ohne Innovation zu behindern.

Mit zunehmender Reife der Systeme und tieferer Integration in operative Workflows steigen die Anforderungen an die Governance. KI-Komponenten können beginnen, regulierte Tätigkeiten, sicherheitskritische Entscheidungen oder kundennahe Interaktionen zu beeinflussen.

Multi-Agenten-Systeme können eingeführt werden, um Aufgaben zu koordinieren oder komplexe Prozesse zu automatisieren. In dieser Phase wechseln Organisationen typischerweise zu DaVinciA⁺ Enterprise, das den vollständigen Lebenszyklus, die Architektur- und Aufsichtsstrukturen umfasst, die zuvor beschrieben wurden. Dieses Modell stellt umfassende Dokumentation, Validierungsevidenz, Audit-Trails und Änderungsmanagementverfahren bereit, die für interne Audits und externe regulatorische Prüfungen geeignet sind. Der Übergang erfolgt nicht abrupt, sondern spiegelt eine natürliche Entwicklung wider, da die Abhängigkeit der Organisation von KI zunimmt.

Die Wahl zwischen cloudbasierten, On-Premise- oder hybriden Bereitstellungsmodellen hat keinen wesentlichen Einfluss auf die Governance-Prinzipien des Rahmens. DaVinciA⁺ ist so konzipiert, dass es unabhängig von spezifischen Plattformen oder Orchestrierungswerkzeugen funktioniert.

Cloud-Umgebungen können Effizienz und Skalierbarkeit bieten, während On-Premise-Implementierungen aus Gründen des Datenschutzes, der Regulierung oder der Sicherheit bevorzugt werden können. Hybride Modelle ermöglichen es Organisationen, sensible Komponenten intern zu halten und externe Infrastruktur für weniger kritische Aufgaben zu nutzen. In allen Szenarien bleibt Governance die steuernde Ebene: Zweck, Grenzen, Schlussfolgerung und Aufsichtspflichten des Systems ändern sich nicht mit dem technischen Substrat.

Die Einführung erfordert zudem Klarheit über Rollen. Eine erfolgreiche Bereitstellung hängt von der Zusammenarbeit zwischen technischen Teams, Qualitäts- und Regulierungsfunktionen, Risikomanagement und operativer Führung ab. DaVinciA⁺ führt ein Governance-Modell ein, das festlegt, wer für die Definition des Systemzwecks, die Validierung des Verhaltens, die Überwachung der Leistung und die Aufsicht über Änderungen verantwortlich ist. Diese Verantwortlichkeiten fügen sich natürlich in bestehende Organisationsstrukturen ein und ermöglichen eine Einführung ohne umfangreiche Reorganisation. Durch frühzeitige Klärung der Erwartungen können Organisationen spätere Unsicherheiten vermeiden, insbesondere wenn Systeme beginnen, compliance-relevante Entscheidungen zu beeinflussen.

Die Skalierung des Rahmens über mehrere Systeme oder Abteilungen hinweg erfordert ein maßvolles Vorgehen. DaVinciA⁺ betont die schrittweise Ausweitung von Governance-Praktiken, unterstützt durch Vorlagen, wiederholbare Verfahren und konsistente Dokumentation. Organisationen können damit beginnen, den Rahmen auf einen einzelnen Anwendungsfall anzuwenden, und ihn anschließend auf weitere Bereiche ausdehnen, sobald der Nutzen nachgewiesen ist. Im Laufe der Zeit entsteht so ein kohärentes Governance-Umfeld, in dem alle KI-Systeme vergleichbar dokumentiert sind, gemeinsame

Aufsichtsmechanismen nutzen und Evidenz erzeugen, die zu einem einheitlichen Risiko- und Leistungsbild aggregiert werden kann.
Wichtig ist, dass die Bereitstellung nicht ausschließlich aus der Perspektive der Compliance betrachtet werden sollte. Unternehmen, die DaVinciA⁺ einführen, berichten typischerweise von geringerer Unklarheit in der Entwicklung, schnelleren internen Freigaben und verbesserter Audit-Bereitschaft. Diese operativen Vorteile werden umso deutlicher, je stärker KI-Systeme funktionsübergreifend skaliert werden.
Organisationen, die DaVinciA⁺ einsetzen, stellen häufig fest, dass die bereitgestellte Struktur die operative Zuverlässigkeit erhöht und Unsicherheit in der Entwicklung reduziert. Klare Grenzen verringern Nacharbeiten aufgrund widersprüchlicher Erwartungen. Explizite Aufsicht stärkt das Vertrauen in Entscheidungen. Umfassende Dokumentation erleichtert die Zusammenarbeit zwischen Teams. Diese Vorteile sind in regulierten Branchen besonders sichtbar, erstrecken sich jedoch auf jeden Bereich, in dem KI Entscheidungen beeinflusst, die von Bedeutung sind.
DaVinciA⁺ fungiert somit sowohl als Governance-Rahmen als auch als operativer Enabler. Es bietet Organisationen einen Weg zur verantwortungsvollen Einführung, ohne Momentum einzubüßen. Durch die Bereitstellung stabiler Strukturen, die skalieren, sich anpassen und Prüfungen standhalten können, unterstützt der Rahmen sowohl Innovation als auch Verantwortlichkeit. Seine Bereitstellungsmodelle spiegeln ein praxisnahes Verständnis organisatorischer Realitäten wider und stellen sicher, dass sich Governance parallel zu den Fähigkeiten und Verantwortlichkeiten der unterstützten Systeme entwickelt.

DaVinciA⁺ Reifegradmodell

Organisationen, die DaVinciA⁺ einführen, durchlaufen definierte Stufen der Governance-Fähigkeit. Diese Ebenen spiegeln zunehmende strukturelle Disziplin, Tiefe der Aufsicht und Audit-Bereitschaft wider. Das nachstehende Reifegradmodell stellt eine Referenzstruktur dar, um die Einführung zu steuern, Fortschritte zu überprüfen und Implementierungen der nächsten Phase zu planen.

Ebene
Beschreibung
Operative Indikatoren

Ebene 1 — Pilot
Erstimplementierung mit grundlegenden Governance-Elementen
• Identität & Zweck definiert
• Grundlegende Auditprotokollierung
• Manuelle Aufsichtskontrollpunkte

Ebene 2 — Strukturiert
Multi-Agenten- und OQ-fähige Systeme
• RACI-Rollen zugewiesen
• Eskalationslogik formalisiert
• Drift-Überwachung aktiviert

Ebene 3 — Enterprise
Implementierung auf reguliertem Niveau mit Lebenszyklusaufsicht
• PQ-Tests abgeschlossen

• Change-Control-Board operativ
• Monatliche Governance-Reviews dokumentiert

Ebene 4 — Audit-bereit
Vollständig ausgereifte Systeme mit vollständiger Nachvollziehbarkeit
• Mindest-Evidenzpaket erstellt
• Externe Audit-Bereitschaft bestätigt
• Konformitätsevidenz (nicht behauptend) verfügbar

Jede Ebene baut auf der vorherigen auf und erhöht Vertrauen, Verteidigungsfähigkeit und Kontrolle. Dieses Modell schreibt weder Geschwindigkeit noch Zeitplan vor, sondern bietet einen strukturierten Pfad, entlang dessen DaVinciA⁺-Governance im Einklang mit Systemkritikalität und regulatorischem Kontext skaliert werden kann.

Kapitel 9 – Fallstudien

Fallstudien sind enthalten, um zu veranschaulichen, wie DaVinciA⁺ in praktischen Umgebungen angewendet werden kann, um Struktur, Nachvollziehbarkeit und Aufsicht in KI-gestützten Systemen zu etablieren. Sie sind nicht dazu bestimmt, Leistungsfähigkeit, Sicherheitseigenschaften, regulatorische Konformität oder Zertifizierungsreife nachzuweisen. Stattdessen liefern sie Beispiele dafür, wie Organisationen den Rahmen eingesetzt haben, um komplexe Schlussfolgerungen zu organisieren, Grenzen zu formalisieren und Lebenszyklus-Governance einzuführen. Die folgenden Fälle spiegeln zwei Projekte in nicht miteinander verbundenen Domänen wider, jeweils in unterschiedlichen Reifegraden, in denen DaVinciA⁺ zur Stärkung der Governance-Disziplin eingesetzt wurde.

Fallstudie 1 — Wissensintensives Expertensystem in einem nicht regulierten Bereich
Eine Organisation, die ein spezialisiertes instruktionales KI-System entwickelte, verfolgte das Ziel, einen umfangreichen Bestand an Expertenwissen in eine konsistente, interpretierbare und auditierbare Mehrkomponenten-Architektur zu überführen. Vor der Einführung von DaVinciA⁺ bestand das Systemdesign aus lose definierten konzeptionellen Modulen, denen dokumentierte Grenzen, Interaktionsregeln oder Aufsichtserwartungen fehlten. Dies führte zu Unklarheiten hinsichtlich des Systemverhaltens und schränkte die Möglichkeit ein, die Lösung verantwortungsvoll zu skalieren. DaVinciA⁺ wurde eingeführt, um eine strukturierte Grundlage bereitzustellen.
Zentrale Aktivitäten umfassten:
Formalisierung von Identität und Zweck, um den Zweck, die Beschränkungen und die Nicht-Ziele des Systems klar zu definieren.

Etablierung von agentenspezifischen Grenzen, um sicherzustellen, dass jede Schlussfolgerungskomponente innerhalb genehmigter Verantwortlichkeiten operierte.

Dokumentation von Wissensquellen und Schlussfolgerungslogik, um transparente Überprüfung und Versionskontrolle zu ermöglichen.

Implementierung von Aufsichtsregeln und kontrollierten Delegationspfaden, sodass Interaktionen zwischen mehreren Komponenten überwacht und rekonstruiert werden konnten.

Ermöglichung einer nachvollziehbaren Weiterentwicklung, sodass spätere Erweiterungen die strukturelle Integrität des Systems nicht beeinträchtigten.

Obwohl diese Umgebung keiner regulatorischen Aufsicht unterlag, ermöglichte die Einführung von DaVinciA⁺ der Organisation, das System von einem informellen Prototyp zu einer stabilen, steuerbaren Struktur weiterzuentwickeln. Der Rahmen schuf Klarheit, Auditierbarkeit und kontrolliertes Wachstum, ohne Innovation einzuschränken.

Fallstudie 2 — Compliance-relevantes Entscheidungsunterstützungssystem in einem regulierten Kontext
Eine separate Organisation, die ein KI-gestütztes Entscheidungsunterstützungswerkzeug für compliance-sensible Workflows entwickelte, benötigte ein Governance-Modell, das eine zukünftige regulatorische Prüfung unterstützen konnte. Das System sollte domänenspezifische Regeln aufnehmen, strukturierte und unstrukturierte Informationen interpretieren und menschliche Prüfer in urteilsbasierten Prozessen unterstützen. Von Beginn an erkannte die Organisation, dass Lebenszyklus-Governance, Nachvollziehbarkeit und menschliche Aufsicht entscheidend sein würden, um einen verantwortungsvollen Betrieb nachzuweisen.
DaVinciA⁺ wurde als internes Governance-Rahmenwerk ausgewählt.
Es wurde angewendet, um:
den vorgesehenen Verwendungszweck, die architektonischen Ebenen und die operativen Grenzen des Systems zu definieren.

kontrollierte Schlussfolgerungsprozesse mit versionierter Logik, genehmigten Werkzeugen und dokumentierten Datenquellen einzuführen.

den Validierungslebenszyklus (IQ, OQ, PQ) zu planen und zu dokumentieren, um strukturelle, verhaltensbezogene und operative Zweckmäßigkeit sicherzustellen.

eine umfassende Auditprotokollierung zu implementieren, die die Rekonstruktion von Entscheidungen, Eskalationsauslösern und Aufsichtsinterventionen ermöglicht.

Mechanismen zur Änderungssteuerung zu etablieren, sodass Aktualisierungen von Modellen, Werkzeugen oder Workflows unter dokumentierter Überprüfung erfolgten.

Durch die Verankerung des Systems in DaVinciA⁺ baute die Organisation eine starke Evidenzgrundlage auf, lange bevor regulatorische Einreichungen oder externe Auditaktivitäten absehbar waren. Der Rahmen stellte sicher, dass die Weiterentwicklung des Systems transparent und kontrollierbar blieb und dass menschliche Aufsicht konsequent in compliance-relevante Entscheidungen eingebettet war.

Zusammenfassung der Erkenntnisse aus den Fallstudien
Über beide Beispiele hinweg – eines nicht reguliert und explorativ, das andere reguliert und compliance-relevant – zeigte sich dasselbe Muster:
Zweck und Grenzen wurden explizit statt implizit.

Schlussfolgerungsprozesse wurden überprüfbar und gesteuert statt intransparent.

Aufsicht wurde strukturiert, wodurch vorhersehbare Human-in-the-Loop-Interventionen ermöglicht wurden.

Nachvollziehbarkeit wurde inhärent, wodurch sowohl interne Absicherung als auch externe Audit-Bereitschaft unterstützt wurden.

Das Systemwachstum blieb kontrolliert und verhinderte unbeabsichtigte Drift im Umfang oder Verhalten.

Diese Beispiele zeigen, wie DaVinciA⁺ die KI-Entwicklung in Governance-Prinzipien verankern kann, ohne Leistungs-, Sicherheits- oder Konformitätsansprüche zu erheben. Ihr Zweck ist illustrativ: zu zeigen, wie der Rahmen Klarheit, Verantwortlichkeit und disziplinierte Weiterentwicklung in unterschiedlichen KI-Umgebungen unterstützen kann.

Kapitel 10 – Technischer Anhang

Daten-Governance-, Datenschutz- und Cybersicherheits-Geltungsbereich
DaVinciA⁺ geht davon aus, dass grundlegende Daten-Governance-Kontrollen – einschließlich Datenschutzrichtlinien, Zugriffsmanagement, Datenminimierung, Aufbewahrung und Cybersicherheitsmaßnahmen – auf der Ebene der Infrastruktur, der Plattform oder des Qualitätsmanagementsystems (QMS) implementiert und betrieben werden. Diese Kontrollen gelten als Voraussetzungen und nicht als Bestandteile des DaVinciA⁺-Rahmens selbst.
Zukünftige Erweiterungen des Rahmens werden Referenzabbildungen zu etablierten Normen und Leitlinien bereitstellen, einschließlich ISO/IEC 27701, DSGVO und NIST SP 800-53, um Organisationen zu unterstützen, die KI-Governance mit umfassenderen Datenschutz- und Sicherheitskontrollumgebungen integrieren möchten. Diese Abbildungen bleiben nicht-normativ und implementierungsneutral.
Der Zweck des technischen Anhangs in einem öffentlichen Whitepaper besteht nicht darin, operative Details bereitzustellen, sondern den Lesern ein klareres Verständnis der Arten von Artefakten und Evidenz zu vermitteln, die ein gesteuertes KI-System unterstützen. In regulierten und unternehmensweiten Umgebungen benötigen Stakeholder häufig Einblick in die Strukturen, die Aufsicht, Auditierbarkeit und Lebenszyklusmanagement ermöglichen. DaVinciA⁺ stellt diese Strukturen durch eine Reihe konzeptioneller Elemente bereit, die das Systemverhalten untermauern, ohne proprietäre Logik oder interne Implementierungsdetails offenzulegen. Der Anhang fasst diese Elemente zusammen, um zu veranschaulichen, wie technische Transparenz in der Praxis erreicht wird.
Ein zentrales Element des Rahmens ist der Auditdatensatz. KI-Systeme erzeugen eine Abfolge von Entscheidungen, Werkzeugaufrufen, Schlussfolgerungspfaden und kontextuellen Interpretationen, die in einer dauerhaften und überprüfbaren Form erfasst werden sollen. DaVinciA⁺ behandelt Auditprotokollierung als kontinuierliche Aktivität und nicht als optionales Diagnosemerkmal. Jeder Systemlauf erzeugt einen strukturierten Datensatz, der es Ermittlern, Auditoren und Aufsichtsteams

ermöglicht, Ereignisse mit Klarheit zu rekonstruieren. Diese Datensätze spiegeln typischerweise den erklärten Zweck des Systems, die empfangenen Eingaben, die Grenzen, innerhalb derer es operierte, sowie die als Reaktion ergriffenen Maßnahmen wider. Während das spezifische Format dieser Datensätze je nach Organisation und Plattform variiert, bleibt die zugrunde liegende Erwartung gleich: Transparenz muss auf grundlegender Ebene in das System eingebettet sein.

Eng damit verbunden ist das Konzept der Metadaten. KI-Systeme sind von zahlreichen kontextuellen Variablen abhängig – Modellversionen, Konfigurationseinstellungen, Datensatzkennungen, Entscheidungsschwellen und Umgebungsbedingungen –, die das Verhalten beeinflussen. Ohne präzise Metadaten können selbst geringfügige Änderungen Unsicherheit darüber erzeugen, wie oder warum ein System zu einem bestimmten Ergebnis gelangt ist. DaVinciA⁺ betont, dass Metadaten systematisch erfasst und als Teil des Audit-Trails aufbewahrt werden. Dieser Ansatz ermöglicht, dass Änderungen nachvollzogen, Verhalten korrekt interpretiert und Evidenz über den gesamten Lebenszyklus des Systems hinweg kohärent bleibt. Metadaten fungieren als verbindendes Gewebe zwischen Konfiguration, Schlussfolgerung und Aufsicht. Metadatenkategorien umfassen typischerweise:

– Modell-Metadaten (Versionen, Parameter, Anbieter)
– Konfigurations-Metadaten (Werkzeugberechtigungen, Umgebungseinstellungen)
– Entscheidungs-Metadaten (Schlussfolgerungspfad, Aktivierung von Leitplanken)
– Aufsichts-Metadaten (Prüfer, Eskalationsgründe, Ergebnisse)

Diese Kategorien schaffen Konsistenz, ohne proprietäre Interna offenzulegen.

Ein weiteres wichtiges Element betrifft die formale Beschreibung von Systemgrenzen. DaVinciA⁺ betont nicht, dass Organisationen ihre interne Logik veröffentlichen, fördert jedoch eine klare Artikulation von Umfang, Beschränkungen und autorisierten Funktionen. Diese Beschreibungen helfen Stakeholdern, den vorgesehenen Einsatz des Systems zu verstehen und zu bewerten, ob sein Verhalten mit diesem Zweck konsistent bleibt. In Multi-Agenten-Umgebungen erstrecken sich diese Grenzen auch auf die Beziehungen zwischen Agenten und legen fest, welche Interaktionen zulässig sind, wie Delegation erfolgt und wo menschliche Aufsicht eingreifen muss. Auch wenn diese Beschreibungen implementierungsspezifisch sind, stellt der Rahmen sicher, dass sie konsistent und überprüfbar erfasst werden.

Die Änderungssteuerung gehört ebenfalls zum technischen Ökosystem, das Governance unterstützt. KI-Systeme entwickeln sich durch Aktualisierungen von Modellen, Werkzeugen, Datensätzen und operativen Regeln weiter. DaVinciA⁺ strukturiert diese Änderungen durch formale Überprüfungsprozesse, die potenzielle Auswirkungen bewerten und festlegen, ob eine Revalidierung erforderlich ist. Der Anhang definiert keine spezifischen Workflows vor, hebt jedoch die Bedeutung hervor, die Begründung jeder Änderung, die sie stützende Evidenz sowie die damit verbundenen Aufsichtsentscheidungen zu dokumentieren. Diese Disziplin ermöglicht, dass die Weiterentwicklung des Systems bewusst und nachvollziehbar erfolgt und nicht inkrementell und ungeprüft.

Abschließend erkennt der Anhang die Test- und Überwachungsmechanismen an, die eine verantwortungsvolle KI-Bereitstellung begleiten. Organisationen können eine Vielzahl von Techniken einsetzen – Verifikationstests, Verhaltensbewertungen, Drift-Überwachung und periodische Evaluierungen –, um sicherzustellen, dass das System weiterhin innerhalb der erwarteten Parameter operiert. DaVinciA⁺ liefert die konzeptionelle Grundlage für diese Aktivitäten, indem es definiert, was beobachtet werden muss, was aufgezeichnet werden muss und wie Entscheidungen über das Systemverhalten zu treffen sind. Die konkreten Ausgestaltungen jeder Methode hängen von der technischen Umgebung, dem regulatorischen Kontext und den operativen Anforderungen der Organisation ab.

In ihrer Gesamtheit veranschaulichen diese Elemente die unterstützende Infrastruktur, die für

rechenschaftspflichtige KI erforderlich ist. Der technische Anhang versucht nicht, Implementierungen in vorschreibender Detailtiefe zu beschreiben; vielmehr vermittelt er ein kohärentes Bild der Artefakte und Prozesse, die Transparenz, Aufsicht und Lebenszyklus-Governance ermöglichen. Er bekräftigt das übergeordnete Ziel des Rahmens: sicherzustellen, dass KI-Systeme während ihrer Entwicklung, Bereitstellung und Weiterentwicklung verständlich und kontrollierbar bleiben.

„Illustrative Aufsichtsstruktur, die die Governance-Prinzipien von DaVinciA⁺ widerspiegelt. Keine erforderliche Systemkonfiguration."

Hinweis:
„Die Eskalationsschwellen-Matrix ist illustrativ und nicht normativ. Organisationen sollen Schwellenwerte entsprechend ihrem Risikomanagementprozess konfigurieren."

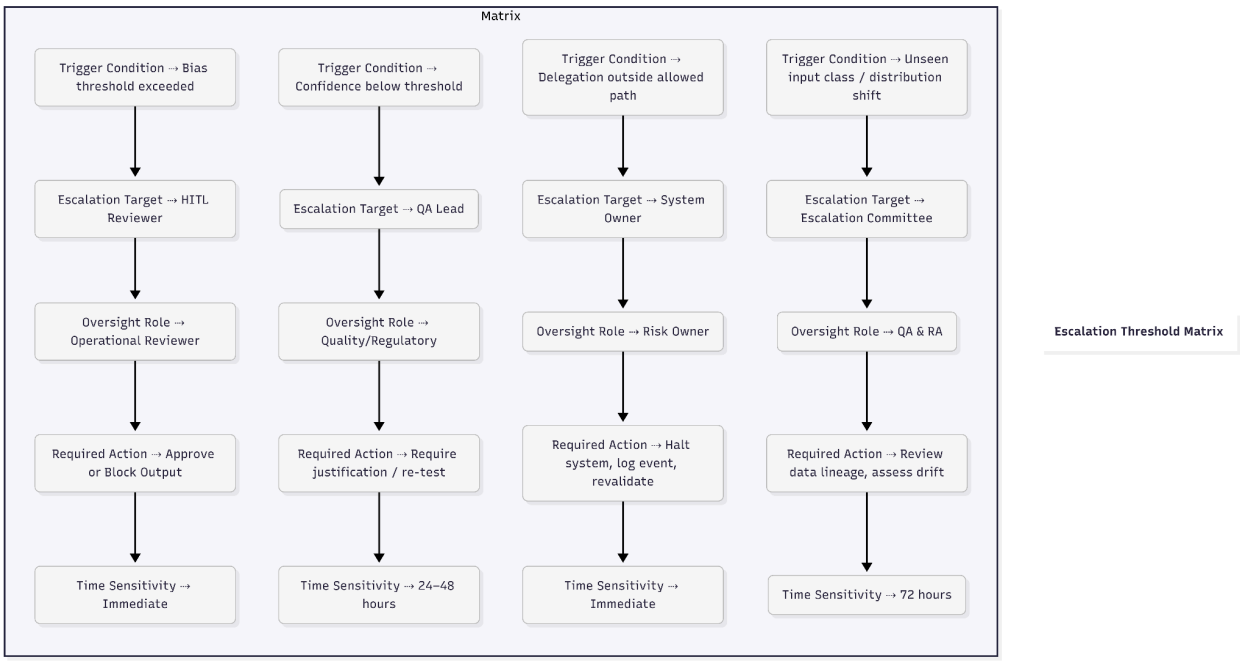

Abbildung 6 — Eskalationsmatrix (illustrativ)
Beispielhafte Zuordnung von Eskalationsbedingungen zu menschlichen Aufsichtsrollen. Diese Abbildung ist nicht normativ und definiert weder automatisiertes Verhalten noch eine erforderliche Systemkonfiguration.

Auditprotokollierungs- und Nachvollziehbarkeitsinfrastruktur
„Illustrativer Audit-Trail-Auszug, der eine mehrstufige Interaktion, eine Leitplankenblockierung und einen Eskalationsauslöser erfasst. Format entspricht dem DaVinciA⁺ Metadatenschema (DMS-AUD-070)."

{
"run_id": "RUN-2025-0415-0371",
"timestamp_utc": "2025-04-15T09:36:18Z",
"agent_id": "AGENT-DECISION-1A",
"user_request": "Generate draft response for regulatory comment letter",
"input_context": {
"data_sources": ["doc://eu-ai-act-v3.4", "doc://client-guidance-notes"],
"risk_tier": "Tier-3",
"governance_mode": "High Oversight"
},
"steps": [
{
"step_id": "STEP-001",
"timestamp_utc": "2025-04-15T09:36:19Z",
"tool_invoked": "summarisation.agent",
"input_summary": "Parse key terms from EU AI Act extract",
"reasoning_snapshot": "Extracting articles relevant to classification scope",
"guardrail_triggered": false,
"escalation_triggered": false,
"output_summary": "Identified Articles 6, 10, and 23 as relevant to request"
},
{
"step_id": "STEP-002",
"timestamp_utc": "2025-04-15T09:36:24Z",

"tool_invoked": "response-generator.model-gpt4",

"input_summary": "Build draft response using regulatory summary",

"reasoning_snapshot": "Synthesising commentary based on compliance structure",

"guardrail_triggered": true,

"guardrail_type": "LegalClaim-Restriction",

"escalation_triggered": true,

"escalation_path": "HITL_Review",

"output_summary": "[BLOCKED] Output contained unverified conformity claim. Routed to System Owner for review."

}

],

"final_output": "[Escalated to Human Reviewer]",

"reviewer_notes": "Model attempted to assert CE conformity. Blocked and returned for rewrite. Escalation logged as E-2025-0349.",

"audit_signoff": {

"reviewed_by": "QA-OVERSIGHT-22",

"review_timestamp": "2025-04-15T09:41:02Z"

}

}

Haftungsausschluss:
„Nicht-normatives Beispiel, das ein mögliches Format eines Auditdatensatzes zeigt. Organisationen können alternative Schemata implementieren, die mit ihrem QMS vereinbar sind.“

Kapitel 11 – Zusammenfassung und Glossar

Zukünftige Roadmap
DaVinciA⁺ ist darauf ausgelegt, sich schrittweise weiterzuentwickeln, während sich Governance-Erwartungen, regulatorische Umfelder und operative Praktiken weiter ausprägen. Geplante ergänzende Veröffentlichungen umfassen:
– Bedrohungsmodellierungs-Muster und Fehlermodus-Bibliotheken
– Leitlinien zur Integration von Daten-Governance und Datenschutz
– Erweiterte Konfigurationsbibliotheken für Human-in-the-Loop-Aufsicht
– Domänenspezifische Risiko-Referenzmodelle (z. B. MedTech, Finanzwesen)
Diese Materialien werden als optionale, nicht-normative Ergänzungen veröffentlicht und verändern nicht den Referenzcharakter des DaVinciA⁺-Rahmens.

Zusammenfassung
Die Entwicklung und der Einsatz von Systemen der künstlichen Intelligenz erfordern ein Maß an Struktur und Verantwortlichkeit, das der Bedeutung der Entscheidungen entspricht, die diese Systeme beeinflussen. DaVinciA⁺ bietet hierfür einen praxisnahen und disziplinierten Ansatz. Es etabliert einen klaren Rahmen, der auf definiertem Zweck, kontrolliertem Schlussfolgern und kontinuierlicher Aufsicht beruht und sicherstellt, dass KI-Systeme in allen Phasen ihres Lebenszyklus transparent, vorhersehbar und unter menschlicher Autorität bleiben.
Die Architektur des Rahmens beschreibt das System durch drei voneinander abhängige Ebenen, die Zweck klären, Verhalten begrenzen und Auditierbarkeit unterstützen. Der Validierungslebenszyklus erstreckt diese Struktur über die Einführung hinaus und betont, dass verantwortungsvoller Betrieb eine fortlaufende Überwachung und Evidenz erfordert, nicht eine einmalige Bewertung. Governance-Praktiken stellen sicher, dass Verantwortlichkeit explizit ist, dass Aufsicht in den täglichen Betrieb eingebettet ist und dass Entscheidungen rekonstruiert und geprüft werden können. Die Compliance-Ausrichtung versetzt Organisationen in die Lage, sich an wandelnde regulatorische Erwartungen anzupassen – durch nachweisbare Prozesse statt deklarativer Behauptungen. Einführungsmodelle ermöglichen die Skalierung des Rahmens über Domänen und Reifegrade hinweg, von frühen Pilotprojekten bis hin zur unternehmensweiten Implementierung. Fallstudien veranschaulichen, wie DaVinciA⁺ bereits in realen Projekten Struktur geschaffen und Klarheit, Nachvollziehbarkeit sowie kontrollierte Weiterentwicklung unterstützt hat.
In ihrer Gesamtheit bilden diese Komponenten einen kohärenten Ansatz für KI-Governance. DaVinciA⁺ schafft eine stabile Grundlage, auf der Organisationen verantwortungsvoll innovieren können, Vertrauen in ihre Systeme wahren und gleichzeitig regulatorische Veränderungen bewältigen. Es bietet ein Mittel, um sicherzustellen, dass KI ein kontrolliertes, transparentes und rechenschaftspflichtiges Instrument bleibt – fähig, komplexe Entscheidungen zu unterstützen, ohne Aufsicht oder organisatorische Integrität zu gefährden.
DaVinciA⁺ ermöglicht es Organisationen, Governance vor Skalierung, Evidenz vor Audit und Klarheit vor Komplexität zu etablieren. Durch die Operationalisierung von Governance-Prinzipien, auf die Regulierungsbehörden häufig verweisen, ermöglicht es Unternehmen, KI-Initiativen mit Zuversicht voranzutreiben und gleichzeitig kontinuierliche Verantwortlichkeit aufrechtzuerhalten.

Glossar

Accountability (Verantwortlichkeit)
Die Verpflichtung identifizierbarer menschlicher Rollen, das Verhalten und die Ergebnisse eines KI-Systems zu überwachen, zu bewerten und zu rechtfertigen.

Agent
Eine spezialisierte Komponente innerhalb eines KI-Systems, die definierte Aufgaben oder Schlussfolgerungsfunktionen unter dokumentierten Grenzen und Aufsicht ausführt.

Audit Logging (Auditprotokollierung)
Die systematische Aufzeichnung von Systemhandlungen, Schlussfolgerungsschritten und kontextuellen Informationen, um die Rekonstruktion und Überprüfung des Verhaltens zu ermöglichen.

Change Control (Änderungssteuerung)
Ein strukturierter Prozess zur Bewertung und Dokumentation von Änderungen an einem KI-System, einschließlich der Bewertung von Auswirkungen und der Anforderungen an eine Revalidierung.

Compliance Alignment (Compliance-Ausrichtung)
Die Praxis, Systeme und Prozesse so zu strukturieren, dass sie die Erwartungen regulatorischer Rahmenwerke unterstützen, ohne Konformität zu behaupten.

Configuration (Konfiguration)
Die dokumentierten technischen und operativen Einstellungen, die festlegen, wie ein KI-System instanziiert ist, einschließlich Modellversionen, Werkzeuge und Berechtigungen.

Continuous Monitoring (Kontinuierliche Überwachung)
Die fortlaufende Bewertung des Systemverhaltens zur Erkennung von Abweichungen, aufkommenden Risiken oder Leistungsänderungen, die ein Eingreifen erfordern können.

Delegation Pathway (Delegationspfad)
Eine autorisierte Interaktion, über die ein Agent innerhalb definierter Grenzen Informationen oder Unterstützung von einem anderen anfordern darf.

Drift
Eine Veränderung des Systemverhaltens oder der zugrunde liegenden Daten, die Ausgaben oder Schlussfolgerungen beeinflusst und eine Überwachung sowie potenzielle Revalidierung erfordert.

Escalation (Eskalation)
Der Prozess, durch den ein KI-System Unsicherheit, Risiko oder Grenzverletzungen identifiziert und die Entscheidungsfindung an menschliche Aufsicht übergibt.

Governance
Die Gesamtheit der Strukturen, Prozesse und Verantwortlichkeiten, die ermöglicht, dass KI-Systeme innerhalb definierter ethischer, operativer und regulatorischer Grenzen betrieben werden.

Identity and Intent (Identität und Zweck)
Eine formale Beschreibung des Zwecks, des Umfangs, der Beschränkungen und der Nicht-Ziele des Systems, die architektonische und operative Entscheidungen verankert.

Lifecycle (Lebenszyklus)
Die vollständige Abfolge von Aktivitäten zur Entwicklung, Validierung, Einführung, Überwachung und Aktualisierung eines KI-Systems.

Metadata (Metadaten)
Kontextuelle Informationen, die beschreiben, wie das System betrieben wurde, einschließlich Modellversionen, Konfigurationsdetails und Umgebungsbedingungen.

Oversight (Aufsicht)
In den Systembetrieb eingebettete menschliche Überwachung zur Bewertung von Ausgaben, zum Umgang mit Unsicherheit und zur Sicherstellung, dass Entscheidungen innerhalb autorisierter Grenzen bleiben.

Performance Qualification (PQ)
Bewertung des Verhaltens eines KI-Systems in seiner realen operativen Umgebung.

Reasoning Process (Schlussfolgerungsprozess)
Die interne Logik, durch die ein KI-System Eingaben interpretiert und Ausgaben erzeugt, einschließlich Entscheidungswegen und Werkzeugnutzung.

Risk Management (Risikomanagement)
Die Identifikation, Bewertung und Minderung potenzieller Schäden, die mit Systemverhalten oder Systemausfällen verbunden sind.

Scope (Umfang)
Der autorisierte Satz von Aufgaben, Verantwortlichkeiten und Domänen, innerhalb dessen ein KI-System operieren darf.

Traceability (Nachvollziehbarkeit)
Die Fähigkeit, Systemverhalten anhand dokumentierter Schlussfolgerungen, Auditprotokolle und kontextueller Metadaten zu rekonstruieren.

Validation (Validierung)
Die strukturierte Bewertung eines KI-Systems zur Bestätigung, dass es über seinen gesamten Lebenszyklus hinweg korrekt, sicher und innerhalb definierter Grenzen betrieben wird.

Anhang A — Mindest-Evidenzpaket für Governance-Review

Mindest-Evidenzpaket für Governance-Review
Regulatorische und unternehmensweite Governance-Bewertungen sind evidenzbasiert. DaVinciA⁺ definiert oder definiert keine spezifischen Artefakte vor; stattdessen etabliert es eine Mindest-Evidenzstruktur, anhand derer Governance, Aufsicht und Lebenszyklusdisziplin geprüft werden können. Die nachstehende Tabelle veranschaulicht einen repräsentativen Mindestsatz an Evidenz, der üblicherweise im Rahmen von Governance- oder Audit-Reviews erwartet wird.

Artefakt
Quellvorlage
Beschreibung

Identitäts- & Zweckaufzeichnung
DMS-GOV-001
Deklarierter Systemumfang, Grenzen und Nicht-Ziele

Aufsichts- & Eskalationsregeln
DMS-GOV-011
Bedingungen, unter denen Human-in-the-Loop-Aufsicht erforderlich ist

Auditprotokoll-Schema & Beispiele
DMS-AUD-070
Erfasste Läufe, Schritte, Leitplankenaktivierungen und Nachvollziehbarkeit

RACI-Matrix
DMS-GOV-010
Definierte Zuordnung von Verantwortung und Rechenschaftspflicht

Änderungssteuerungsregister
DMS-CC-061
Dokumentierte Systemänderungen und Auswirkungsbewertungen

IQ / OQ / PQ-Berichte
DMS-VAL-021 / 031 / 041
Evidenz zur Installations-, Verhaltens- und Realwelt-Validierung

Drift-Überwachungsprotokoll
DMS-MON-050
Aufzeichnungen zur statistischen und verhaltensbezogenen Drift-Erkennung

Diese Artefakte entsprechen den in dem DaVinciA⁺ Validierungs-Toolkit und dem Deployment Playbook definierten Strukturen und werden ausschließlich zu illustrativen Governance-Zwecken dargestellt.