

Holger Reibold

One Model Does Not Fit All

Warum domänenspezifische
Sprachmodelle zur strategischen
Notwendigkeit werden

BRAIN-MEDIA.DE

Holger Reibold

One size does not fit all

Warum domänenspezifische Sprachmodelle zur
strategischen Notwendigkeit werden

BRAIN-MEDIA.DE

Alle Rechte vorbehalten. Ohne ausdrückliche, schriftliche Genehmigung des Verlags ist es nicht gestattet, das Buch oder Teile daraus in irgendeiner Form durch Fotokopien oder ein anderes Verfahren zu vervielfältigen oder zu verbreiten. Dasselbe gilt auch für das Recht der öffentlichen Wiedergabe. Der Verlag macht darauf aufmerksam, dass die genannten Firmen- und Markennamen sowie Produktbezeichnungen in der Regel marken-, patent- oder warenrechtlichem Schutz unterliegen.

Verlag und Autor übernehmen keine Gewähr für die Funktionsfähigkeit beschriebener Verfahren und Standards.

© 2025 Brain-Media.de

ISBN: 978-3-95444-300-0

Cover: zincamerastock/Alamy

Brain-Media.de

Dr. Holger Reibold – Hubert-Müller-Str. 52c – 66113 Saarbrücken

info@brain-media.de – www.brain-media.de

Inhaltsverzeichnis

VORWORT	1
1 ONE MODEL DOES NOT FIT ALL	5
1.1 Wo generalistische Sprachmodelle scheitern	13
1.1.1 Terminologische Präzision als Schwachstelle	13
1.1.2 Halluzinationen sind kein Randphänomen	15
1.1.3 Fehlende Regel- und Normdurchsetzung	16
1.1.4 Kontextverlust in komplexen Dokumenten	17
1.1.5 Mangelnde Reproduzierbarkeit	17
1.1.6 Fehlkosten und Haftungsrisiken	18
1.2 Kosten, Haftung und Risiko	19
1.2.1 Direkte und indirekte Kosten	19
1.2.2 Fehlkosten und ihre Asymmetrie	21
1.2.3 Haftung und Verantwortungsdiffusion	21
1.2.4 Regulatorische Anforderungen	23
1.2.5 Risikoaggregation im Systemkontext	24
1.2.6 Ökonomische Neubewertung	24
2 WAS EINE DOMÄNE WIRKLICH AUSMACHT	27
2.1 Was ist eine Domäne?	27

2.2	Warum Transfer Learning oft nicht reicht	32
2.2.1	Transfer Learning adressiert Aufgaben	33
2.2.2	Die Grenzen der Terminologie-Adaption.....	34
2.2.3	Implizites Domänenwissen nicht nachrüstbar	35
2.2.4	Normen und Regeln sind kein Trainingssignal.....	36
2.2.5	Overfitting statt Verständnis.....	36
2.2.6	Wann Transfer Learning dennoch sinnvoll ist	37
2.3	Wann Spezialisierung zwingend ist	38
3	ENTSCHEIDUNGARCHITEKTUREN FÜR DSLMS.....	43
3.1	Die drei Zugangsmodelle zu LLMs	43
3.2	External Augmentation.....	47
3.2.1	RAG als pragmatischer Einstieg	47
3.2.2	RAG erweitert Zugriff – nicht Verständnis.....	49
3.2.3	Wissensaktualität versus Wissensverstehen	50
3.2.4	Fehlannahmen bei RAG-Architekturen	51
3.2.5	Wann RAG nicht mehr ausreicht.....	52
3.3	Prompting und Grey-Box-Strategien	53
3.3.1	Was Prompting leisten kann.....	53
3.3.2	Prompting wirkt nicht auf Wissen	54
3.3.3	Grenzen bei Terminologie und implizitem Wissen	55
3.3.4	Prompting und Regelverbindlichkeit	56
3.3.5	Wann Prompting sinnvoll bleibt.....	56
3.3.6	Prompting als Teil einer Grey-Box-Strategie	57

3.4	White Box: der Weg zu echten DSLMs	57
3.5	Build, Buy oder Adapt?	61
4	DER SCHLÜSSEL: DATEN UND GOVERNANCE	65
4.1	Daten als limitierender Faktor	65
4.2	Governance, Compliance und Ownership	69
4.3	Betrieb von DSLMs	72
5	WO DSLMS FUNKTIONIEREN – UND WARUM	77
5.1	Biomedizin und Gesundheitswesen	78
5.1.1	Warum Generalisten hier gefährlich sind	79
5.1.2	Lessons Learned	81
5.2	Bauwesen und Infrastruktur	82
5.2.1	Fachterminologie als Risiko	82
5.2.2	Normengetriebene KI statt plausibler Sprache	83
5.2.3	Reale Effekte statt theoretischer Fehler	84
5.2.4	Lessons Learned	84
5.3	Sicherheit	85
5.3.1	Warum Sicherheit eine DSLM-Domäne ist	86
5.3.2	DSLMS für Security Operations	86
5.3.3	Angriff und Verteidigung	87
5.3.4	KI-Sicherheit als neue Unterdomäne	87

5.3.5	Hohe Markt- und Zukunftsrelevanz	88
5.4	Weitere Domänen im Vergleich.....	89
5.4.1	Finanzen.....	89
5.4.2	Recht	90
5.4.3	Software Engineering.....	91
5.4.4	Telefonie und Callcenter	91
6	EVALUATION UND ENTSCHEIDUNGSHILFEN	93
6.1	Evaluation von DSLMs	93
6.1.1	Warum generische Benchmarks versagen.....	94
6.1.2	Domänenspezifische statt allgemeiner Scores	95
6.1.3	Erfolg messbar machen	96
6.2	Die Zukunft der Domain AI.....	97
6.2.1	Modularisierung statt monolithischer Modelle	97
6.2.2	Multimodalität als Domänenanforderung	98
6.2.3	Open Source versus Kontrolle ist kein Widerspruch.....	98
6.2.4	Strategische Implikationen	99
6.3	Checkliste	100
6.3.1	Brauche ich überhaupt ein DSLM?.....	101
6.3.2	Welches Niveau an Spezialisierung ist notwendig?.....	102
6.3.3	Welche Risiken gehe ich ein?	103
6.3.4	Entscheidungsleitlinie	104
	ZUM SCHLUSS.....	105

ANHANG – GLOSSAR.....	109
QUELLENVERZEICHNIS.....	115
STICHWORTVERZEICHNIS	119
MEHR VON BRAIN-MEDIA	123

Vorwort

Große Sprachmodelle haben in kürzester Zeit eine technologische Zäsur ausgelöst. Systeme wie GPT, LLaMA oder vergleichbare Foundation Models zeigen eindrucksvoll, wie leistungsfähig statistische Sprachverarbeitung inzwischen geworden ist. Diese Erfolge haben jedoch eine problematische Vorstellung befördert: die Annahme, ein einziges, universelles Modell könne langfristig alle fachlichen, organisatorischen und regulatorischen Anforderungen gleichermaßen erfüllen.

In der Praxis zeigt sich zunehmend das Gegenteil. Dort, wo präzises Fachwissen, terminologische Eindeutigkeit, normative Regeln und haftungsrelevante Entscheidungen eine Rolle spielen, stoßen generalistische Sprachmodelle systematisch an ihre Grenzen. In der Medizin, im Bauwesen, im Recht, in der Finanzwelt oder in sicherheitskritischen technischen Domänen reichen „meist plausible“ Antworten nicht aus. Fehlinterpretationen, Halluzinationen oder semantische Ungenauigkeiten sind hier nicht nur technische Mängel, sondern reale Risiken.

Vor diesem Hintergrund gewinnen Domain-Specific Language Models (DSLMS) an Bedeutung. Sie stehen für einen grundlegenden Paradigmenwechsel: weg vom One-Model-Fits-All-Denken, hin zu spezialisierten, kontrollierbaren und domänenpräzisen KI-Systemen. DSLMS sind keine kleinere Variante großer Modelle und auch kein reines Fine-Tuning-Produkt. Sie sind Ausdruck einer anderen Denkweise – einer Denkweise, die Domänen als komplexe Wissens-, Regel- und Entscheidungs-

systeme begreift. Dieses Buch richtet sich an Leserinnen und Leser, die vor einer konkreten Frage stehen:

Wann reicht ein generalistisches Sprachmodell – und wann ist domänenspezifische KI zwingend erforderlich?

Es richtet sich an technische Entscheider, Architekten, fortgeschrittene Praktiker und organisationsnahe Experten, die nicht nach Hype, sondern nach belastbaren Entscheidungsgrundlagen suchen. Der Fokus dieses Buches liegt daher nicht auf einer weiteren Einführung in Large Language Models. Stattdessen verfolgt es drei zentrale Ziele:

1. Erstens entwickelt es ein präzises Verständnis davon, was eine Domäne im technischen und organisatorischen Sinn tatsächlich ausmacht – jenseits bloßer Thementzuordnung oder Branchenlabels.
2. Zweitens zeigt es, welche architektonischen Optionen für domänenspezifische Sprachmodelle existieren, welche Vor- und Nachteile sie besitzen und unter welchen Rahmenbedingungen sie sinnvoll eingesetzt werden können. Dabei stehen nicht Tools oder Frameworks im Vordergrund, sondern Entscheidungslogiken, Trade-offs und langfristige Tragfähigkeit.
3. Drittens beleuchtet es anhand ausgewählter Leitdomänen, wie DSLMs in der Praxis funktionieren, wo sie messbaren Mehrwert liefern und welche Fehler bei Planung, Training und Betrieb vermieden werden müssen.

Ein zentrales Anliegen dieses Buches ist es, die Diskussion um DSLMs zu entemotionalisieren. Weder sind sie Allheilmittel, noch sind sie ein akademisches Randthema. Sie sind ein strategisches Instrument für Organisationen, die Verantwortung, Haftung, Qualität und Wettbewerbsfähigkeit ernst nehmen. Entsprechend behandelt dieses Buch nicht nur Trainingsmethoden, sondern auch Fragen von Governance, Datenhoheit, Betriebskosten und regulatorischer Einbettung.

Gleichzeitig ist dem Autor bewusst, dass dieses Werk – wie jedes Buch zu einem sich schnell entwickelnden Technologiefeld – eine Momentaufnahme darstellt. Umso mehr wurde darauf geachtet, zeitstabile Konzepte, Muster und Entscheidungsrahmen in den Mittelpunkt zu stellen und kurzfristige Tool-Trends bewusst zu vermeiden.

Dieses Buch lädt nicht dazu ein, jedes Problem mit einem eigenen Modell zu lösen. Es lädt dazu ein, bewusst zu entscheiden, wann Spezialisierung notwendig ist, wie tief sie gehen sollte und welche Verantwortung mit ihr einhergeht.

Wenn es dazu beiträgt, Domain-Specific Language Models nicht als modische Abzweigung, sondern als konsequenten nächsten Schritt der KI-Entwicklung zu verstehen, hat es sein Ziel erreicht.

Herzlichst

Holger Reibold

(Dezember 2025)

1 One Model Does Not Fit All

Die rasante Verbreitung großer Sprachmodelle hat in vielen Organisationen eine zentrale Frage aufgeworfen:

Reicht ein generalistisches Large Language Model aus – oder braucht es eine domänenspezifische Lösung?

In den frühen Phasen der aktuellen KI-Welle wurde diese Frage häufig pragmatisch beantwortet. Allgemeine Modelle waren leicht verfügbar, leistungsfähig und beeindruckend vielseitig. Sie ermöglichten schnelle Prototypen, erste Produktivsysteme und neue Formen der Wissensarbeit. Doch mit zunehmender Integration in operative Prozesse zeigt sich, dass Vielseitigkeit allein kein ausreichendes Qualitätsmerkmal ist.

Sobald Sprachmodelle in fachlich präzise, regulierte oder sicherheitskritische Kontexte eingebettet werden, verschiebt sich der Bewertungsmaßstab. Terminologische Genauigkeit, nachvollziehbare Entscheidungslogik, Aktualität von Wissen und regulatorische Konformität werden wichtiger als fluide Sprache oder kreative Generalisierung. In solchen Umgebungen sind Fehler nicht nur ärgerlich, sondern potenziell haftungsrelevant.

Genau an diesem Punkt geraten generalistische Modelle systematisch an ihre Grenzen. Sie verfügen weder über ein konsistentes domänenspezifisches Vokabular noch über explizite Wissens- oder Regel-

strukturen. Ihr Verhalten basiert auf statistischer Approximation, nicht auf fachlicher Verankerung. Das führt zu bekannten Phänomenen wie Halluzinationen, semantischen Fehlinterpretationen oder inkonsistenten Antworten – insbesondere dort, wo Fachbegriffe, Normen oder implizite Annahmen eine zentrale Rolle spielen.

Domain-Specific Language Models sind eine Antwort auf diese strukturellen Defizite. Sie verfolgen einen anderen Ansatz: Statt möglichst viel allgemeines Wissen abzubilden, konzentrieren sie sich auf klar abgegrenzte fachliche Kontexte. Sie internalisieren Terminologie, Datenformate, Wissensstrukturen und normative Regeln einer Domäne und ermöglichen dadurch eine deutlich höhere Präzision und Kontrolle.

Dabei ist es wichtig zu betonen, dass DSLMs kein monolithisches Konzept darstellen. Sie reichen von relativ leichten Spezialisierungsformen – etwa der Einbindung externer Wissensquellen – bis hin zu vollständig eigenständigen Modellen mit domänenspezifischem Vokabular und Training von Grund auf. Welche Form sinnvoll ist, hängt nicht von technischer Machbarkeit ab, sondern von fachlichen Anforderungen, regulatorischem Rahmen, verfügbaren Daten und organisatorischer Verantwortung.

Dieses Buch setzt genau an dieser Entscheidungsstelle an. Es versteht sich nicht als Einführung in maschinelles Lernen oder als Katalog aktueller KI-Tools. Stattdessen richtet es sich an Leser, die vor konkreten Architektur-, Betriebs- und Governance-Fragen stehen:

Wie viel Spezialisierung ist notwendig? Welche Risiken lassen sich dadurch reduzieren? Und welche neuen Verpflichtungen entstehen durch domänenspezifische Modelle?

Um diese Fragen zu beantworten, folgt das Buch einer klaren, schrittweisen Struktur. Zunächst wird gezeigt, warum generalistische Sprachmodelle in vielen Fachdomänen nicht nur praktisch, sondern strukturell an Grenzen stoßen und weshalb das Paradigma eines universell einsetzbaren Modells nicht tragfähig ist. Im Mittelpunkt stehen dabei Fehlkosten, normative Bindung und implizites Fachwissen als zentrale Treiber für Spezialisierung.

Darauf aufbauend entwickelt das Buch ein präzises Verständnis davon, was eine Domäne im technischen, organisatorischen und governancebezogenen Sinn ausmacht. Es wird erläutert, welche Anforderungen sich daraus für Daten, Verantwortung, Auditierbarkeit und Betrieb ergeben – und warum Domänentiefe nicht nachträglich „aufgepfropft“ werden kann.

Im weiteren Verlauf werden die konkreten Architektur- und Umsetzungspfade systematisch eingeordnet – von Grey-Box-Ansätzen wie Prompting und External Augmentation bis hin zu White-Box-Strategien und echten Domain-Specific Language Models. Dabei stehen nicht Werkzeuge, sondern Entscheidungslogiken, Risiken und Kontrollgrade im Vordergrund.

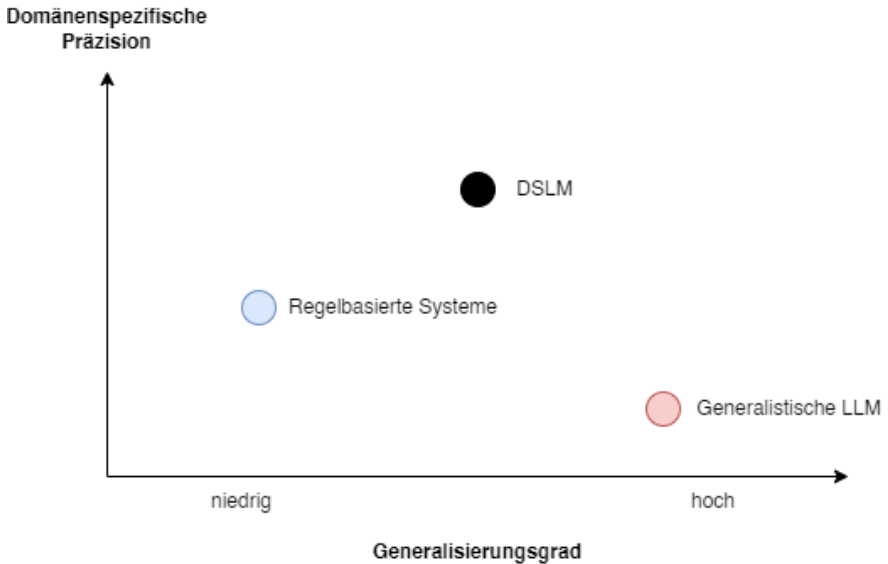
Abschließend werden diese Konzepte anhand ausgewählter Praxisdomänen zusammengeführt. Beispiele aus Biomedizin, Bauwesen, Sicherheit, Recht und Finanzen zeigen, wo domänenspezifische Sprachmodelle funktionieren, wo sie zwingend erforderlich sind – und wo bewusst darauf verzichtet werden sollte.

Ziel dieses Buches ist es, Orientierung zu geben – nicht durch Vereinfachung, sondern durch Struktur. Es soll Leserinnen und Leser in die Lage versetzen, fundierte Entscheidungen über den Einsatz, den Aufbau und den Betrieb von Domain-Specific Language Models zu treffen. Nicht jedes Problem erfordert ein eigenes Modell. Doch dort, wo Präzision, Verantwortung und Nachvollziehbarkeit entscheidend sind, ist Spezialisierung kein Luxus, sondern eine Notwendigkeit.

Die rasanten Fortschritte großer Sprachmodelle haben eine Vorstellung geprägt, die auf den ersten Blick plausibel erscheint: Je größer und allgemeiner ein Modell ist, desto besser eignet es sich für alle Anwendungsfälle. Skalierung, so die implizite Annahme, ersetze Spezialisierung. In der Praxis zeigt sich jedoch zunehmend, dass dieses Paradigma an strukturelle Grenzen stößt – insbesondere dort, wo Sprache nicht Selbstzweck, sondern Träger fachlicher Entscheidungen ist.

Generalistische Sprachmodelle sind darauf optimiert, statistische Regularitäten in großen, heterogenen Textkorpora zu erfassen. Ihre Stärke liegt in der Breite: Sie können über viele Themen sprechen, Kontexte flexibel wechseln und auch mit unvollständigen oder unscharfen Eingaben umgehen. Diese Eigenschaften sind in offenen, kreativen oder explorativen Szenarien von großem Wert. Sie werden jedoch

problematisch, sobald präzise Fachsprache, normatives Wissen oder haftungsrelevante Entscheidungen ins Spiel kommen.



Generalisierung vs. domänenspezifische Präzision: Der grundlegende Zielkonflikt zwischen Generalisierungsgrad und fachlicher Präzision. Generalistische Sprachmodelle maximieren Abdeckung, während domänenspezifische Modelle gezielt Präzision und Kontrollierbarkeit erhöhen. Skalierung allein ersetzt keine Spezialisierung.

Der zentrale Widerspruch liegt darin, dass Generalisierung und fachliche Präzision gegensätzliche Optimierungsziele darstellen. Ein Modell, das auf maximale Abdeckung ausgelegt ist, muss zwangsläufig domänenspezifische Feinheiten vernachlässigen. Es kann Begriffe korrekt

wiedergeben, ohne ihre exakte Bedeutung zu verstehen. Es kann Regeln beschreiben, ohne sie zuverlässig anzuwenden. Und es kann plausible Antworten generieren, ohne zwischen zulässigen und unzulässigen Schlussfolgerungen zu unterscheiden.

Dieses Spannungsfeld wird häufig durch kurzfristige Erfolge überdeckt. In frühen Pilotprojekten wirken generalistische Modelle erstaunlich leistungsfähig. Sie beantworten Fachfragen scheinbar korrekt, fassen Dokumente zusammen oder unterstützen bei der Analyse komplexer Texte. Erst mit zunehmender Nutzung werden ihre Grenzen sichtbar: Inkonsistenzen häufen sich, Randfälle werden falsch behandelt, und Entscheidungen lassen sich nicht reproduzierbar begründen. Was zunächst wie ein Randproblem erscheint, entpuppt sich in produktiven Systemen als strukturelles Risiko.

Besonders deutlich zeigt sich dies in Domänen mit hoher terminologischer Dichte. Fachbegriffe besitzen dort keine umgangssprachliche Bedeutung, sondern sind präzise definiert und in formale Wissenssysteme eingebettet. Ein generalistisches Modell behandelt solche Begriffe jedoch als statistische Einheiten, nicht als konzeptuelle Bausteine. Das führt zu subtilen, aber folgenschweren Fehlern – etwa, wenn ein medizinischer Begriff korrekt formuliert, aber im falschen klinischen Kontext verwendet wird, oder wenn eine juristische Norm zwar zitiert, jedoch falsch interpretiert wird.

Ein weiterer struktureller Schwachpunkt generalistischer Modelle ist ihr Umgang mit Normen und Regeln. In vielen Fachdomänen sind Entscheidungen nicht interpretativ, sondern regelgebunden. Ob eine

Aussage korrekt ist, hängt nicht von sprachlicher Plausibilität ab, sondern von der Einhaltung expliziter Vorgaben. Generalistische Modelle besitzen jedoch keine Mechanismen, um solche Regeln zuverlässig zu internalisieren oder durchzusetzen. Sie approximieren normatives Wissen, ohne es zu garantieren.

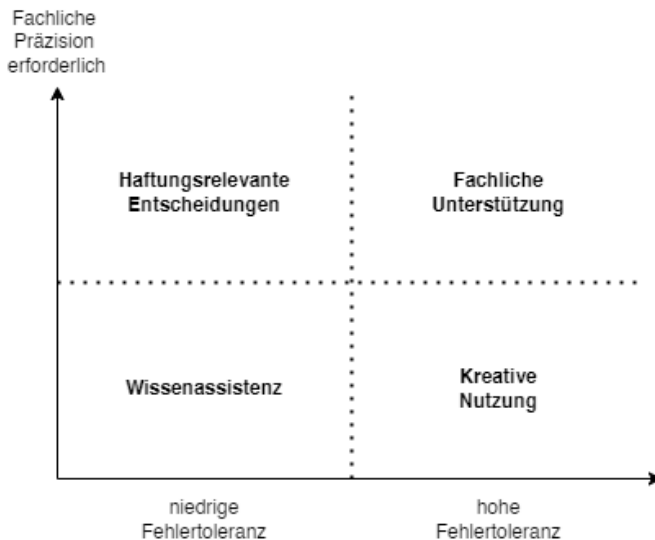
Hinzu kommt, dass große Sprachmodelle keine inhärente Kenntnis über die Konsequenzen ihrer Ausgaben besitzen. Sie optimieren auf Wahrscheinlichkeiten, nicht auf Verantwortung. In sicherheitskritischen oder regulierten Umgebungen ist diese Eigenschaft problematisch. Dort müssen Entscheidungen nachvollziehbar, überprüfbar und im Zweifel korrigierbar sein. Ein System, das nicht erklären kann, warum es eine bestimmte Antwort generiert hat, ist in solchen Kontexten nur eingeschränkt einsetzbar – unabhängig von seiner durchschnittlichen Leistungsfähigkeit.

Diese Einschränkungen lassen sich nicht allein durch Skalierung überwinden. Größere Modelle reduzieren zwar bestimmte Fehler, beseitigen jedoch nicht die zugrunde liegenden strukturellen Probleme. Mehr Parameter ersetzen keine explizite Domänenlogik. Mehr Trainingsdaten erzeugen kein normatives Verständnis. Und höhere Sprachflüssigkeit verbessert nicht automatisch die fachliche Verlässlichkeit.

Domain-Specific Language Models setzen genau hier an. Sie verfolgen nicht das Ziel maximaler Allgemeinheit, sondern gezielter Passgenauigkeit. Durch die bewusste Einschränkung auf klar definierte Domänen können sie Terminologie konsistent abbilden, Wissensstrukturen stabilisieren und Regeln zuverlässiger berücksichtigen. Diese Fokussierung

ist kein Rückschritt, sondern eine notwendige Spezialisierung – vergleichbar mit der Entwicklung von Fachsprachen oder spezialisierten Werkzeugen in anderen technischen Disziplinen.

Das bedeutet nicht, dass generalistische Modelle obsolet werden. Im Gegenteil: Sie bleiben ein wertvolles Fundament für viele Anwendungen. Entscheidend ist jedoch, zu erkennen, wann Generalisierung ausreicht – und wann sie zum Risiko wird. Diese Entscheidung lässt sich nicht abstrakt treffen, sondern nur im Kontext konkreter fachlicher, organisatorischer und regulatorischer Anforderungen.



Einsatzkontexte nach Fehlertoleranz: Nicht alle Anwendungsfälle stellen dieselben Anforderungen an Präzision und Fehlertoleranz. Die Matrix verdeutlicht, in welchen Kontexten generalistische Sprachmodelle sinnvoll einsetzbar sind – und wo spezialisierte Modelle erforderlich werden.

Dieses Buch geht von der Prämisse aus, dass es keinen universellen Königsweg gibt. Stattdessen plädiert es für eine differenzierte Betrachtung: Generalistische Modelle dort, wo Flexibilität und Breite gefragt sind; domänenspezifische Modelle dort, wo Präzision, Verantwortung und Nachvollziehbarkeit im Vordergrund stehen. Die folgenden Kapitel entwickeln die Kriterien, um diese Abgrenzung systematisch vorzunehmen.

1.1 Wo generalistische Sprachmodelle scheitern

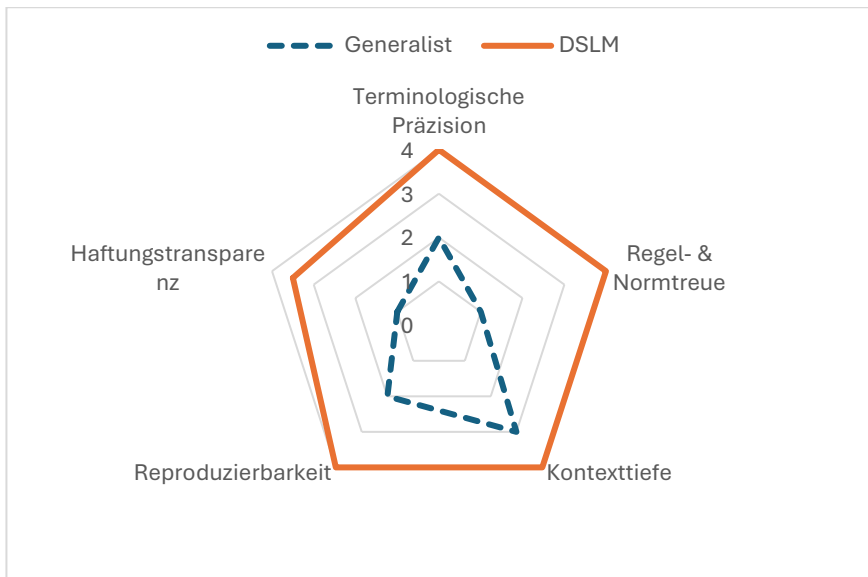
Die Grenzen generalistischer Sprachmodelle werden selten durch spektakuläre Fehlfunktionen sichtbar. In der Praxis scheitern sie meist leise, schrittweise und in Bereichen, die auf den ersten Blick unkritisch erscheinen. Gerade diese unauffälligen Fehler machen sie in fachlich anspruchsvollen Domänen gefährlich. Dieses Kapitel analysiert die typischen Versagensmuster generalistischer Modelle dort, wo Sprache nicht Selbstzweck, sondern Träger von Wissen, Regeln und Entscheidungen ist.

1.1.1 Terminologische Präzision als Schwachstelle

In vielen Fachdomänen ist Sprache kein Mittel zur Beschreibung, sondern ein formales Werkzeug. Begriffe sind eindeutig definiert, oft normiert und kontextabhängig. Generalistische Sprachmodelle behandeln Terminologie jedoch probabilistisch. Sie erkennen, welche Wörter

häufig gemeinsam auftreten, nicht welche Bedeutung sie in einem bestimmten fachlichen Kontext besitzen.

Das führt zu Fehlern, die sprachlich korrekt, fachlich jedoch falsch sind. Ein Modell kann einen medizinischen Begriff korrekt formulieren, ihn aber in einem klinisch unzulässigen Zusammenhang verwenden. Es kann juristische Termini korrekt zitieren, ohne die dahinterliegende Norm richtig zu interpretieren. Solche Fehler sind besonders problematisch, weil sie schwer zu erkennen sind: Sie klingen plausibel, widersprechen aber impliziten fachlichen Regeln.



Halluzination als strukturelles Risiko: Generalistische Sprachmodelle weisen wiederkehrende strukturelle Schwächen auf, insbesondere in terminologischer Präzision, Regelbindung und Nachvollziehbarkeit. Domänenspezifische Modelle reduzieren diese Risiken durch gezielte Einschränkung und Fokussierung.

Quellenverzeichnis

- Aghaei, E., Niu, X., Shadid, W., & Al-Shaer, E. (2022). Securebert: A domain-specific language model for cybersecurity. In International Conference on Security and Privacy in Communication Systems (pp. 39-56). Cham: Springer Nature Switzerland.
- Alvarez, G., & Paulman, T. (2025). Top 10 Strategic Technology Trend 2026. Online: <https://www.gartner.com/en/articles/top-technology-trends-2026>.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33.
- Chen, S. F., & Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Computer Speech & Language*, 13(4), 359–393.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.
- Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., & Poon, H. (2021). Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1), 1-23.
- Haslum, P., Lipovetzky, N., Magazzeni, D., Muise, C., Brachman, R., Rossi, F., & Stone, P. (2019). An introduction to the planning domain definition language (Vol. 13, pp. 1-187). Morgan & Claypool Publishers.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hulsen, T. (2022). Literature analysis of artificial intelligence in biomedicine. *Annals of translational medicine*, 10(23), 1284.

- Köckemann, U. (2025). AIDDL: The AI Domain Definition Language for integrated AI systems. *SoftwareX*, 31, 102259.
- Ling, C., Zhao, X., Lu, J., Deng, C., Zheng, C., Wang, J., ... & Zhao, L. (2025). Domain specialization as the key to make large language models disruptive: A comprehensive survey. *ACM Computing Surveys*, 58(3), 1-39.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv:1907.11692.
- Lötsch, J., Kringel, D., & Ultsch, A. (2021). Explainable artificial intelligence (XAI) in biomedicine: Making AI decisions trustworthy for physicians and patients. *BioMedInformatics*, 2(1), 1-17.
- Matarazzo, A., & Torlone, R. (2025). A survey on large language models with some insights on their capabilities and limitations. arXiv preprint arXiv:2501.04040.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Technical Report.
- Suzuki, M., Sakaji, H., Hirano, M., & Izumi, K. (2023). Constructing and analyzing domain-specific language model for financial text mining. *Information Processing & Management*, 60(2), 103194.
- Takyar, A. (2024). How to train an open-source foundation model into a domain-specific LLM? Online: <https://www.lee-wayhertz.com/how-to-train-domain-specific-llms/>.
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., & Scialom, T. (2023). LLaMA: Open and efficient foundation language models. arXiv:2302.13971.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Wang, B., Wang, Z., Wang, X., Cao, Y., A Saurous, R., & Kim, Y. (2023). Grammar prompting for domain-specific language generation

with large language models. *Advances in Neural Information Processing Systems*, 36, 65030-65055.

Winograd, T. (1972). *Understanding natural language*. Academic Press.

Woods, W. A. (1973). An experimental parsing system for transition network grammars. In R. Rustin (Ed.), *Natural Language Processing* (pp. 111–154). Algorithmics Press.

ZHAO, X., LU, J., DENG, C., ZHENG, C., WANG, J., CHOWDHURY, T., ... & Gu, Q. (2023). Beyond one-model-fits-all: A survey of domain specialization for large language models. *arXiv preprint arXiv*, 2305.

Zheng, Z., Lu, X. Z., Chen, K. Y., Zhou, Y. C., & Lin, J. R. (2022). Pre-trained domain-specific language model for natural language processing tasks in the AEC domain. *Computers in Industry*, 142, 103733.

Stichwortverzeichnis

A

Adapt	61
Adapter	60
Analyse	10
Approximation	6
Assistenzsystem	45
Auditierbarkeit	45, 71

B

Bauwesen	82
Benchmark	94
Betrieb	72
Biomedizin	78
Black-Box	47
Build	61
Buy	61

C

Callcenter	91
Checkliste	100
Compliance	69

D

Daten	65
-------------	----

Datenfokussierung	68
Datenmenge	60
Datenqualität	60
Domain Councils	105
Domain-Specific Language Model	1
Domäne	2, 27
Domänenspezifische Präzision	9
Domänentiefe	31
Domänenwissen	35
DSL	1

E

Ebenen	28
Eignung	78
Einsatzkontext	12
Erfolg	96
Evaluation	93
External Augmentation	47

F

Fachdomäne	10
Fachterminologie	82
Fehlannahmen	51
Fehlertoleranz	12
Fehlfunktion	15

Fehlkosten.....	18
Finanzen.....	89
Fine-Tuning.....	1
Flexibilität.....	13
Foundation Model.....	1

G

Generalisierung.....	9
Generalistisches Sprachmodell	8
Gesundheitswesen.....	78
Governance.....	65, 69
GPT	1
Grey-Box	47, 53, 57

H

Haftungsrisiken	18
Halluzination	15

I

Implikationen	99
Implizites Wissen.....	55
Infrastruktur	82

J

Juristisch Norm.....	16
----------------------	----

K

Kontextverlust.....	17
Korrektheit	16
Kosten	19, 73

L

Large Language Model.....	5
LLaMA.....	1
LLMOps	75, 106

M

Machine-Learning	72
MLOps	74
Modularisierung.....	97
Multimodalität	98

N

Nachvollziehbarkeit	17
NLP.....	29
Normdurchsetzung	16
Normen.....	36
Normengetriebene KI	83
Nutzungsmonitoring.....	74

O

Ökonomische Bewertung	24
One-Model-Fits-All.....	1

Open Source	98
Overfitting.....	36
Ownership.....	69

P

Passgenauigkeit.....	11
Prompting.....	53

R

RAG.....	45
Rechnische Norm	16
Recht.....	90
Regeln	36
Regelverbindlichkeit	56
Regulatorische Anforderungen.....	23
Reifezyklus	106
Repräsentation	29
Reproduzierbarkeit	17
Retrieval-Augmented Generation	45
Risikoaggregation	24

S

Schwächen.....	14
Security Operations	86
Sicherheit	85
Skalierung	11
Software Engineering	91
Spezialisierung	7, 38

Sprache	5
Sprachflüssigkeit	28
Sprachmodell	1
Subtokenisierung	29

T

Telefonie	91
Terminologie-Adaption	34
Terminologische Präzision.....	13
Total Cost of Ownership	73
Training from Scratch	60
Trainingsinfrastruktur	60
Transfer Learning	32

V

Verantwortungsdiffusion	21
Verständnis.....	36
Vokabularabdeckung	29
Vollständigkeit	16

W

White-Box	47, 57
Wissensarbeit	5
Wissensquelle	6
Wissensstruktur.....	11
Wissenssystem.....	32

Z

Zukunft 97

Zugangsmodell 46