Mathematik für Künstliche Intelligenz

Eine Landkarte der relevanten Gebiete

Dr. Jo F.R. Granaria

ORCID: 0009-0006-5203-5100 | jo@va.gmbh

F.R.Granaria AC

August 2026

Einleitung

Wer sich der Künstlichen Intelligenz nähert, steht vor einer scheinbar uferlosen mathematischen Landschaft: Von linearer Algebra über Maßtheorie bis zur Differentialgeometrie scheint alles irgendwie „relevant“. Diese Unübersichtlichkeit ist ein didaktisches Problem, kein sachliches. Denn die Landschaft ist geordnet — sie zerfällt in drei klar unterscheidbare Ebenen: einen *Kern* von vier Gebieten, ohne den kein Verständnis moderner Modelle möglich ist; eine Schicht *wichtiger Ergänzungen*, die vor allem in der Praxis trägt; und ein Feld von *Spezialgebieten*, deren Relevanz von der gewählten Forschungsrichtung abhängt.

Die These dieses Beitrags ist quantifizierbar: Die vier Kerngebiete — lineare Algebra, mehrdimensionale Analysis, Wahrscheinlichkeitstheorie und Optimierung — decken rund neunzig Prozent dessen ab, was zum Verstehen und Implementieren moderner Modelle nötig ist (Deisenroth et al., 2020). Der Rest wird themenspezifisch nachgeladen. Wer diese Ordnung kennt, lernt nicht weniger, aber in der richtigen Reihenfolge.

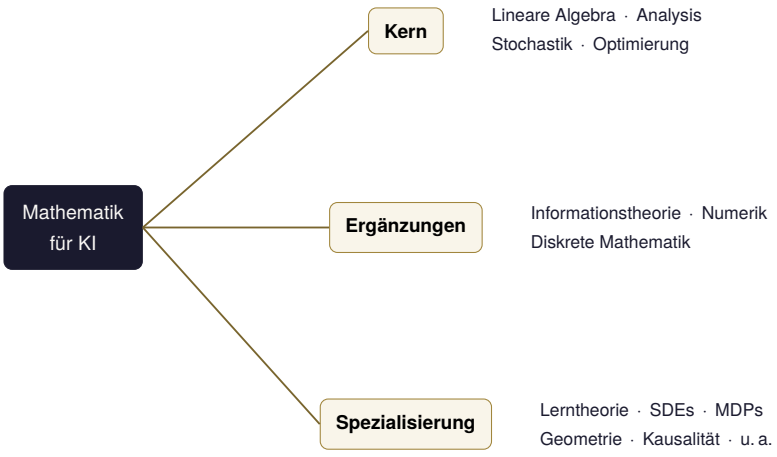


Abbildung 1: Drei Ebenen mathematischer Relevanz für die KI.

1 Der Kern: vier Gebiete

1.1 Lineare Algebra

Die lineare Algebra ist das wichtigste Einzelgebiet — nicht als eines unter mehreren, sondern als die Sprache, in der neuronale Netze überhaupt formuliert sind. Jede Schicht eines Netzes ist eine affine Abbildung, gefolgt von einer Nichtlinearität; der Attention-Mechanismus moderner Sprachmodelle ist im Kern ein Produkt dreier Matrizen; die Hauptkomponentenanalyse ist eine Eigenwertzerlegung der Kovarianzmatrix. Wer Vektorräume, Matrixmultiplikation, Eigenwerte und Eigenvektoren, die Singulärwertzerlegung, Rangbegriff, Normen und Projektionen beherrscht, liest die Architektur eines Modells wie eine Partitur.

Hinzu tritt ein praktischer Punkt, der oft unterschätzt wird: Für die tägliche Arbeit ist *Fließsicherheit* wertvoller als Breite. Wer Tensor-Shapes, Broadcasting-Regeln und Indexstrukturen im Kopf verfolgen kann, debuggt in Minuten, wofür andere Stunden brauchen.

1.2 Mehrdimensionale Analysis

Training ist Differentialrechnung in hoher Dimension. Der Gradient zeigt die Richtung des steilsten Anstiegs der Verlustfunktion; die Kettenregel, systematisch über einen Berechnungsgraphen angewandt, *ist* der Backpropagation-Algorithmus (Goodfellow et al., 2016). Jacobi- und Hesse-Matrizen beschreiben lokale Geometrie und Krümmung der Verlustlandschaft; Taylor-Entwicklungen begründen, warum Verfahren zweiter Ordnung konvergieren und wann Approximationen erster Ordnung genügen. Ohne diese Begriffe bleibt das Training neuronaler Netze eine Beschwörungsformel; mit ihnen wird es ein nachvollziehbarer Rechenvorgang.

1.3 Wahrscheinlichkeitstheorie und Statistik

Das gesamte moderne Framing des maschinellen Lernens — ein Modell ist eine Wahrscheinlichkeitsverteilung über Daten — lebt von der Stochastik. Verteilungen, bedingte Wahrscheinlichkeit, der Satz von Bayes, Erwartungswert und Varianz bilden das Fundament; Maximum-Likelihood-Schätzung erklärt, warum die üblichen Verlustfunktionen so aussehen, wie sie aussehen; Konzentrationsungleichungen und Schätztheorie begründen, warum aus endlichen Stichproben überhaupt verallgemeinert werden kann. Generative Modelle — Diffusionsmodelle, Variational Autoencoders — sind angewandte Wahrscheinlichkeitstheorie in Reinform.

1.4 Optimierung

Lernen heißt minimieren. Die Optimierungstheorie liefert das Vokabular dafür: Gradientenabstieg und seine stochastischen Varianten (SGD, Adam), Konvexität und ihre Grenzen — denn die Verlustlandschaften tiefer Netze sind gerade *nicht* konvex —, Lagrange-Multiplikatoren und KKT-Bedingungen für restringierte Probleme, Konvergenzverhalten und Lernraten-Schedules (Boyd & Vandenberghe, 2004). Wer Optimierung versteht, versteht, warum ein Training divergiert, warum Momentum hilft und was ein Warmup-Schedule leistet.

2 Wichtige Ergänzungen

2.1 Informationstheorie

Entropie, Kullback-Leibler-Divergenz und Kreuzentropie sind keine Randbegriffe, sondern die Standard-Verlustfunktionen der Klassifikation und des Sprachmodellierens (Cover & Thomas, 2006). Die Perplexität eines Sprachmodells ist exponentiierte Kreuzentropie; Mutual Information strukturiert Repräsentationslernen und Feature-Auswahl. Die Informationstheorie beantwortet die Frage, *was* ein Modell eigentlich lernt: die Kompression von Regelmäßigkeiten.

2.2 Numerik

Zwischen mathematischer Idealwelt und Fließkomma-Realität liegt die Numerik. Konditionierung, numerische Stabilität und die Eigenheiten reduzierter Präzision (`fp16`, `bf16`) erklären, warum Trainingsläufe explodieren, Gradienten verschwinden und Normalisierungsschichten nötig sind. Wer je einen `NaN`-Loss debuggt hat, weiß: Numerik ist keine Theorieübung, sondern Überlebenswissen.

2.3 Diskrete Mathematik und Graphentheorie

Graph Neural Networks, Knowledge Graphs, kombinatorische Suche und Constraint-Solving beruhen auf diskreten Strukturen. Auch klassische Informatik-Mathematik — Komplexität, Rekursion, Zählprinzipien — gehört hierher. Für symbolisch-neuronale Hybridsysteme ist dieses Gebiet zentral, für reine Deep-Learning-Praxis flankierend.

3 Spezialisierung nach Richtung

Jenseits von Kern und Ergänzungen hängt der mathematische Bedarf von der Forschungsrichtung ab. Tabelle 1 ordnet die wichtigsten Richtungen ihren Gebieten zu.

Zwei Beobachtungen zu dieser Tabelle. Erstens: Die Spezialgebiete setzen den Kern voraus, ersetzen ihn aber nicht — wer Optimal Transport ohne solide Wahrscheinlichkeitstheorie studiert, baut auf Sand. Zweitens: Die Zuordnung ist eine Relevanzordnung, keine Ausschlussliste; die Grenzen sind durchlässig, und gerade an den Übergängen — etwa zwischen Diffusionsmodellen und stochastischer Analysis — entsteht derzeit die interessanteste Forschung.

4 Der Lernpfad

Aus der Landkarte folgt eine Reihenfolge. Wer die Gebiete systematisch erarbeiten will, geht so vor:

Lineare Algebra ⟶ Analysis ⟶ Stochastik ⟶ Optimierung ⟶ Informationstheorie

Die Reihenfolge ist nicht willkürlich, sondern folgt den Abhängigkeiten: Die Analysis in hoher Dimension braucht die Vektorraumsprache der linearen Algebra; die moderne Stochastik formuliert ihre Objekte als Vektoren und Matrizen (Kovarianz!); die Optimierung setzt Gradienten und damit die Analysis voraus; die Informationstheorie schließlich verbindet Stochastik und Verlustfunktionen zu einem Gesamtbild.

Tabelle 1: Spezialgebiete nach Forschungsrichtung.

Richtung	Zusätzliche Mathematik
Theorie des Deep Learning	Maßtheorie, Funktionalanalysis, statistische Lerntheorie (VC-Dimension, PAC, Rademacher-Komplexität)
Generative Modelle	Stochastische Differentialgleichungen, Markov-Ketten, Variationsrechnung, Optimal Transport
Reinforcement Learning	Markov-Entscheidungsprozesse, dynamische Programmierung, Kontrolltheorie (Sutton & Barto, 2018)
Geometric Deep Learning	Gruppen- und Darstellungstheorie, Differentialgeometrie, Riemannsche Mannigfaltigkeiten (Bronstein et al., 2021)
Effizienz und Kompression	Tensor-Zerlegungen, Approximationstheorie, Quantisierungstheorie
Kausale Inferenz	Kausale Graphen (DAGs), Do-Kalkül, Identifizierbarkeitstheorie (Pearl, 2009)
Interpretierbarkeit	Vertiefte lineare Algebra (Superposition, Sparse Coding), Spieltheorie (Shapley-Werte)

Drei praktische Ratschläge. *Erstens:* Tiefe vor Breite — die vier Kerngebiete gründlich zu beherrschen ist wertvoller, als sieben Gebiete anzulesen. *Zweitens:* Implementieren als Verstehensprobe — wer Backpropagation einmal von Hand in dreißig Zeilen NumPy geschrieben hat, versteht die Kettenregel anders als aus dem Lehrbuch. *Drittens:* Die Spezialgebiete nicht auf Vorrat lernen, sondern dann, wenn ein konkretes Problem sie verlangt — mit solidem Kern ist jedes von ihnen in Wochen zugänglich, ohne ihn in keinem Zeitraum.

Gesamtbild

Die Mathematik der Künstlichen Intelligenz ist kein uferloses Gelände, sondern eine geordnete Landschaft mit einem kompakten Zentrum. Vier Gebiete tragen fast alles: die lineare Algebra als Sprache der Modelle, die Analysis als Theorie ihres Trainings, die Stochastik als Rahmen ihrer Deutung, die Optimierung als Mechanik ihres Lernens. Zwei bis drei Ergänzungen — Informationstheorie, Numerik, diskrete Strukturen — machen aus dem Verständnis Praxistauglichkeit. Alles Weitere ist Spezialisierung: legitim, lohnend, aber nachgeordnet.

Wer diese Ordnung ernst nimmt, gewinnt zweierlei. Erstens einen realistischen Lernplan, der in Monaten zu einem tragfähigen Fundament führt statt in Jahren zu fragmentarischem Überblickswissen. Zweitens ein Urteilskriterium: Wo immer ein neues Paper, eine neue Architektur, ein neues Paradigma auftaucht, lässt sich fragen, auf welcher Ebene der Landkarte es operiert — und damit, was man bereits mitbringt und was gezielt nachzuladen ist. Die Landkarte ersetzt das Gehen nicht. Aber sie verhindert, dass man in die falsche Richtung läuft.

Quellen

Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges*. arXiv: 2104.13478. Verfügbar 10. August 2026 unter https://arxiv.org/abs/2104.13478

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2. Aufl.). Wiley.

Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for Machine Learning*. Cambridge University Press. Verfügbar 10. August 2026 unter https://mml-book.github.io

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Verfügbar 10. August 2026 unter https://www.deeplearningbook.org

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2. Aufl.). Cambridge University Press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2. Aufl.). MIT Press.