KATALYSIA RESEARCH · PUBLIC

RESEARCH

Glossar -- Abkuerzungen des LLM-Trainings-Lebenszyklus

Wiederverwendbares Modul (Pre-Training, SFT, RLHF, DPO, RLAIF, RLVR, Safety/Red-Team)

Eigenstaendiges Glossar-Modul, einsetzbar in Vollfassung, Executive Summary und Presentation

Stand: 12. Mai 2026

Basis: Claude Opus 4.6 · KATALYSIA Research

GLOSSAR v1.0

Wiederverwendbares Modul für die KATALYSIA-Studie *Human-AI-Training und Integration von Spezialwissen in Large Language Models* (v1.0, 12. Mai 2026).

> *Dieses Glossar löst die Abkürzungen aus der Trainings-Pipeline auf, die in Folie 4 der Executive Presentation, in Kapitel 2.1 der Vollfassung und in der Markt-Analytik (Kapitel 5) konsistent verwendet werden. Es ist als eigenständiges Modul angelegt, damit es ohne Änderung in die Executive Summary, die Vollfassung oder künftige Folien-Versionen übernommen werden kann.*

Trainings-Phasen und Verfahren

Abk. / Phase	Langform	Kurzbeschreibung
Pre-Training	Pre-Training	Selbst-überwachtes Lernen auf großen, unstrukturierten Textkorpora (Common Crawl, Bücher, Code, Web). Liefert ein generelles Sprach- und Weltmodell.
Mid-Training	Mid-Training	Continued Pre-Training auf kuratierten Hochqualitätsdaten und Synthetic Data; erschließt fachliche Tiefe (Lehrbücher, Code-Repositorien, mathematische Beweise).
SFT	Supervised Fine-Tuning	Nachtraining auf 10 000–500 000 hochwertigen Instruction-Response-Paaren. Liefert Instruction-Following, Stil und Anweisungsbefolgung.
RLHF	Reinforcement Learning from Human Feedback	Drei-Stufen-Pipeline (SFT → Reward-Modell aus Präferenzpaaren → RL-Training, meist PPO oder GRPO). Prägt Helpfulness, Honesty, Harmlessness.
DPO	Direct Preference Optimization	Präferenzbasiertes Training ohne explizites Reward-Modell und ohne RL-Phase. Methodisch einfacher und compute-ärmer als RLHF; Varianten IPO, KTO, ORPO, SimPO.
RLAIF	Reinforcement Learning from AI Feedback	Bewertung durch ein anderes (oft größeres) LLM statt durch Menschen. Skaliert Volumina; Anthropic kombiniert RLAIF mit Constitutional AI zu RLHAIF.
RLVR	Reinforcement Learning from Verifiable Rewards	Reasoning-Training auf Aufgaben mit verifizierbaren Belohnungen (Mathe, Code, Logik). Dominanter Hebel der Frontier-Performance 2024–26.
Safety / Red-Team	Sicherheits- und Adversarial-Training	Refusal-Datensätze, Jailbreak-Tests, Capability-Evaluierung. EU-AI-Act Art. 55 verlangt dokumentiertes adversariales Testen für GPAI mit systemischem Risiko.

Ergänzende Begriffe und Varianten

Abk.	Langform	Kontext
PPO	Proximal Policy Optimization	Klassischer RL-Algorithmus in RLHF (OpenAI).
GRPO	Group Relative Policy Optimization	RL-Variante, in DeepSeek-R1 prominent verwendet.
IPO	Identity Preference Optimization	DPO-Variante, robuster gegen Overfitting auf deterministische Präferenzen.
KTO	Kahneman-Tversky Optimization	Arbeitet mit Einzelannotationen statt Paaren.
ORPO	Odds Ratio Preference Optimization	Kombiniert SFT und Präferenz-Lernen in einem Schritt.
SimPO	Simple Preference Optimization	Längen-normalisierte DPO-Variante.
Constitutional AI (CAI)	Constitutional AI	Prinzipiengeleitetes Alignment (Anthropic): Verfassung steuert Kritik und Revision; mit RLHF kombiniert ergibt RLHAIF.
GPAI	General-Purpose AI	EU-AI-Act-Klassifikation; bei systemischem Risiko Pflicht zu adversarialem Testen (Art. 55).
SME	Subject Matter Expert	Fachexperte als Datenquelle in SFT- und Domain-Trainings-Programmen.
PaaS	Platform-as-a-Service	Geschäftsmodell, z. B. Snorkel, Labelbox, Toloka.

KATALYSIA Research · Public · Stand: Mai 2026 · Erstellt mit Claude (Anthropic) · Opus 4.6.