

Big Data & Künstliche Intelligenz

Prüfungsrelevante Zusammenfassung der 5 Lernergebnisse

LZ1: Trends & Big Data

LZ2: ML-Methoden

LZ3/4: Frontier-Modelle

LZ5: Risiken & AI Act

Was ist Big Data & Künstliche Intelligenz?

Volume

Datenmenge

Petabytes täglich — allein Common Crawl archiviert Milliarden Webseiten

Velocity

Geschwindigkeit

Echtzeit-Datenströme: Sensoren, Social Media, Transaktionen

Variety

Vielfalt

Text, Bild, Audio, Video, Sensordaten — strukturiert & unstrukturiert

Paradigmenwechsel seit 2022: Von diskriminativer zu generativer KI

- › Früher: KI klassifiziert & vorhersagt (diskriminativ) → z.B. Spam-Filter, Kreditscoring
- › Heute: KI erstellt neue Inhalte (generativ) → GPT, Claude, Gemini, Mistral
- › Neu 2025/26: Agentic AI — KI plant & führt mehrstufige Aufgaben selbst aus

Woher lernt die KI — und was bedeutet das für Europa?

Datenquellen für LLM-Training

- **Web-Crawl:** Common Crawl, C4 — bis zu 80% des Trainingskorpus
- **Bücher & Wissenschaft:** Urheberrechtskonflikte aktiv (OpenAI, Meta)
- **Lizenzierte Daten:** Reddit-API (seit 2023 kostenpflichtig)
- **RLHF-Daten:** Menschliches Feedback für Feinabstimmung
- **Synthetische Daten:** KI generiert Trainingsdaten für sich selbst

Transparenz-Paradox

Kein Anbieter — auch Mistral nicht — legt Trainingsdaten vollständig offen. Stanford FMTI: Mistral Score 18/100.

CLOUD Act

US-Behörden können auf Daten US-amerikanischer Anbieter zugreifen — unabhängig vom Speicherort in der EU. Direkter DSGVO-Konflikt.

Open Weights ≠ Souveränität

Mistral veröffentlicht Modellgewichte, nicht Trainingsdaten. Auf BRZ-Infrastruktur betrieben = echte Souveränität.

Welche Methode für welchen Anwendungsfall?

Supervised Learning

Wann: Gelabelte Daten vorhanden

Bsp: Spam, Kredit-scoring, Förder-bewertung

Unsupervised Learning

Wann: Muster ohne Labels suchen

Bsp: Kunden-segmentierung, Betrugs-erkennung

Reinforcement Learning

Wann: Sequenz. Entscheidungen

Bsp: Roboter, RLHF bei LLMs

Generative AI (LLMs)

Wann: Inhalte erzeugen/transformieren

Bsp: GovGPT, KAPA, Rundschreiben

Transfer Learning

Wann: Wenig eigene Daten

Bsp: Behörden-chatbot, Fine-tuning

Entscheidungsregel: Gelabelte Daten? → Supervised | Muster suchen? → Unsupervised | Inhalte erzeugen? → GenAI | Wenig Daten? → Transfer Learning

Welches Modell ist das beste — und für wen?

Claude Opus 4.7

Anthropic  | SWE-Bench: 87,6%

Constitutional AI — Ethik eintrainiert. GPAI systemisches Risiko.

Coding #1

GPT-5

OpenAI  | AIME 2026: perfekt

Bestes Mathe-Reasoning. Größtes Ökosystem.

Mathe #1

Gemini 3.1 Pro

Google  | Kontext: 1M Token

Stärkstes Multimodal-Modell nativ.

Multimodal #1

Mistral Large 3

Mistral  | 675B MoE-Parameter

Einziges EU-Frontier-Modell. Apache 2.0 — selbst hostbar auf BRZ.

EU #1

DeepSeek V4 Pro

DeepSeek  | 34× günstiger

MIT-Lizenz. API über CN-Infrastruktur → DSGVO-Konflikt!

⚠ Public AI: NEIN

Kein Modell gewinnt in jeder Kategorie. Die richtige Frage: Welches Modell für welche Aufgabe, unter welchem Rechtsrahmen, mit welchem Hosting?

Was bedeutet KI für den österreichischen Arbeitsmarkt?

68%

nutzen KI
gelegentlich (AT, 2026)

50%

erwarten
Jobverluste durch KI

16%

Arbeitsinhalte
automatisierbar (AT)

-1,1%

Produktivität
2023–24 (vs. EU +0,5%)

Paradigmenwechsel: Wer ist betroffen?

Früher: Blue-Collar-Routinejobs (Fließband, Dateneingabe)

Heute: White-Collar-Jobs: Übersetzer, Programmierer,
Analysten, Juristen

Treiber: GenAI kann komplexe kognitive Aufgaben erledigen

AT-spezifisch: Hoher Anteil an öffentlicher Verwaltung &
Industriesektor besonders exponiert

3 Szenarien für Österreich

✓ Optimistisch:

KI kompensiert Demografierückgang, BRZ als souveräne Plattform

~ Realistisch:

Polarisierung; Sozialstaat (Gini -43%) federt ab, aber unter Druck

! Pessimistisch:

Zu langsame Adoption + Jobverlust; Umschulungssystem überfordert

Was der EU AI Act konkret verbietet und reguliert



Verboten seit Feb 2025 (Art. 5)

- Social Scoring durch staatliche Stellen
- Biometrische Echtzeit-Massenüberwachung
- Unterschwellige Verhaltensmanipulation
- Emotionserkennung am Arbeitsplatz / in Schulen
- Predictive Policing per demografischem Profiling
- Massensammlung von Gesichtserkennungsdaten

Hochrisiko ab Aug 2026 (Anhang III)

- Biometrie & Gesichtserkennung
- Kritische Infrastruktur (Strom/Wasser/Verkehr)
- Bildung & Prüfungsbewertung
- HR / Recruiting & Personalentscheidungen
- Kreditvergabe & Bonitätsscoring
- Strafverfolgung, Justiz, Migration

KI-Konklave März 2026: Umsetzungsfahrplan

Bundesverwaltung

Shared-Service-First

Gemeinsame BRZ-Infrastruktur für alle Ressorts — keine Insellösungen

Souveränität & Sicherheit

DSGVO-konforme, vertrauenswürdige Infrastruktur ohne unkontrollierten Datentransfer

Human-in-the-Loop

KI unterstützt — finale Entscheidung verbleibt beim Menschen (AI Act Art. 14)

Konkrete Vorhaben im Fahrplan

GovGPT:	Generischer LLM-Zugang für alle Bundesbediensteten
KI im ELAK:	Chatten mit Akten, Zusammenfassungen, IFG-Unterstützung
KAPA:	KI-Assistent für parlamentarische Anfragen
KI-Rundschreiben:	Automatisierte Erstellung aus bestehenden Datenbanken
Bundesscribe:	Meeting-Transkription, Protokoll, Übersetzung

Risikobasierte Modellwahl

Unkritisch:	US-Hyperscaler + AVV → KI-SUN, Barrierefreiheit
Operativ:	EU-Hosting / Sovereignty-Cloud + TIA → ELAK, Rundschreiben
Hochsensitiv:	On-Premise BRZ + Mistral Open Weight → KAPA, Legistik, Förderung

Was bleibt — 5 Lernziele auf einen Blick

01

Paradigmenwechsel:

GenAI ist kein besseres ML — es ist eine neue Kategorie. Agentic AI ist ab 2025 operational.

02

Trainingsdaten = Blackbox:

Kein Anbieter legt vollständig offen. Open Weights \neq Souveränität. CLOUD Act kollidiert mit DSGVO.

03

Kein Modell gewinnt alles:

Claude #1 Coding/Ethik | GPT-5 #1 Mathe | Mistral #1 EU-Souveränität | DeepSeek/Qwen: Public AI verboten.

04

White-Collar-Disruption:

16% AT-Arbeitsinhalte automatisierbar. Übersetzer, Programmierer, Analysten am stärksten betroffen.

05

AI Act schützt — aber begrenzt:

8 Verbote + Hochrisiko-Pflichten ab Aug 2026. Makroökonomische Risiken (Ungleichheit) adressiert er nicht.