LIMINAR ACADEMY · MBA · MODUL 3

Liminar

Künstliche Intelligenz und Gesellschaft

Das Überkapitel: der rote Faden über LV 3.1 (Technikethik) und LV 3.2 (Trustworthy AI) — Kompetenzaufbau in zwei Stufen

6 ECTS · Perspektive öffentlicher Sektor · APA 7 · Juli 2026

Die Kompetenztreppe — vom Verstehen zum Gestalten

Zwei Lehrveranstaltungen, ein Aufbau: LV 3.1 legt das Fundament, LV 3.2 operationalisiert es.

L1	Technikbegriff evaluieren	LV 3.1
L2	Technik-/Maschinenethik differenzieren	LV 3.1
L3	Moralische Algorithmen beurteilen	LV 3.1
L4	Roboterethik beurteilen	LV 3.1
L5	Prinzipien evaluieren (HLEG)	LV 3.2
L6	Nach Maximen einstufen	LV 3.2
L7	Fair-AI-Werkzeuge gestalten	LV 3.2

MERKSATZ *3.1 = Haltung (Werte, Verantwortung). 3.2 = Handwerk (prüfen, messen, verankern). Jede Stufe setzt die vorige voraus.*

TEIL I · LV 3.1

Technikethik — das ethische Fundament

Lernergebnisse 1–4: die Urteilsfähigkeit aufbauen

Technikbegriff & Ethik differenzieren

Instrumentalismus

Technik ist neutral — Werkzeug

Substantivismus

Technik trägt eigene Wertlogik („Gestell“)

Konstruktivismus

Technik ist sozial konstruiert — Werte im Design

Technik- vs. Maschinenethik

Technikethik: wie setzt der Mensch Technik ein? Maschinenethik: können Maschinen moralisch handeln? Value Alignment (Bostrom) → Human Oversight (Art. 14).

Drei Ethiktheorien

- Deontologie (Kant) → AI Act Art. 5 Verbote
- Konsequentialismus → risikobasierter Ansatz
- Tugendethik (Aristoteles) → Responsible Innovation

VERNETZUNG Merksatz: Je nach Technikbegriff verschiebt sich die Verantwortung — vom Artefakt zum politischen System.

Moralische Algorithmen & Roboterethik

L3 · Bias & Impossibility

- Bias-Taxonomie: Historical, Representation, Measurement, Aggregation, Deployment
- Impossibility Theorem (Chouldechova 2017): Calibration, Demographic Parity & Equalized Odds nicht zugleich
- → Fairness-Wahl ist politisch, nicht technisch
- Interpretability vs. Explainability (SHAP/LIME)

L4 · Responsibility Gap

- Sparrow (2007): bei Vollautonomie ist niemandem Verantwortung zurechenbar
- nicht durch bessere Technik lösbar
- Felder: Pflegeroboter, LAWS, autonome Fahrzeuge, RPA Public Sector
- → keine Vollautonomie bei Rechtsfolgen

KOMPETENZBRÜCKE Kompetenzbrücke 3.1 → 3.2: diese vier Urteilskompetenzen sind Voraussetzung, um die 7 Prinzipien zu verstehen.

TEIL II · LV 3.2

Trustworthy AI — die Operationalisierung

Lernergebnisse 5–7: aus Haltung wird prüfbares Handwerk

Die 7 HLEG-Prinzipien — lawful · ethical · robust

P1 Human Oversight

AI Act: Art. 14

P2 Robustness

AI Act: Art. 9, 15

P3 Privacy & Data

AI Act: Art. 10

P4 Transparency

AI Act: Art. 13

P5 Fairness

AI Act: Art. 10

P6 Societal Wellbeing

AI Act: schwach

P7 Accountability

AI Act: Art. 17, 26

MERKSATZ *HLEG (Soft Law 2019) = ethisches Fundament; AI Act (Hard Law) = rechtliche Operationalisierung — Übersetzung unvollständig.*

VERNETZUNG Trustworthy AI muss gleichzeitig lawful, ethical UND robust sein.

Einstufung in sechs Filtern & Fair-AI-Werkzeuge

L6 · Sechs sequenzielle Filter

- 1 deontologisch (Art. 5 Verbote)
- 2 konsequentialistisch (Nutzen > Schaden, Rawls)
- 3 tugendethisch (Art. 17)
- 4 regulatorisch (Annex III, DSGVO)
- 5 ALTAI (7 Prinzipien)
- 6 Gap-Analyse (Ethik-Compliance-Gap)

L7 · Werkzeuge über den Lebenszyklus

- Value Sensitive Design (Problemdefinition)
- Datasheets for Datasets (Daten)
- Bias-Audit — 80-%-Regel
- Model Cards (Modell)
- Algorithmic Impact Assessment (Deployment)
- Continuous Fairness Monitoring

MERKSATZ *Ethik-Compliance-Gap: Compliance ≠ Ethik. Aufgabe der Beratung: den Gap benennen und schließen.*

ChatGPT · Claude · Mistral an den 7 HLEG-Prinzipien

Prinzip	ChatGPT	Claude	Mistral	Sieger
P1 Human Oversight	●●○	●●●	●●○	Claude
P2 Robustness & Safety	●●●	●●●	●●○	ChatGPT = Claude
P3 Privacy & Data	●●○	●○○	●●●	Mistral
P4 Transparency	●●○	●●●	●●○	Claude
P5 Fairness	●○○	●○○	●●○	Mistral (knapp)
P6 Societal Wellbeing	●○○	●●○	●●●	Mistral
P7 Accountability	●●○	●●●	●○○	Claude

●●● stark · ●●○ mittel · ●○○ schwach · Bewertung teils wertend, Stand Juli 2026

MERKSATZ *Kein System erfüllt alle 7 — die Folge des Impossibility Theorems (L3). Die Systemwahl ist eine Wertentscheidung.*

SYNTHESE

Von der Haltung zum Handwerk

Teil I · LV 3.1

Urteilsfähigkeit: Werte, Verantwortung, Bias, Responsibility Gap

Teil II · LV 3.2

Instrumentarium: Prinzipien, Einstufung, Fair-AI-Werkzeuge

Fallstudie

Anwendung: kein System erfüllt alle 7 — Wahl ist normativ

Roter Faden: echte statt symbolische Vertrauenswürdigkeit — Substanz, nicht Geste.

KERNBOTSCHAFT

3.1 gibt die Haltung, 3.2 das Handwerk — zusammen machen sie KI in der Gesellschaft verantwortbar.