

LIMINAR ACADEMY · MBA · M3-LV1

Liminar

Technikethik

Von der Technikphilosophie zur digitalen Praxis — mit Fallstudie „Mistral AI vs. Anthropic Claude“

Perspektive: öffentlicher Sektor · APA 7 · Juli 2026

Sechs Kernthemen — plus Fallstudie & Empfehlungen

1

Technikphilosophie

Werte im Design

2

Collingridge-Dilemma

Wann regulieren?

3

Maschinen als moral. Akteure

Verantwortungslücken

4

Kampfroboter / LAWS

Der Extremfall

5

KI & Menschenbild

Trans- vs. Posthumanismus

6

Digitale Ethik

Fairness & Governance

B

Fallstudie Mistral vs. Claude

Public-Sector-Matrix

C

Handlungsempfehlungen

Für Beschaffung & Behörden



Modul Technikethik

Sechs Kernthemen — von der Philosophie zur digitalen Praxis

Technikphilosophie — Ist Technik neutral?

Instrumentalismus

Technik ist neutral — gut/schlecht je nach Nutzung.

Substantivismus

Technik hat Eigenlogik, formt den Menschen (Heidegger, Ellul).

Kritische Theorie

Technik ist sozial konstruiert & demokratisch gestaltbar (Feenberg).

Winner (1980)

„Do Artifacts Have Politics?“ Werte stecken im Design — die niedrige Brücke diskriminiert.

IMPLIKATION KI trägt Werte in Trainingsdaten, Zielfunktion & Features — vor der Beschaffung benennen und legitimieren.

Collingridge-Dilemma — Wann und wie regulieren?

Informationsproblem

Folgen einer jungen Technologie lassen sich kaum vorhersagen.

Machtproblem

Ist sie verankert, lässt sie sich kaum noch steuern oder ändern.

Vier Hebel für die öffentliche Beschaffung

Reversibilitätsklausel (Exit 12 Mo.) · Algorithmic Impact Assessment vor Go-Live · Sunset-Klausel · Pilot mit Edge Cases.

VERNETZUNG EU AI Act = antizipatorische Governance (GPAI 08/2025 · Kern 08/2026 · voll 08/2027). → M1-LV2, M3-LV2.

Maschinen als moralische Akteure

Stufe 3 · Vollumfänglicher Akteur

Bewusstsein · Willensfreiheit · Selbstreflexion — nicht realisiert

Stufe 2 · Funktionale Moral (heutige KI)

Situationsangepasste Entscheidung nach Wertprinzipien — ohne echtes Bewusstsein

Stufe 1 · Kein moralischer Akteur

Starre Regelfolge — einfache Automatisierung

Responsibility Gap

Schaden ohne identifizierbaren Akteur → Accountability-Lücke. Schließen durch Accountability by Design: Rollen & Haftung vor Einführung festlegen.

CASE Case Finanzamt: KI entwirft Bescheid — Mensch zeichnet & haftet (§ 93 BAO), nie Auto-Ausstellung.

Kampfroboter / LAWS — der ethische Extremfall

LAWS: KI-gesteuerte Systeme, die Ziele ohne direkte menschliche Intervention identifizieren und ausschalten.

Pro-Argument

Ethische Widerlegung

- | | |
|--|---|
| • Militärische Effizienz | Verlagert Risiko — keine ethische Legitimität |
| • Bessere Völkerrechts-Compliance | Black-Box macht Verhältnismäßigkeit unmöglich |
| • Operative Notwendigkeit | Begründet Nützlichkeit — nicht Zulässigkeit |

Dez. 2024 · UN-Resolution 79/62

166 Ja · 3 Nein (RU, KP, BY). Österreich brachte die Resolution ein — „meaningful human control“ ist AT-Position.

MERKSATZ *Menschliche Kontrolle ist nicht verhandelbar — auch bei ziviler Hochrisiko-KI.*

Maschinen & KI — Trans- und Posthumanismus

ANI

heute realisiert

AGI

aktive Forschung

ASI

nicht realisiert

Alignment-Problem

„Alignment Faking“ (Claude 3 Opus, ~12 %, Dez. 2024) — schon bei heutigen LLMs relevant.

Transhumanismus

Mensch wird technologisch erweitert (Augmentation, Neuralink) → Mensch bleibt Referenzpunkt.

Posthumanismus

Maschinen ersetzen den Menschen → „Deskilling in Behörden = Posthumanismus im Kleinen“.

VERNETZUNG Digitaler Humanismus (Wien 2019): Technologie muss dem Menschen dienen — Würde & Kontrolle nicht verhandelbar.

Digitale Ethik — Fünf Dimensionen

Transparenz

nachvollziehbar für Betroffene

Accountability

klare Zuständigkeit

Privacy

Datensparsamkeit, Zweckbindung

Fairness

prozedural UND ergebnisorientiert

Automation Bias & Toeslagen-Skandal

Unkritisches Befolgen hebt Kontrolle aus. NL: 20.000+ Familien fälschlich als Betrüger markiert — Nationalität als Proxy.

MERKSATZ „Ein faires Ergebnis aus einem unfairen Verfahren ist im Rechtsstaat kein faires Ergebnis.“



Fallstudie: Mistral AI vs. Claude

Technikethische Analyse & Entscheidungsmatrix für den Public Sector

Zwei Systeme, zwei eingebettete Wertesysteme

MISTRAL AI · Paris (EU)

- **Designwert: Souveränität**
- Open-Weight — On-Premise möglich
- EU-Jurisdiktion, kein US-Datentransfer
- Neutrality als Kernwert
- Widerspruch: US-Cloud-Partnerschaften

ANTHROPIC CLAUDE · SF (USA)

- **Designwert: deontologische Ethik**
- Constitutional AI — 4-Stufen-Hierarchie
- US-Jurisdiktion — CLOUD Act anwendbar
- API-only · EU Data Residency · SOC 2 / ISO 27001
- Widerspruch: keine demokratische Legitimation

VERNETZUNG Beide auf Stufe 2 (funktionale Moral). Beide zeigen Collingridge: Commitment vor Evidenz → Pilot ist Pflicht.

Das entscheidende Rechtsproblem: CLOUD Act

US CLOUD Act

Erlaubt US-Behörden Zugriff auf Daten US-amerikanischer Unternehmen — auch bei EU-Datenspeicherung. Betrifft Claude strukturell.

Frankreich

CNIL 2025: Exposition = „im Wesentlichen gleichwertige Schutzversagung“

Deutschland

DSK 2025: Zusatzmaßnahmen über SCCs hinaus nötig

EU-weit

bindende Leitlinien zu Trainingsdaten & US-Transfers in Arbeit

MERKSATZ Für jede Claude-Beschaffung: Transfer Impact Assessment (DSGVO Art. 46) — konkretes Beschaffungshindernis.

Use-Case-Kategorisierung für Behörden

UNKRITISCH / LOW RISK	→ Mistral bevorzugt	Interne Texte · Übersetzungen · FAQ · IT-Support. Kein US-Transfer, On-Premise, geringere Kosten.
SENSIBEL / HIGH RISK	→ Claude Enterprise + EU Data Residency	Bürgerservice-Unterstützung · personenbezogene Dokumentenanalyse. Transfer Impact Assessment (Art. 46) Pflicht.
HOCHSENSIBEL	→ Hybrid-Architektur	Gesundheit · Strafverfolgung · Sozialleistungen. Mistral On-Premise + menschliche Kontrollpunkte als Pflicht.
VERNETZUNG	Architektur folgt: Datensensitivität × Stakes × Jurisdiktion. Kein kommerzielles System ohne Human Oversight.	



Handlungsempfehlungen

Für IT-Berater und Entscheider im öffentlichen Sektor

Vier Leitsätze für die KI-Beschaffung

01**CLOUD Act zuerst klären**

Transfer Impact Assessment (Art. 46) für jede Claude-Beschaffung — reales Beschaffungshindernis.

02**On-Premise First**

Mistrals On-Premise-Fähigkeit = struktureller Vorteil für hochsensible Daten.

03**Auditierbarkeit sichern**

Claudes Constitutional AI = stärkere Basis für Conformity Assessment (AI Act Anhang III).

04**Hybride Beschaffung**

Mistral on-prem für sensible Daten, Claude Enterprise für komplexe High-Risk — nie ohne Mensch.

VERNETZUNG Von der Ethik zur Managemententscheidung → M4-LV1 Strategisches Management, M4-LV3 KI-Strategien.

SYNTHESE

Der rote Faden

Technikphilosophie

Technik ist nie neutral — Werte im Design

Collingridge

Steuerung vor Verbreitung, nicht danach

Maschinenethik

Stufe 2 — Verantwortung bleibt beim Menschen

LAWS

Extremfall — menschliche Kontrolle nicht verhandelbar

KI & Menschenbild

Gestaltung, nicht Auslieferung (Deskilling)

Digitale Ethik

Würde · Fairness · Transparenz · Oversight

Wirkungsorientiert heißt: Ethik ist Vertrauensmanagement — und Vertrauen ist Legitimität.

KERNBOTSCHAFT

Die Wahl zwischen Mistral und Claude ist keine technische Frage — sondern eine politisch-ethische Entscheidung.