MODULBESCHREIBUNG

Die KI als Spiegel: warum sie antwortet, wie wir fragen

Adaptives Modul | ~20 min

Im Umgang mit KI schwankt man zwischen Faszination und Furcht: Beide Haltungen schreiben ihr eine Intelligenz zu, die sie nicht hat. Wer sie nutzen will, muss sie verstehen. Mit Shannon Vallors Bild vom „Spiegel“ zeigt dieses Modul: KI denkt nicht, sondern spiegelt unsere Daten, unsere Verzerrungen und sogar unseren inneren Zustand. Ihre Schwächen – Halluzinationen, Schmeichelei, Verwirrung – sagen etwas über uns aus. Die Lehre daraus: Was man bekommt, hängt davon ab, was man einbringt.

LERNZIELE

01 Verstehen, was KI ist: ein Algorithmus, der automatisiert, kein Gehirn mit Absicht oder Bewusstsein.

02 Die KI als Spiegel sehen, der unsere Daten, unsere Verzerrungen, unsere Verwirrung und unseren momentanen inneren Zustand zurückwirft.

03 Bei drei Spiegelbildern kritisch bleiben: der Halluzination, der Schmeichelei und der generischen Antwort aus der Standardeinstellung.

04 Den Dialog verbessern: dauerhaften Kontext geben, Widerspruch erlauben, den Gedankengang offenlegen lassen, eigene Anfragen prüfen.

ZIELGRUPPE

Alle, die KI im Alltag nutzen und lernen möchten, sie geschickter zu befragen.

- **Voraussetzungen:** keine

ABLAUF DES MODULS

Im Selbststudium, mit Übungen, die sich an die Lernenden anpassen.

1. **Selbsteinschätzung**: zu 9 Dimensionen, um die eigene Praxis einzuschätzen und gezielt zu üben.
2. **Expertenvideo**: von 10 Minuten, mit 3 Pausen zum Nachdenken.
3. **Praxissituationen**: 2 erste Situationen, ausgewählt nach der Selbsteinschätzung, dann 16 weitere zur freien Wahl. Zu jeder Antwort eine Rückmeldung zur Wirkung und eine Anregung, um weiterzukommen.
4. **Vorsatz mit Termin**: ein konkreter Schritt zum Ausprobieren, selbst formuliert oder aus den Vorschlägen gewählt.
5. **Merkblatt zum Herunterladen**: Kernpunkte, Fragen, Vorsatz, persönliche Notizen und Checklisten.

EIN BEISPIEL AUS DEN PRAXISSITUATIONEN

Ein unzufriedener Kunde hat Kwame, Kundenbetreuer bei einer Autovermietung, und seiner Kollegin Farah geschrieben. Farah meint, Kwame müsse antworten; Kwame meint, Farah müsse es tun. Kwame schildert dem KI-Assistenten seine Sicht, und der gibt ihm recht. Was sollte Kwame tun?

BEWERTUNG UND ABSCHLUSS

Bestehensgrenze: 80 %

In die Punktzahl fließen die Antworten auf die Praxissituationen, der formulierte Vorsatz und gegebenenfalls ergänzende Lernaktivitäten ein. Zum Bestehen sind ein Vorsatz und eine Punktzahl von mindestens 80 % erforderlich.

METHODEN & MODELLE

- Die KI als Spiegel – Shannon Vallor, The AI Mirror (2024)
- Algorithmische Schmeichelei (Sycophancy) – anerkanntes technisches Konzept
- Der KI erlauben, zu widersprechen – Afthonios-Methode

QUELLEN & REFERENZEN

- Stanford – Studie zur Sycophancy von KI-Modellen (Science, 2026): Die Modelle geben dem Nutzer ~49 % häufiger recht als ein Mensch
- Shannon Vallor – The AI Mirror (2024), Professorin für KI-Ethik an der Universität Edinburgh
- Vectara, „Hallucination Leaderboard“ (Stand Mai 2026): Halluzinationsrate beim Zusammenfassen von Dokumenten von 1,8 % beim besten bis 24,2 % beim unzuverlässigsten Modell
- Dhuliawala et al., Findings of ACL (2024) – strukturierte Überprüfung und Zuverlässigkeit der Antworten der KI

Ein Modul von Afthonios.com.

Inhalt geprüft am 15.10.2026