Halderstone

Trainingsmodul

Grenzen & Fehlermodi von KI

Unsicherheit, Grenzen und typische Schwachstellen
in prädiktiven und generativen KI-Systemen

Woran können KI-Systeme scheitern?

Überblick

KI-Systeme arbeiten unter Unsicherheit. Ihr Verhalten verändert sich mit Daten, Kontext, Integrationsentscheidungen und menschlicher Interaktion. Wenn diese Grenzen nicht verstanden werden, vertrauen Organisationen KI entweder zu stark oder blockieren sie vorschnell, was Governance- und operationelle Risiken schafft.

Dieses Modul entwickelt ein klares, strukturiertes Bild der Grenzen und Fehlermodi von KI in prädiktiven und generativen Systemen. Die Teilnehmenden lernen, wo Unsicherheit entsteht, wie Fehlfunktionen in realen Umgebungen auftreten und wie sich KI-Ausgaben und technische Nachweise mit angemessenem Urteilsvermögen einordnen lassen.

Zielgruppe

- Verantwortliche für KI-Managementsysteme und Umsetzende, die mit technischen Teams zusammenarbeiten
- Fachleute für Governance, Risiko und Compliance, die KI-Fachwissen benötigen
- Produktverantwortliche und prozessverantwortliche Personen, die für KI-gestützte Dienstleistungen verantwortlich sind
- Auditfachleute, die ein gemeinsames Basisverständnis von KI-Systemen brauchen (nicht Audittechnik)
- Alle, die grundlegende Kenntnisse zu KI erwerben wollen

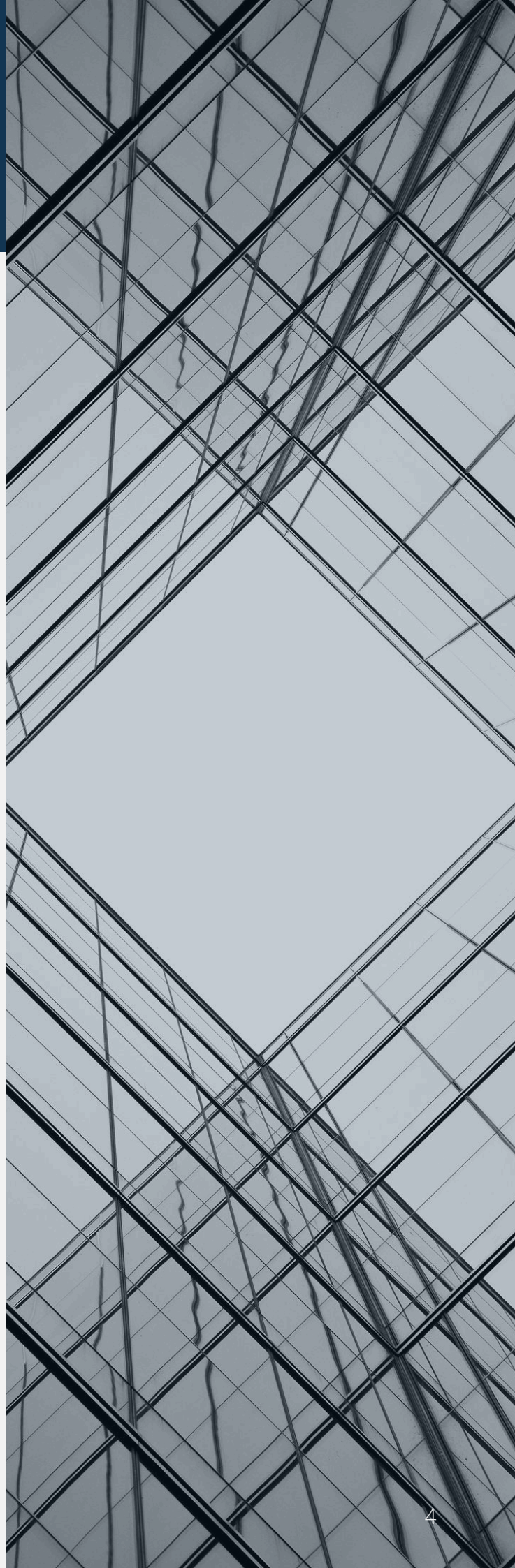
Ist dieses Modul für Sie das Richtige?

Es passt gut für Sie, wenn Sie...

- die Gründe verstehen wollen, weshalb KI-Ausgaben als fehlbare Signale zu behandeln sind und ihr Inhalt nicht automatisch faktisch korrekt ist.
- ein realistisches Denkmodell für Unsicherheit in prädiktiver und generativer KI brauchen.
- typische Fehlermodi von KI in realen operativen Kontexten erkennen wollen.
- Zuverlässigkeit, Grenzen und Vertrauen ohne falsche Gewissheit beurteilen müssen.
- das Verhalten von KI ohne Abstützung auf Anbieterbehauptungen oder Tools einordnen wollen.

Es passt möglicherweise weniger gut für Sie, wenn Sie...

- Methoden zur KI-Risikoanalyse oder zum Steuerungsdesign suchen.
- statistische Tiefenanalysen oder Optimierungs-techniken auf Modellebene erwarten.
- Umsetzungshilfen, Überwachungs-Setups oder Lebenszyklusprozesse wollen.
- bereits über fortgeschrittene Forschung in KI oder Data Science verfügen.



Lernergebnisse

Zentrale Lernergebnisse

- Hauptquellen von Unsicherheit in KI-Systemen und ihre Auswirkungen auf Ergebnisse erläutern
- Typische Fehlermodi in Modellen des prädiktiven maschinellen Lernens und in generativen KI-Systemen erkennen
- Soziotechnische Fehlermodi beschreiben, bei denen menschliche und organisatorische Faktoren mit KI zusammenwirken

Zusätzliche Fähigkeiten

- Probleme von Datenpipelines erkennen, die zu Modelldrift und Verzerrung beitragen
- Strukturierte Durchgänge zu Fehlermustern durchführen, um eine mögliche Verschlechterung der KI-Leistung früh zu erkennen
- Grenzen und Unsicherheit gegenüber Anspruchsgruppen so kommunizieren, dass ein verantwortungsvoller KI-Einsatz unterstützt wird

Agenda

Warum Grenzen und Unsicherheit zentral für die KI-Governance sind

Wie KI-Ausgaben als fehlbare probabilistische Signale zu behandeln sind, deren Inhalt nicht automatisch verlässlich oder faktisch korrekt ist, und wie Organisationen typischerweise scheitern, wenn sie KI-Systemen entweder zu stark vertrauen oder sie vorschnell blockieren

Woher Unsicherheit in KI-Systemen kommt

Wie Unsicherheit aus Daten, Labels und Kontextverschiebungen entsteht und wie sich Modellunsicherheit von Unsicherheit auf Systemebene durch Integration, Workflows und menschliche Faktoren unterscheidet

Grenzen des Modellverhaltens bei prädiktivem maschinellem Lernen

Wie prädiktive Modelle durch Grenzen der Generalisierung, Scheinkorrelationen, Verschiebungen in der Verteilung und Zielkonflikte bei der Leistung begrenzt sind, die von Schwellenwerten und dem Nutzungskontext abhängen

Grenzen des Modellverhaltens bei generativer KI

Wie generative Modelle Halluzinationen, Grenzen beim Befolgen von Anweisungen, Prompt-Sensitivität und Einschränkungen durch Kontextfenster und Wissensaktualisierungen zeigen

Datenbezogene Fehlermodi

Wie Fehlfunktionen aus Abweichungen zwischen Trainings- und Produktivdaten, Datenleckagen, Änderungen in Datenpipelines und Datenproblemen wie fehlenden Werten, Verzerrung und Herkunftsnachweisproblemen entstehen

Systemische und soziotechnische Fehlermodi

Wie KI-Systeme durch Automatisierungsbias, Fehlanwendung, Rückkopplungsschleifen, Gaming und fragile Schnittstellen versagen, die von Latenz, Integrationsgrenzen und Ausweichverhalten beeinflusst werden

Praxisworkshop

Anwendung der erlernten Konzepte, Methoden und Ansätze in einem realistischen Praxisfall

Enthaltene Unterlagen

Lernunterlagen

- Foliensatz
- Workbook für Teilnehmende

Vorlagen & Werkzeuge

- KI-Unsicherheitskarte (Quellen und beobachtbare Signale)
- Katalog für Fehlermodi (prädiktiv, generativ und auf Systemebene)
- Canvas für den Weg von Daten über Modell und Integration bis zur Nutzung
- Leitfragen für technische Walkthroughs

Bestätigung

- Teilnahmebestätigung

Vorbereitungshinweise

Vorausgesetzter Hintergrund

Dieses Modul setzt grundlegende Vertrautheit mit zentralen KI-Konzepten und Systemtypen voraus (Daten, Training im Vergleich zu Inferenz und gängige Architekturmustern von KI). Die Teilnehmenden sollten zudem in der Lage sein, technische Beschreibungen auf hoher Ebene zu lesen (Dienstleistungen, APIs, Datenspeicher).

Hilfreiche Vorkenntnisse sind:

- Grundverständnis digitaler Dienstleistungen und Abhängigkeiten (Anwendungen, Schnittstellen, Datenflüsse)
- Vertrautheit mit gängigen IT-Steuerungsmassnahmen (Zugriffsschutz, Protokollierung, Verschlüsselung) auf konzeptioneller Ebene

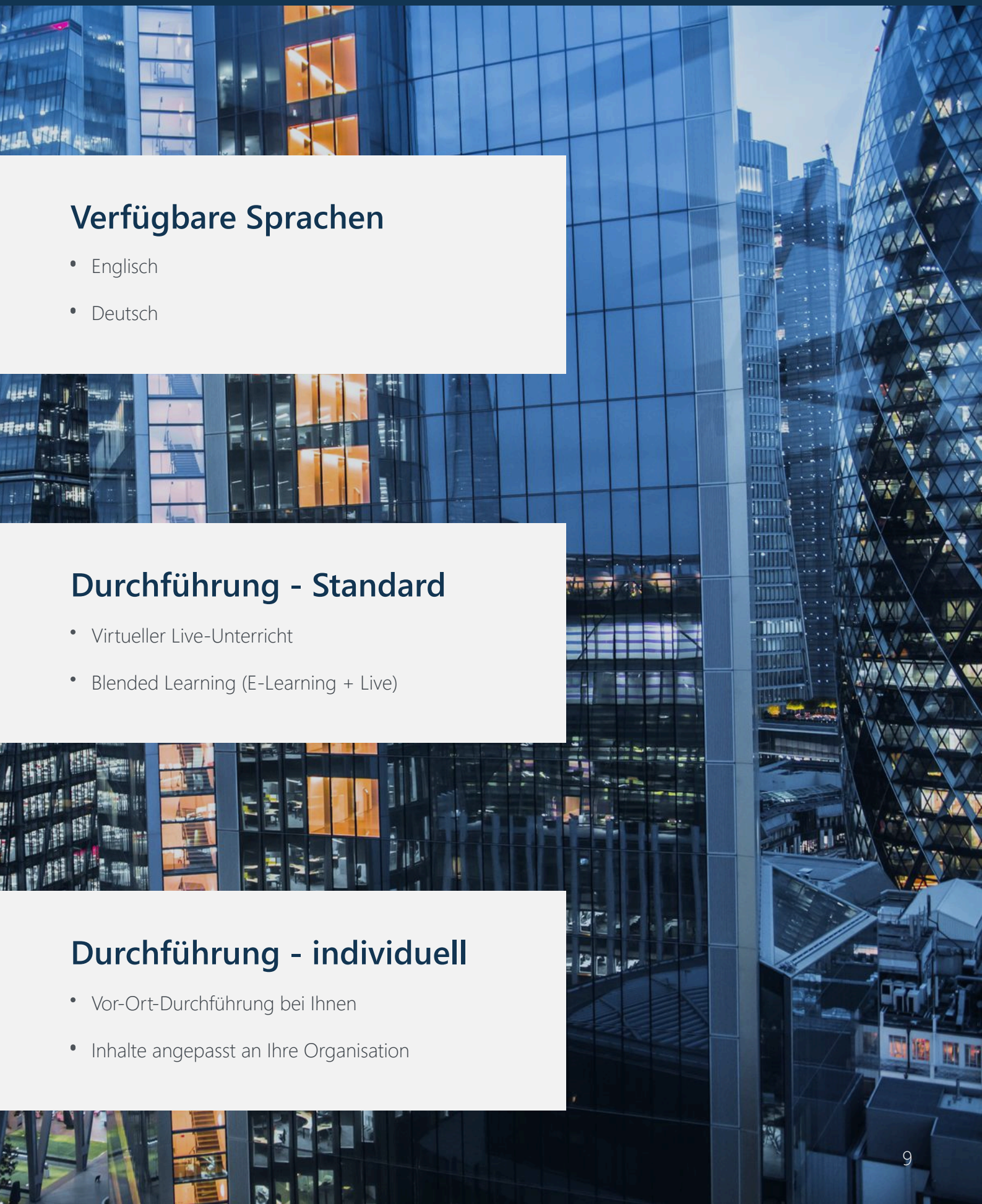
Vorbereitungsmodule

Unterstützend (optional)

Hilfreich, aber nicht erforderlich, um wirksam teilnehmen zu können

- KI-Systeme & Architekturen

Organisatorisches



Verfügbare Sprachen

- Englisch
- Deutsch

Durchführung - Standard

- Virtueller Live-Unterricht
- Blended Learning (E-Learning + Live)

Durchführung - individuell

- Vor-Ort-Durchführung bei Ihnen
- Inhalte angepasst an Ihre Organisation

Halderstone

Halderstone by Langer & Co
Zürcherstrasse 2
CH-8852 Altendorf
Schweiz
info@halderstone.com
www.halderstone.com