



FernUniversität  
in Hagen

# Experimentieren mit KI und Bildbeschreibung

Ein kurzer Einblick in einen Prototyp  
für die Erstellung von Alternativtexten

GLOBAL ACCESSIBILITY AWARENESS DAY 2026

Axel Nattland – Zentrum für Lernen und Innovation (ZLI)

## Worum geht es?

Warum ein solcher Prototyp für die Hochschule interessant ist

- Alternativtexte sind für blinde und sehbehinderte Studierende wichtig – ihre Erstellung braucht aber Zeit und Kontextwissen.
- Der Prototyp soll keine redaktionelle Arbeit ersetzen, sondern einen ersten Formulierungsvorschlag liefern.
- Interessant wird das besonders dort, wo viele Bilder anfallen: in Lehrmaterialien, Dokumenten oder Arbeitsabläufen.

## Im Mittelpunkt

- bessere erste Entwürfe
- schnellere Weiterarbeit
- lokale, datensensible Nutzung
- menschliche Qualitätskontrolle

**KI als Einstieg – nicht als letzte Instanz**

## Wie funktioniert der Prototyp?

Der Ablauf in wenigen Schritten

### 1. Bild auswählen

Einzelbild oder mehrere Dateien

### 2. Profil festlegen

z. B. knapper Alt-Text oder ausführlicher

### 3. Vorschlag erzeugen

Die KI erstellt einen ersten Entwurf

### 4. Prüfen und exportieren

Menschen prüfen, korrigieren und übernehmen

**Wichtig:** Die Ausgabe ist ein Vorschlag. Qualität, Kontext und Freigabe bleiben in menschlicher Verantwortung.

## Was der Prototyp schon leistet

Ein kurzer Blick auf die praktischen Stärken

- **Flexible Ausgabepprofile**  
Je nach Einsatz kann der Text knapper oder ausführlicher formuliert werden.
- **Lokale Verarbeitung**  
Die Ausführung kann lokal erfolgen – das ist für sensible Inhalte ein klarer Vorteil.
- **Einzelbild oder mehrere Bilder**  
Der Prototyp unterstützt nicht nur einzelne Bilder, sondern auch Serien oder Sammelverarbeitung.
- **Export für die Weiterarbeit**  
Ergebnisse lassen sich speichern, überarbeiten und in vorhandene Prozesse übernehmen.

## Was wir daraus mitnehmen

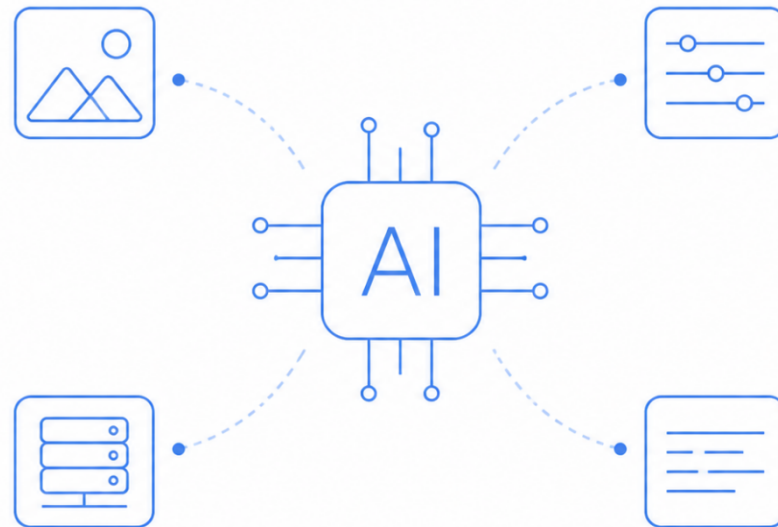
### Nutzen, Grenzen und offene Fragen

- KI kann den Einstieg in gute Bildbeschreibungen deutlich erleichtern.
- Der entscheidende Mehrwert entsteht dort, wo Vorschläge redaktionell geprüft und verbessert werden.
- Für die Hochschule interessant sind vor allem Qualität, Datenschutz und praktische Einbindung in Arbeitsabläufe.

### Offene Punkte

- Wie konsistent sind die Ergebnisse?
- Wie viel Nachbearbeitung ist nötig?
- Welche Profile sind im Alltag wirklich hilfreich?
- Wo braucht es zusätzliche redaktionelle Regeln?

# Tech-Info



## **KI-generierter Alternativtext:**

Das Bild zeigt ein zentrales Symbol für Künstliche Intelligenz (AI), das mit vier verschiedenen Icons verbunden ist: ein Bild, Einstellungen, eine Datenbank und Text.

## Eingesetztes Modell

### Pixtral-Large-Instruct-2411

- Multimodales LLM bzw. Vision-Language-Model (Bild + Text)
- ca. 124 Milliarden Parameter
- Basis: Mistral Large 2 + Vision Encoder (~1B Parameter) + Cross-Attention Layers
- EXL2 Quantisierung (4.5bpw) für effizienteren Speicherverbrauch
- Lokaler Betrieb (datenschutzkonform)
- Mistral Research License (MRL)

## Eigene Anpassungen

- Web-App für Upload, Auswahl, Prüfung und Export
- Ausgabeprofile für unterschiedliche Beschreibungstypen
- zusätzliche Kontextübergabe aus der Anwendung
- optionale Spezialinstruktionen, z. B. Formel-Verbalisierung
- Einzelbild- und Bulk-Verarbeitung
- Export für die redaktionelle Weiterarbeit
- Queue-, Status- und Download-Logik

## Prompting für barrierefreie Bildbeschreibung

- profilgesteuerte Prompt-Bausteine
- System- und User-Prompt werden aus Profilen geladen
- zusätzlicher Kontext wird im User-Teil ergänzt
- optionale Zusatzinstruktionen, z. B.  
*Personen ignorieren* oder *Formeln verbalisieren*
- Spezialisierung erfolgt über Prompting, nicht über Modelltraining

## Inference & Ausführung

### ExLlamaV2

- Optimierte Inference Engine für große KI-Modelle
- Hohe Performance auf NVIDIA GPUs
- Multi-GPU Unterstützung

<https://github.com/turboderp-org/exllamav2>

## Hardware

### 2 × NVIDIA L40

- 48 GB VRAM pro GPU
- 96 GB Gesamt-GPU-Speicher

### Aktuelle Auslastung bei EXL2 Quantisierung (4.5bpw)

- GPU 0: ~45 GB VRAM
- GPU 1: ~35 GB VRAM
- Gesamt: ~80 GB VRAM

**Vielen Dank für Ihre Aufmerksamkeit!**