

Taxonomische Klassifikation kognitiver Komplexität bei KI-generierten Multiple-Choice-Fragen

Zielsetzung

Dieses Dokument untersucht, inwieweit kognitive Anspruchsniveaus auf Grundlage der Bloom'schen Taxonomie durch generative Sprachmodelle – exemplarisch GPT-4.5/5.1 – bei der automatisierten Erstellung von Multiple-Choice-Fragen (MC-Fragen) zuverlässig operationalisierbar sind. Die verbreitete Kategorisierung nach subjektivem Schwierigkeitsgrad ("leicht/mittel/schwer") wird dabei zugunsten einer instruktionsbasierten, kognitiv fundierten Klassifikation ersetzt. Ziel ist die Entwicklung eines differenzierten, pädagogisch validen Stufenmodells, das eine konsistente Annotation, adaptive Itemauswahl sowie lernzielbezogene Feedbackmechanismen in digitalen Assessmentsystemen ermöglicht.

1. Theoretischer Rahmen: Bloom-Taxonomie und KI-gestützte Fragengenerierung

Die ursprüngliche Taxonomie von Bloom et al. (1956) unterscheidet sechs hierarchisch angeordnete kognitive Prozessebenen:

1. Erinnern (Remember)
2. Verstehen (Understand)
3. Anwenden (Apply)
4. Analysieren (Analyze)
5. Bewerten (Evaluate)
6. Erschaffen (Create)

In der revidierten Version (Anderson & Krathwohl, 2001) wurden diese Prozesse durch aktionsorientierte Verben ersetzt und um eine Wissensdimension ergänzt (factual, conceptual, procedural, metacognitive knowledge). Die Taxonomie hat sich als Werkzeug zur Lernzieldefinition, Aufgabenentwicklung und Curriculumsanalyse etabliert. Im Kontext KI-gestützter MC-Fragengenerierung stellt sich jedoch die Frage, welche dieser Stufen durch ein sprachbasiertes Modell, das keine Weltverankerung besitzt, zuverlässig abgedeckt werden können.

Zentrale Befunde:

- **Stufen 1–3** (Reproduktion, Anwendung) sind zuverlässig generierbar, da sie auf explizit strukturiertem Wissen, definitorischen Begriffssystemen und regelgeleiteten Handlungsmustern beruhen.
- **Stufe 4** (strukturierte Analyse) ist in begrenztem Umfang abbildbar, sofern die Aufgabenstellung eindeutig und fehlerresistent formuliert ist.
- **Stufen 5–6** (Bewertung, Kreation) übersteigen die semantischen und heuristischen Fähigkeiten aktueller Sprachmodelle, insbesondere im geschlossenen MC-Format.

Es ist zu betonen, dass das hier entwickelte Modell nicht den Lösungsaufwand (difficulty), sondern den **inhaltlich-kognitiven Anforderungsgrad** der Frage adressiert. Schwierigkeit und Komplexität sind orthogonale Dimensionen.

2. Operationalisierbare Kognitionsstufen in der Praxis

Nach Analyse linguistischer Operatoren, Aufgabenstrukturen und empirischer Validität lassen sich drei Stufen als für KI geeignet identifizieren.

Stufe 1: Reproduktion (Recall)

- **Deutsch:** Erinnern, Faktenwissen, Wiedergeben
- **Englisch:** Recall, Remember, Factual knowledge
- **Merkmale:**
 - Reproduktion deklarativen Wissens ohne Transformation
 - Abfrage definitorischer Inhalte, Vokabeln, Normwerte, Gesetze
- **Beispielformate:**
 - „Welche Aussage beschreibt korrekt das Prinzip der Osmose?“
 - „Was ist die SI-Einheit der elektrischen Spannung?“

Diese Items sind deterministisch generierbar und durch große Trainingsmengen gestützt. Fehler treten selten auf.

Stufe 2: Anwendung (Application)

- **Deutsch:** Umsetzen, Ausführen, regelgeleitete Umsetzung
- **Englisch:** Apply, Use, Execute
- **Merkmale:**
 - Übertragung bekannter Verfahren auf neue, aber strukturverwandte Kontexte
 - Nutzung von Formeln, Algorithmen, heuristischen Regeln
- **Beispielformate:**
 - „Welche Berechnung ist geeignet, um den Stromfluss in Schaltung X zu ermitteln?“
 - „Welche didaktische Maßnahme passt zu der geschilderten Situation?“

Solche Items sind automatisiert generierbar, solange Problemstellung und Lösungspfad eindeutig sind. Kontextfreie Abstrahierung gelingt zuverlässig.

Stufe 3: Strukturierte Analyse (Structured Analysis)

- **Deutsch:** Vergleichen, Zuordnen, Schlussfolgern
- **Englisch:** Analyze, Infer, Differentiate, Classify
- **Merkmale:**
 - Zerlegung von Sachverhalten in funktionale Komponenten
 - Identifikation von Argumentstruktur, Fehlern oder logischen Beziehungen
- **Beispielformate:**
 - „Welche Aussage widerspricht der Argumentation in Absatz A?“
 - „Welche Hypothese lässt sich aus dem dargestellten Befund ableiten?“

Die Generierung ist möglich, wenn die Struktur der Aufgabe vorgegeben ist. Bei offenen Problemstellungen steigt das Fehlerrisiko durch Ambiguität oder inadäquate Distraktorlogik.

3. Nicht operationalisierbare Stufen

Stufe 4 (komplexe Analyse)

Erfordert interkontextuelle Verknüpfungen, metakognitive Bewertungen und semantisch validierte Argumentationslinien. Sprachmodelle generieren zwar formal plausible Inhalte, jedoch ohne tatsächliche Verstehensgrundlage. Dadurch entstehen inkonsistente Distraktoren und unplausible Schlussfolgerungen.

Stufe 5 (Bewerten)

Die Abwägung konkurrierender Lösungen nach Kriterien wie Ethik, Effizienz oder Nachhaltigkeit setzt Normverständnis, Gewichtung und Kontextwissen voraus – Eigenschaften, die LLMs nicht besitzen. Bewertungsfragen wirken daher oft arbiträr oder suggerieren Objektivität, wo keine existiert.

Stufe 6 (Kreation)

Die Entwicklung originärer Ideen, Konzepte oder Synthesen ist inhärent offen und nicht mit geschlossenen Antwortformaten vereinbar. Zwar kann die KI Lösungsvorschläge generieren, doch MC-Formate lassen keine kreative Variation zu. Eine automatische Bewertung kreativer Qualität ist methodisch nicht belastbar.

4. Schlussfolgerung und didaktisch-technische Integration

Sprachmodelle sind in der Lage, Multiple-Choice-Fragen mit unterschiedlichen kognitiven Anspruchsniveaus **bis einschließlich strukturierter Analyse** zuverlässig zu generieren. Damit liegt ein praktikabler Handlungsrahmen für die Entwicklung kognitiv differenzierter Fragenformate vor.

Empfehlungen zur Systemintegration:

- Etablierung der drei Stufen (Reproduktion, Anwendung, Analyse) als standardisierte **cognitive_level**-Kategorien im Datenmodell
- Didaktisch orientierte Autorenunterstützung durch Glossare, Operatorlisten und Beispiele
- Optionale heuristische Klassifikationsvorschläge auf Basis von Promptstruktur und Formulierungen
- Expliziter Ausschluss automatisierter Generierung für Bloom-Stufen 5 und 6; ggf. manuelle Annotation möglich

Die vorgeschlagene Taxonomie ersetzt die unscharfe Schwierigkeitsklassifikation durch eine kognitiv valide Beschreibung der Verarbeitungsanforderung. Sie ermöglicht sowohl differenzierte Lernanalytik als auch personalisierte Feedbackpfade entlang nachweisbarer Denkopoperationen. Langfristig stellt sie eine Grundlage für adaptive Aufgabensteuerung, Kompetenzprofiling und metakognitiv anschlussfähige Lernarchitekturen dar.