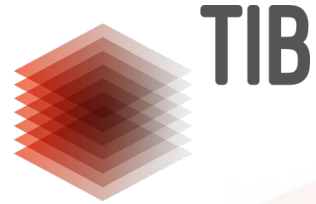


---

LEIBNIZ-INFORMATIONSZENTRUM  
TECHNIK UND NATURWISSENSCHAFTEN  
UNIVERSITÄTSBIBLIOTHEK



# Automatische Generierung von Fragen für Lernvideos

Markos Stamatakis  
Hannover, 08.11.2024

# Motivation

## Lernvideos

- Unterstützung der Lernenden mittels visuellen und auditiven Inhalten (multimodal)
- Zeit- und Ortsunabhängig [Merkt, 2022]
- Bessere Lernerfahrung [Palaiageorgiou, 2019; Leisner, 2020]

## Automatische Fragengenerierung

- Manuelle Fragenerstellung sehr zeitaufwändig
- Überprüfung von Wissen
  - Verständnisfragen
  - Vorbereitung auf Klausuren/Tests
  - Personalisiertes Lernen
- Finden von geeigneten Fakten
  - Automatische Extraktion der wichtigsten Inhalte
  - Didaktisch wertvolle Fragen

## Mögliche Fragenformate

### Multiple-Choice Fragen

- Antworten bereits gegeben (Auswahl reicht aus)
- Sinnvolle Distraktoren müssen definiert werden

Beispiel:

Welches der folgenden Moleküle wird während der Photosynthese als Nebenprodukt freigesetzt?

A) Stickstoff ( $N_2$ ), B) Wasserstoff ( $H_2$ ), C) Sauerstoff ( $O_2$ ), D) Methan ( $CH_4$ )

### Offene Fragen

- Antwort nicht gegeben
- Beantwortung durch Fließtext
- Komplexere Auswertung

Beispiel:

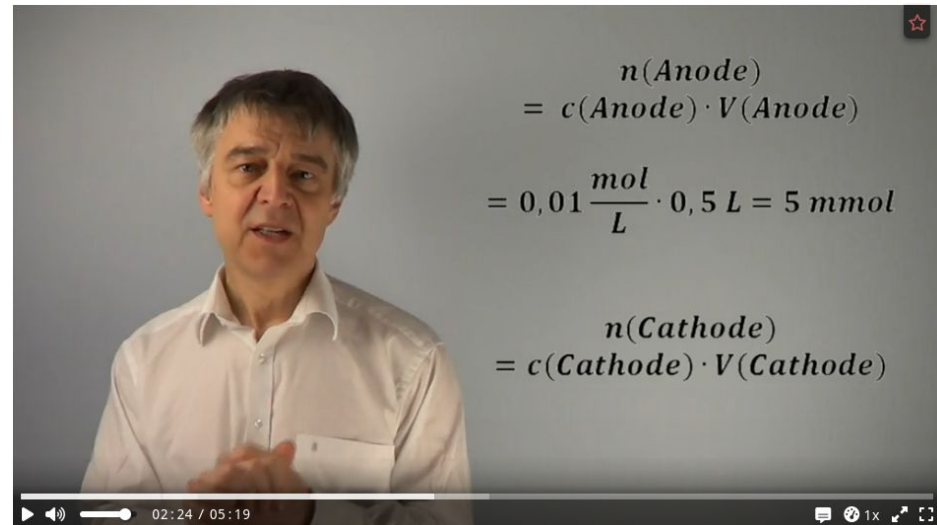
Wie verläuft der Prozess der Photosynthese, und welche Rolle spielt das Chlorophyll dabei?

## Varianten von Lernvideos

- Videos können auf verschiedene Art und Weise dargestellt werden [Shoufan, 2017]
- Videos können auf verschiedenen Sprachen sein

### Mögliche Videoarten

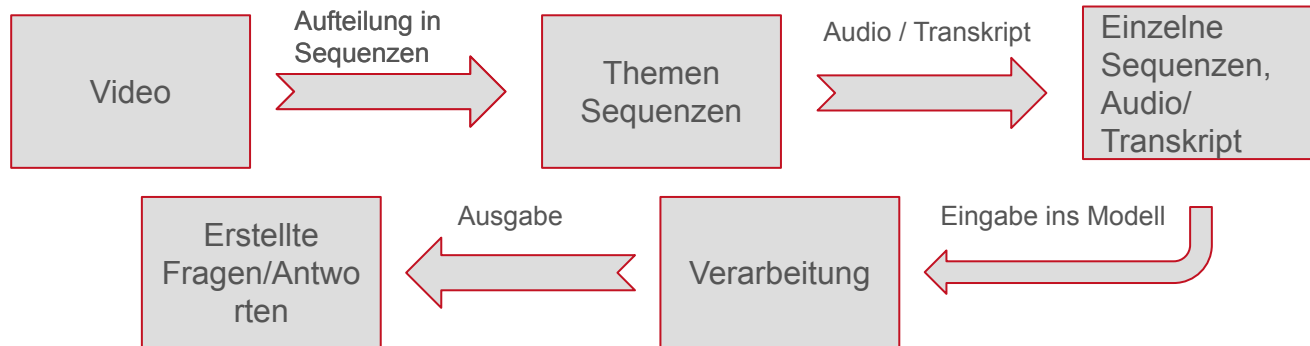
- Powerpoint-Präsentation
- Khan-Style
- Darstellung des Sprechers
- Verwendung eines Whiteboards
- Klassenraumvideos
- Übertragung des Bildschirms
- Animation

A video player interface showing a lecture. On the left, a man in a white shirt is speaking. On the right, a whiteboard displays chemical equations. The equations are: 
$$n(\text{Anode}) = c(\text{Anode}) \cdot V(\text{Anode})$$
$$= 0,01 \frac{\text{mol}}{\text{L}} \cdot 0,5 \text{ L} = 5 \text{ mmol}$$
$$n(\text{Cathode}) = c(\text{Cathode}) \cdot V(\text{Cathode})$$
The video player controls at the bottom show a play button, a volume icon, a progress bar at 02:24 / 05:19, and icons for chat, settings, and full screen.

Example video from the TIB AV-Portal

## Forschungsziele

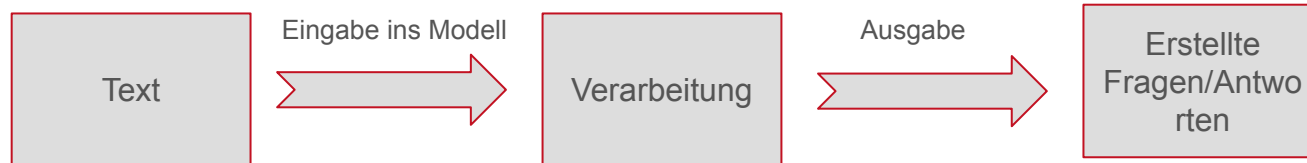
- Multimodale Analyse der Videos (kaum erforscht [Das, 2021])
- Generierung von Multiple-Choice-Fragen zu Videoinhalten
- Generierung von Fragen für Freitextantworten zu Videoinhalten
- Generierung dazugehöriger Antworten
- Integration in einem Echtzeit-System (z.B. TIB-AV-Portal)



## Aktueller Fortschritt

### Evaluation von Large-Language-Modellen (LLMs)

- SQuAD Datensatz [Rajpurkar, 2018] als Basis
  - Über 100.000 Frage-Antwort-Paare für über 500 Wikipedia-Artikel
  - Fokus auf antwortunabhängiges Trainieren
  - Zusätzlich Sprachtranskripte von Videos aus TIB-AV-Portal
- Automatische und manuelle Evaluation
  - Textqualität, Fragenqualität und Kontext manuell evaluiert
  - Einordnung in Problemklassen manuell/automatisch
  - ROUGE-L [Lin, 2004] als Metrik für den Vergleich zum Ground-Truth
- Qualitative Fragengenerierung möglich



## Aktueller Fortschritt

### Evaluation von Vision-Language-Modellen (VLMs)

- LearningQ Datensatz [Chen, 2018] als Basis (Zero-Shot, Fine-Tune)
- Verwendung von VLMs mit unterschiedlichen Architekturen
  - Multimodal (Frames + Audio/Sprachtranskript)
  - Monomodal (Frames)
- LLM als Zero-Shot-Baseline
- Automatische und manuelle Evaluation
  - ROUGE-L [Lin, 2004], BERT-Score [Zhang, 2020], eigene Metrik
  - Relevanz, Abdeckung und Blooms Taxonomie [Conklin, 2005] manuell evaluiert
- Fragengenerierung benötigt Verbesserungen
  - Fragen passen teilweise nicht zum Inhalt
  - Level der Blooms Taxonomie nicht komplett abgedeckt
  - Large-Language-Modell im Zero-Shot präziser

# Ergebnisse

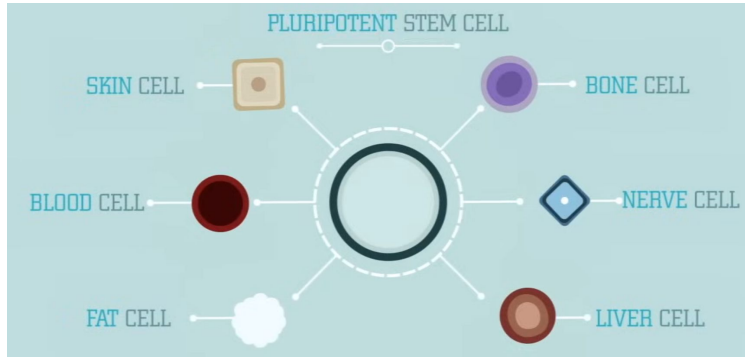


- Verteilung der Fragewörter bei Fragen in %
- M-7B: Mistral-7B [Jiang, 2023], V-LM: Video-LLaMA [Zhang, 2023], GT: Ground-Truth, \*: Fine-Tune
- Keine gleichmäßige Verteilung vorhanden

Modell	Where	Who	When	What	Why	Whose	Whom	Which	How	None
M-7B	16.42	2.17	14.89	44.37	4.17	0.09	0.00	4.80	11.08	2.01
V-LM	3.04	1.49	1.31	41.91	0.59	0.00	0.00	5.32	25.11	21.24
V-LM*	2.41	0.97	3.72	42.56	3.91	0.01	0.00	2.31	28.47	15.63
V-LM* (Nur visuell)	1.92	1.07	4.10	45.67	3.22	0.01	0.00	2.31	30.86	10.84
V-LM* (Nur Audio)	1.58	1.17	3.43	40.95	3.22	0.00	0.00	3.30	32.92	13.43
GT	2.62	0.88	7.67	19.90	13.58	0.05	0.01	2.25	17.29	35.76



## Beispiele generierter Fragen



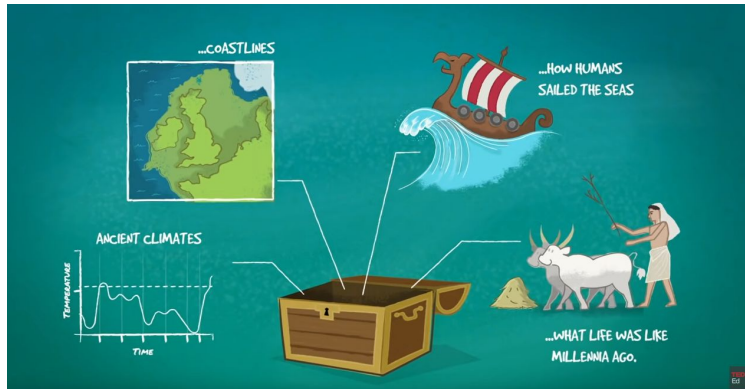
What are stem cells? - Craig A. Kohn - <https://www.youtube.com/watch?v=evH017Coc54>

**Prompt:** Create a question about the video content.

**Answer:** So do platelets only form when the platelet clot forms in the arteries?

**Prompt:** Develop a question that tests comprehension of the video's main idea.

**Answer:** Which of these does the video say are the two stages of mitosis?



How much of human history is on the bottom of the ocean? - Peter Campbell - <https://www.youtube.com/watch?v=kSBB5PsRV-k>

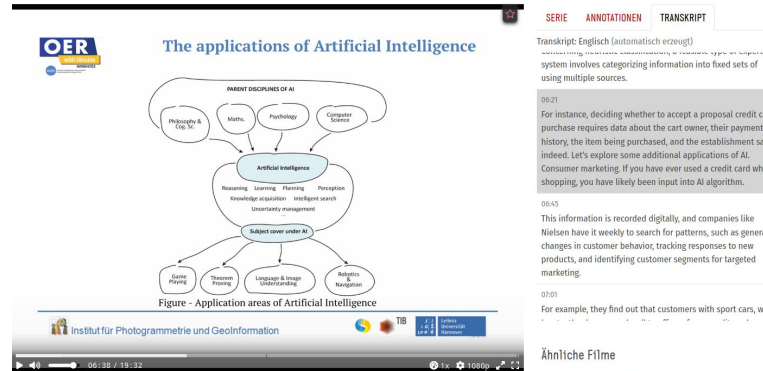
**Prompt:** Create a question about the video content.

**Answer:** Why did the ancient Phoenicians have to know so much about the seas?

**Prompt:** Develop a question that tests comprehension of the video's main idea.

**Answer:** what are the differences and similarities between hinduism and other religions?

- Erstellung eines eigenen Datensatzes
  - Mehr Daten erforderlich für hochwertige Fragen (multimodal)
  - TIB-AV-Portal als Basis
- Untersuchung verschiedener Arten von Videorepräsentationen
  - Modell kann unterschiedlich beeinflusst werden
  - Unterschiedlicher Fokus auf Modalitäten
- Verwendung von Large-Language- und Vision-Language-Modelle als Evaluation
- Generierung von Antworten und Distraktoren



The screenshot shows a video player interface. The main content is a slide titled "The applications of Artificial Intelligence" from OER. The slide features a diagram with "Artificial Intelligence" at the center, surrounded by "Parent Disciplines of AI" (Philosophy & Cog. Sc., Maths, Psychology, Computer Science) and "Subject cover under AI" (Game Playing, Telecom. Pricing, Logistics & Image Optimization, Robotics & Navigation). Below the diagram is the caption "Figure - Application areas of Artificial Intelligence". To the right of the slide is a transcript in German, with the first line indicating it was automatically generated. The transcript discusses information categorization and provides an example of a credit card purchase decision.

**OER** **The applications of Artificial Intelligence**

**PARENT DISCIPLINES OF AI**

- Philosophy & Cog. Sc.
- Maths
- Psychology
- Computer Science

**Artificial Intelligence**

Reasoning Learning Planning Perception  
Knowledge acquisition Intelligent search  
Uncertainty management

**Subject cover under AI**

- Game Playing
- Telecom. Pricing
- Logistics & Image Optimization
- Robotics & Navigation

Figure - Application areas of Artificial Intelligence

**Transkript: Englisch (automatisch erzeugt)**

... system involves categorizing information into fixed sets of using multiple sources.

06:31

For instance, deciding whether to accept a proposal credit card purchase requires data about the cart owner, their payment history, the item being purchased, and the establishment sale indeed. Let's explore some additional applications of AI. Consumer marketing. If you have ever used a credit card while shopping, you have likely been input into AI algorithm.

06:45

This information is recorded digitally, and companies like Nielsen have it weekly to search for patterns, such as general changes in customer behavior, tracking responses to new products, and identifying customer segments for targeted marketing.

07:01

For example, they find out that customers with sport cars, who

Ähnliche Filme

Example video from the TIB AV-Portal

# Referenzen



[Merk, 2022]: Merkt, M., Hoppe, A., Bruns, G., Ewerth, R., and Huff, M. (2022). Pushing the button: Why do learners pause online videos? *Comput. Educ.*, 176:104355.

[Leisner, 2020]: Leisner, D., Zahn, C. G., Ruf, A., and Cattaneo, A. A. P. (2020). Different ways of interacting with videos during learning in secondary physics lessons. In *HCI International 2020 - Posters - 22nd International Conference, HCII 2020, Copenhagen, Denmark, 2020, Proceedings, Part II, volume 1225 of Communications in Computer and Information Science*, pp. 284–291. Springer.

[Palaigeorgiou, 2019]: Palaigeorgiou, G. and Papadopoulou, A. (2019). Promoting self-paced learning in the elementary classroom with interactive video, an online course platform and tablets. *Educ. Inf. Technol.*, 24(1):805–823.

[Shoufan, 2017]: Shoufan, Abdulhadi & Mohamed, Fatma. (2017). On the Likes and Dislikes of YouTube's Educational Videos.

[Das, 2021]: Das, B., Majumder, M., Phadikar, S., and Sekh, A. A. (2021). Automatic question generation and answer assessment: a survey. *Research and Practice in Technology Enhanced Learning*, 16(1):5.

[Chen, 2018]: Guanliang Chen, Jie Yang, Claudia Hauff, and Geert-Jan Houben. 2018. LearningQ: A Large-Scale Dataset for Educational Question Generation. In *International Conference on Web and Social Media, ICWSM 2018, Stanford, (California), USA, June 25-28, 2018. AAAI Press, Stanford, (California), USA, 481–490. <https://doi.org/10.1609/icwsm.v12i1.14987>*

[Lin, 2004]: Lin, C.Y.: ROUGE: A package for automatic evaluation of summaries. In: *Text Summarization Branches Out*. pp. 74–81. Association for Computational Linguistics, Barcelona, Spain (2004)

[Zhang, 2020]: Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In *International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net, Addis Ababa, Ethiopia. <https://openreview.net/forum?id=SkeHuCVFDt>*

[Conklin, 2005]: Jack Conklin. 2005. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives complete edition. *Educational Horizons* 83, 3 (2005), 154–159. <http://www.istor.org/stable/42926529>

[Rajpurkar, 2018]: Pranav Rajpurkar, Robin Jia, & Percy Liang (2018). Know What You Don't Know: Unanswerable Questions for SQuAD. *CoRR*, abs/1806.03822.

[Zhang, 2023]: Hang Zhang, Xin Li, Lidong Bing: Video-LLaMA: An Instruction-tuned Audio-Visual Language Model for Video Understanding. *EMNLP (Demos) 2023*: 543-553

[Jiang, 2023]: Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, L  lio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timoth  e Lacroix, William El Sayed: Mistral 7B. *CoRR* abs/2310.06825 (2023)