Towards AI-Assisted Cognitive Load Estimation In Consumer-Grade VR

von Vincent François

Betreut von:

Prof. Florian Heinrichs
Fachbereich 9 Medizintechnik
FH Aachen

Patrick Querl
Lehr- und Forschungsgebiet Ingenieurhydrologie
RWTH Aachen

Hydrology
Lehr- und
Forschungsgebiet
Ingenieurhydrologie

RWTH AACHEN UNIVERSITY

Gliederung

- **Motivation**
  - Avatar basiertes Lernen und Lehren
  - Cognitive Load
  - Cognitive Load Messung
- **Forschungsfragen**
- **Methoden**
  - Systematische Literaturrescherche
  - Textanalyse (Feaures)
- **Ergebnisse (Features)**
- **Methoden**
  - Textanalyse (Machine Learning Modelle)
  - Elo Ratings
- **Ergebnisse (Machine Learning Modelle)**
  - Architektur
  - Elo
- **Wozu** Cognitive Load messen?

Motivation – Avatar basiertes Lehren und Lernen am LFI

Avatar basiertes Lehren und Lernen

Unity blender

> Another relevant perspective to consider is the hype factor often associated with educational VR. In general, emerging technologies are said to progress through different levels of hype

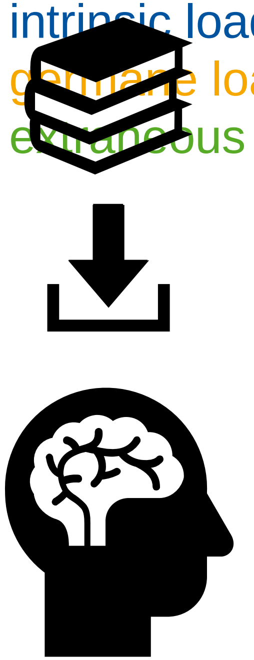
Makransky & Petersen, 2021

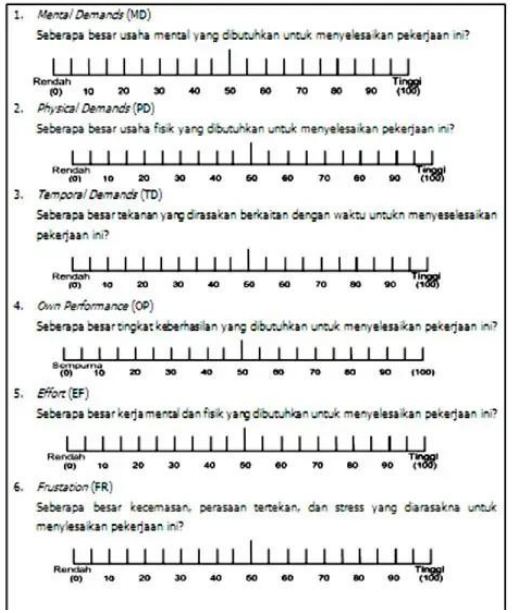
https://myscore.lfi.rwth-aachen.de/de/

Motivation – Cognitive Load

John Sweller

intrinsic load
germane load
extraneous load





1. *Mental Demands* (MD)
Seberapa besar usaha mental yang dibutuhkan untuk menyelesaikan pekerjaan ini?
Rendah (0) 10 20 30 40 50 60 70 80 90 Tinggi (100)
2. *Physical Demands* (PD)
Seberapa besar usaha fisik yang dibutuhkan untuk menyelesaikan pekerjaan ini?
Rendah (0) 10 20 30 40 50 60 70 80 90 Tinggi (100)
3. *Temporal Demands* (TD)
Seberapa besar tekanan yang dirasakan berkaitan dengan waktu untukn menyeselesaikan pekerjaan ini?
Rendah (0) 10 20 30 40 50 60 70 80 90 Tinggi (100)
4. *Own Performance* (OP)
Seberapa besar tingkat keberhasilan yang dibutuhkan untuk menyelesaikan pekerjaan ini?
Sempurna (0) 10 20 30 40 50 60 70 80 90 (100)
5. *Effort* (EF)
Seberapa besar kerja mental dan fisik yang dibutuhkan untuk menyelesaikan pekerjaan ini?
Rendah (0) 10 20 30 40 50 60 70 80 90 Tinggi (100)
6. *Frustation* (FR)
Seberapa besar kecemasan, perasaan tertekan, dan stress yang diarasakna untuk menylesaikan pekerjaan ini?
Rendah (0) 10 20 30 40 50 60 70 80 90 Tinggi (100)

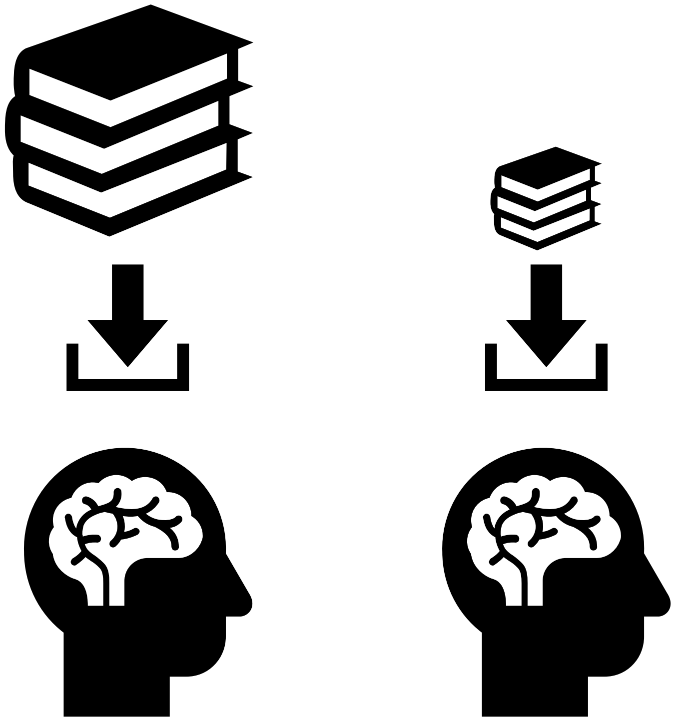
Lernpsychologie
Belastung des Arbeitsgedächtnisses

Motivation – Cognitive Load Messung

Fragebögen

Leistungsmaße
- Ergebnis
- Zeit

Physiologische Maße
- Herzratenvariabilität
- Eye Tracking
- EEG
- …



Motivation – Cognitive Load Messung

Fragebögen

Leistungsmaße
- Ergebnis
- Zeit

Physiologische Maße
- Herzratenvariabilität
- Eye Tracking
- EEG
- …

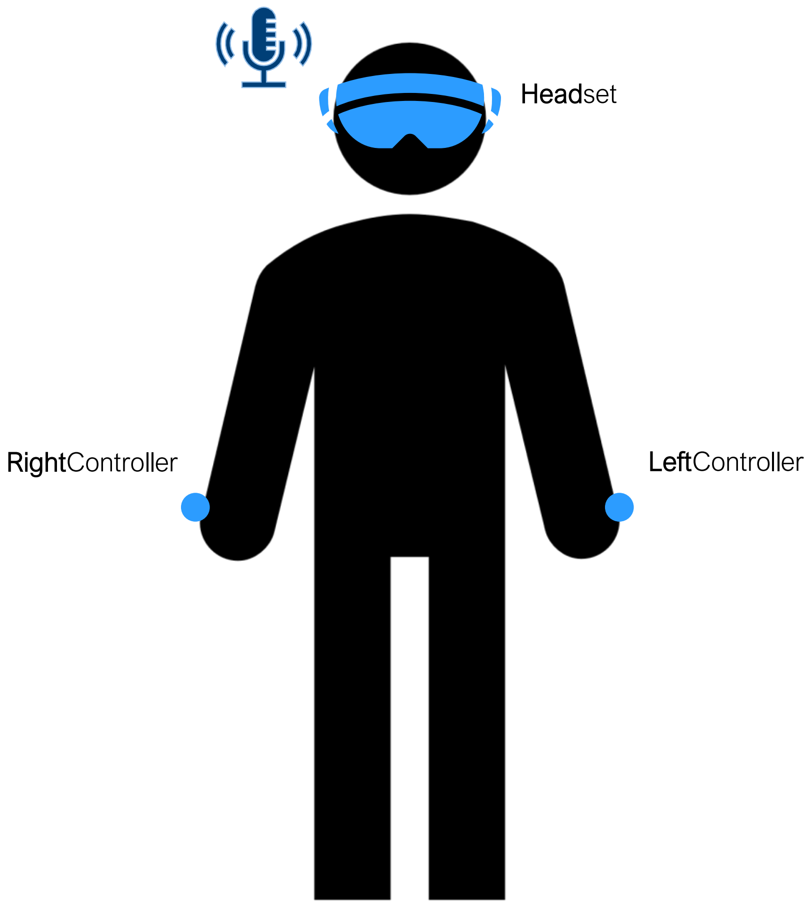
?

Motivation – Cognitive Load Messung in VR

Fragebögen

Leistungsmaße
- Ergebnis
- Zeit

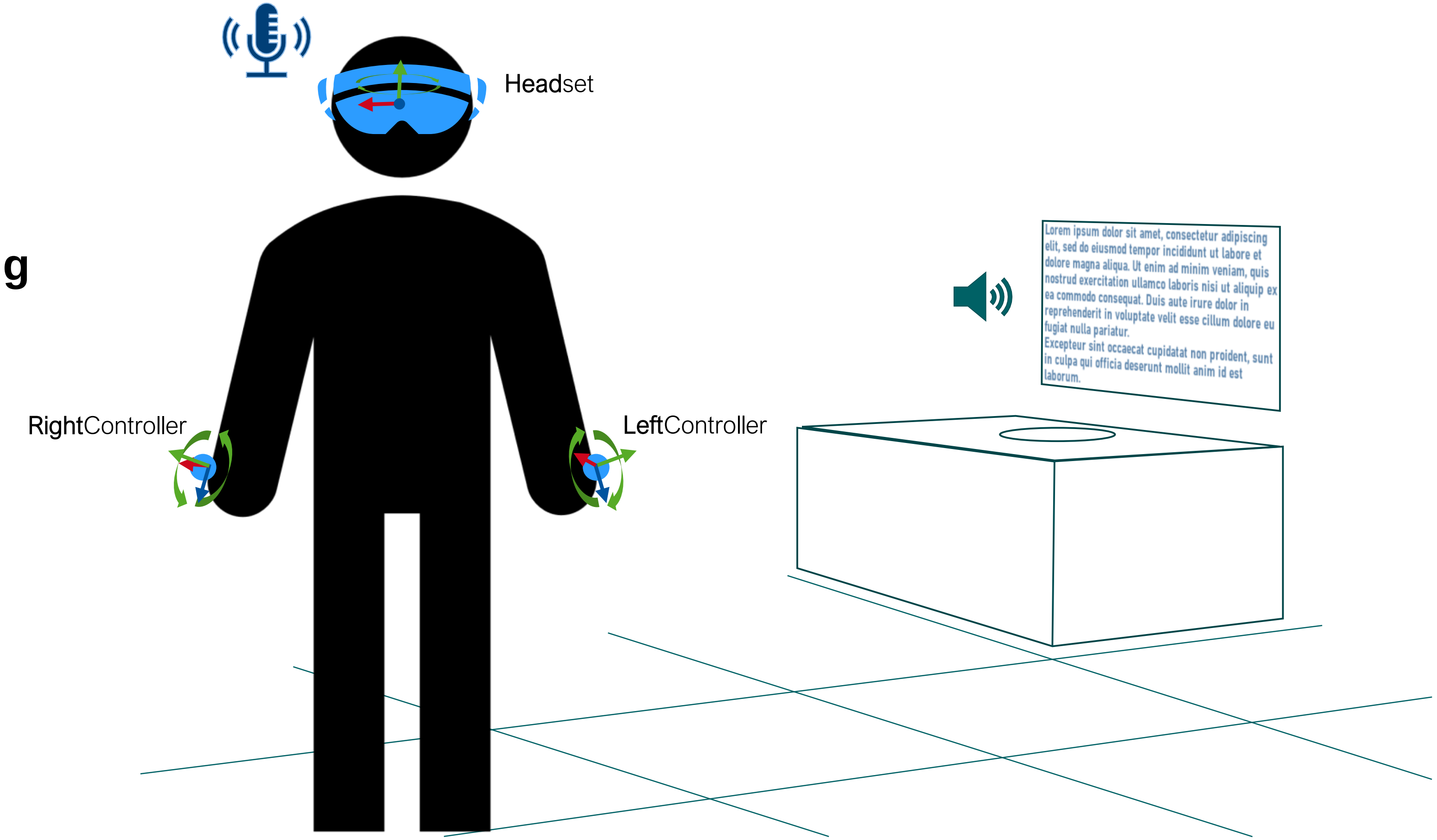
Physiologische Maße
- Herzratenvariabilität
- Eye Tracking
- EEG
- …



Motivation – Cognitive Load Messung in VR

Nutzerverhalten
- Bewegung
- Audio
- Controller Inputs

Virtuelle Umgebung
- Kamera
- Audio
- 3D Umgebung

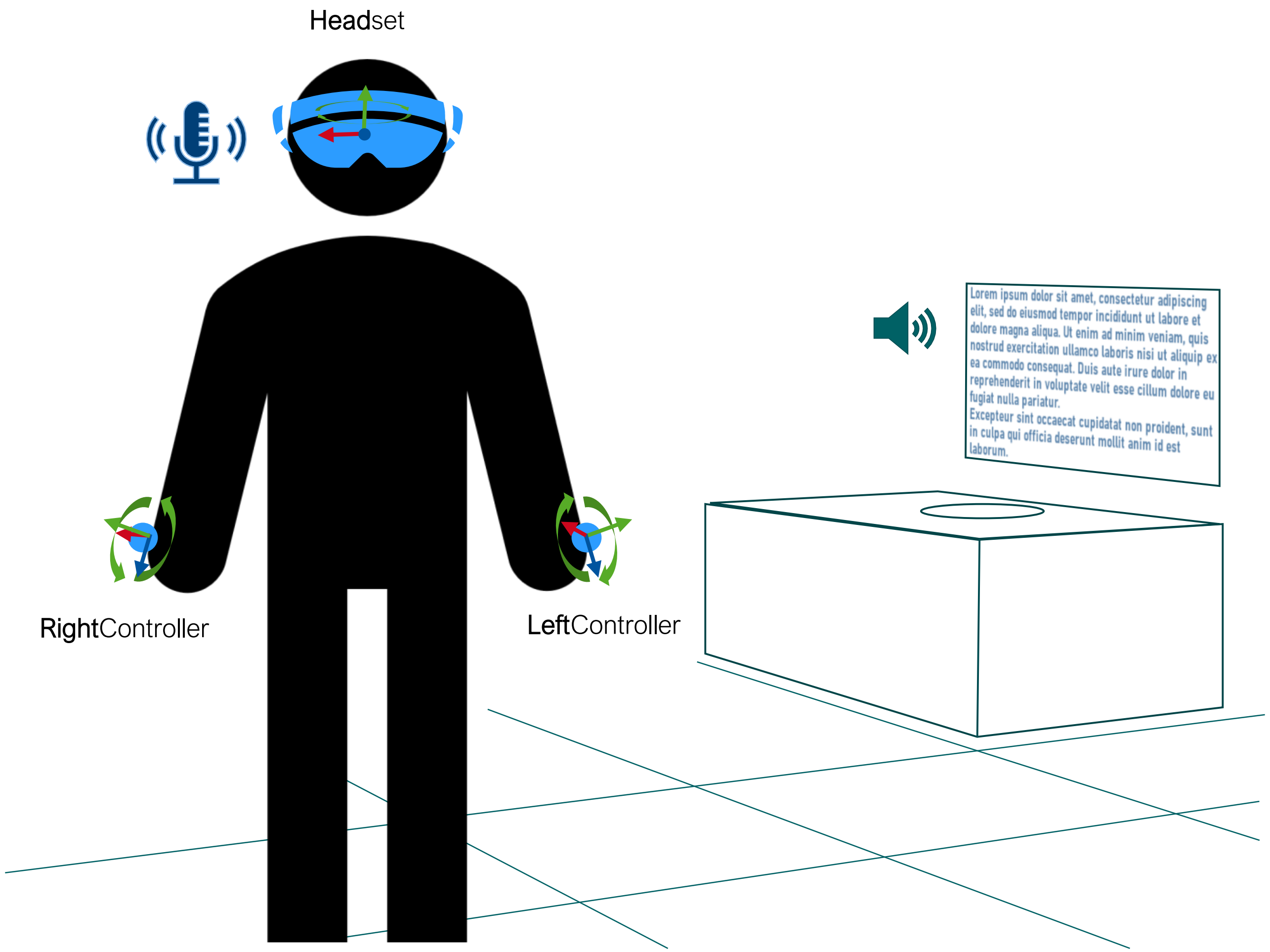


Motivation – Cognitive Load Messung mit Machine Learning in VR

Nutzerverhalten
- Bewegung
- Audio
- Controller Inputs

Virtuelle Umgebung
- Kamera
- Audio
- 3D Umgebung

Machine Learning
- Viele Daten
- Komplexer Zusammenhang



Forschungsfragen

Nutzerverhalten
- Bewegung
- Audio
- Controller Inputs

Virtuelle Umgebung
- Kamera
- Audio
- 3D Umgebung

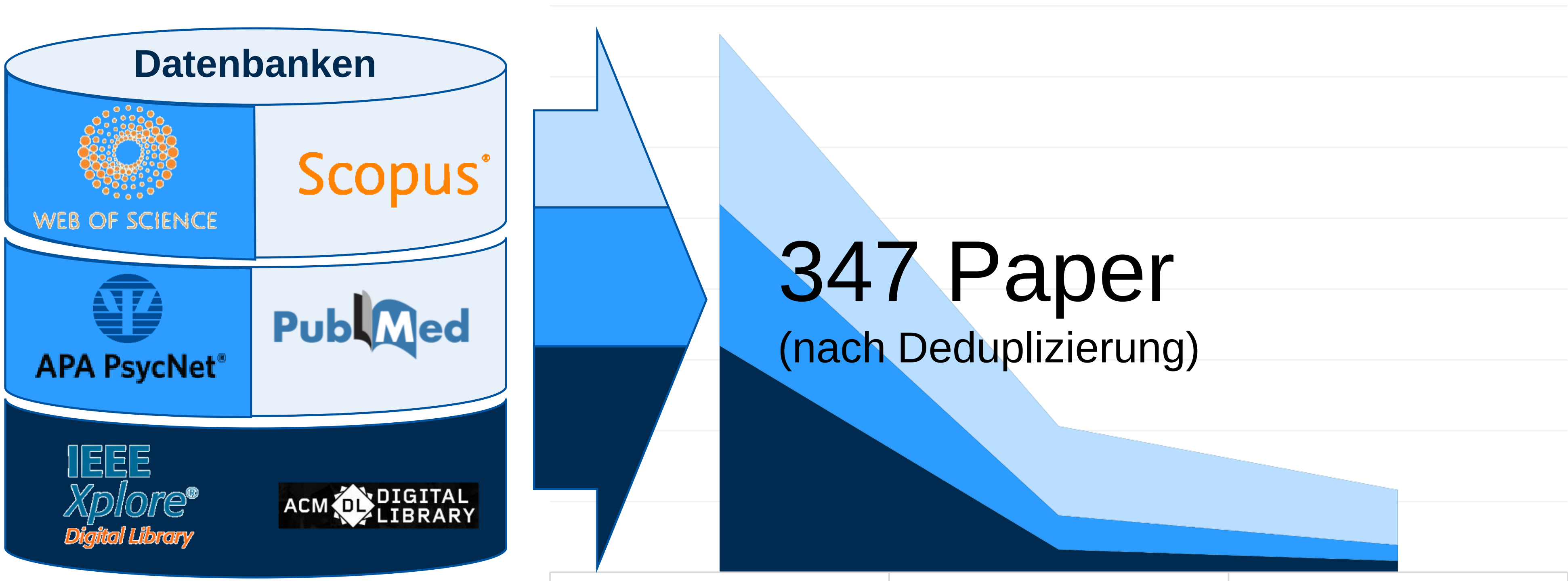
Welche **Nutzerdaten** kommen in Frage?

Welche **Umgebungsdaten** kommen in Frage?

Machine Learning
- Daten
- Zusammenhänge

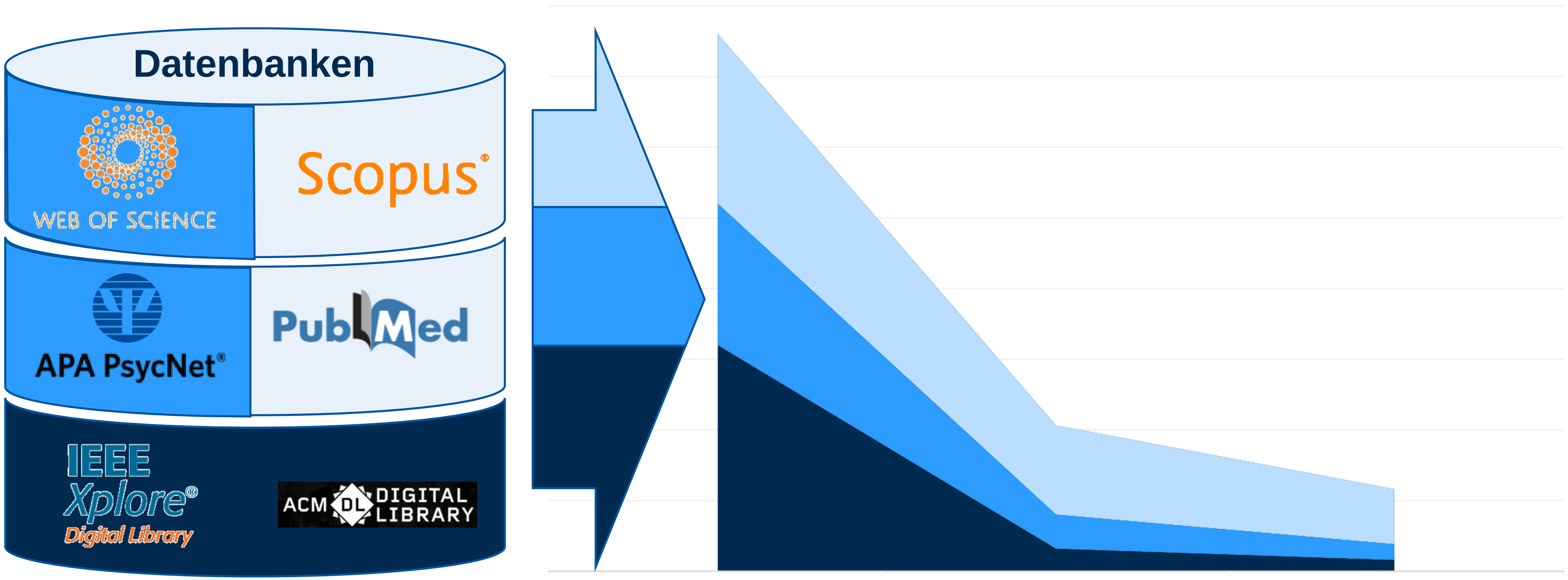
Welche **ML Modelle** kommen in Frage?

Methoden – systematische Literaturrecherche



	Rücklauf	Screening	Volltext Validierung
ML Modelle	120	63	39
Virtuelle Umgebung	100	24	11
Nutzerverhalten	160	16	8

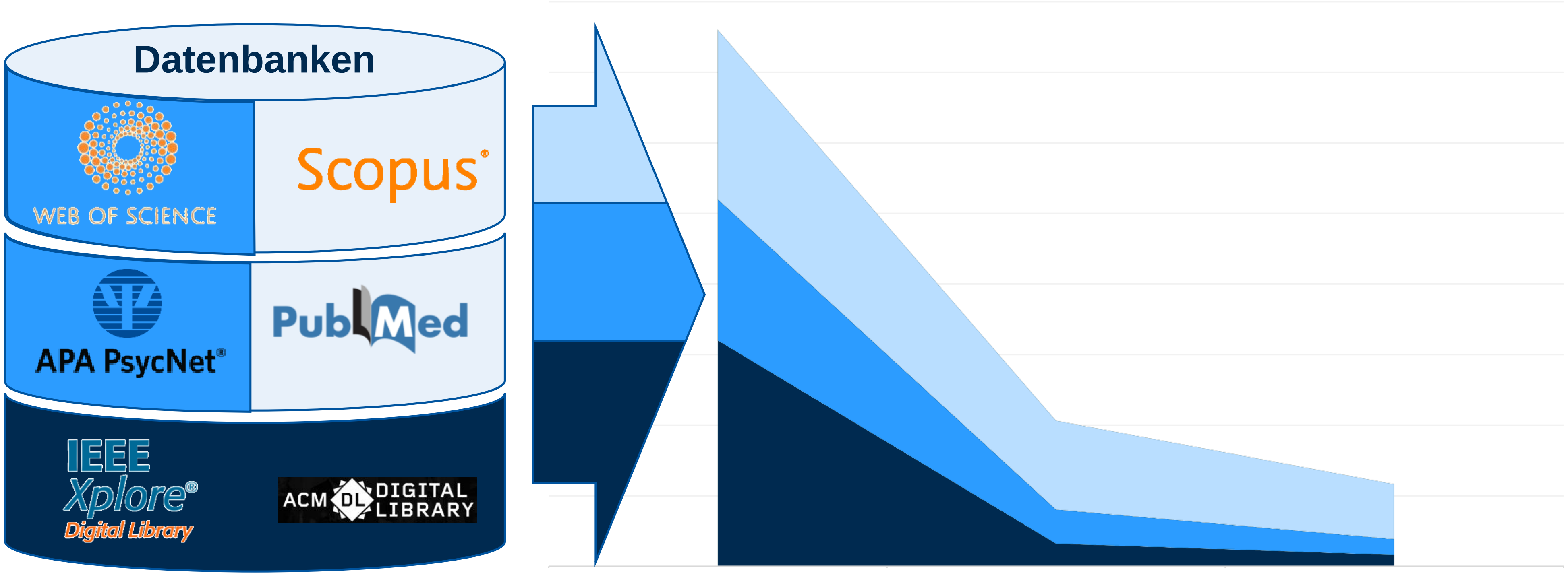
Methoden – systematische Literaturrecherche



Kriterien:

1) **Messung** von **Cognitive Load**
2) In **VR**
3) **Zusammenhang** von Cognitive Load und Verhalten
4) **Stichprobe**nkriterium

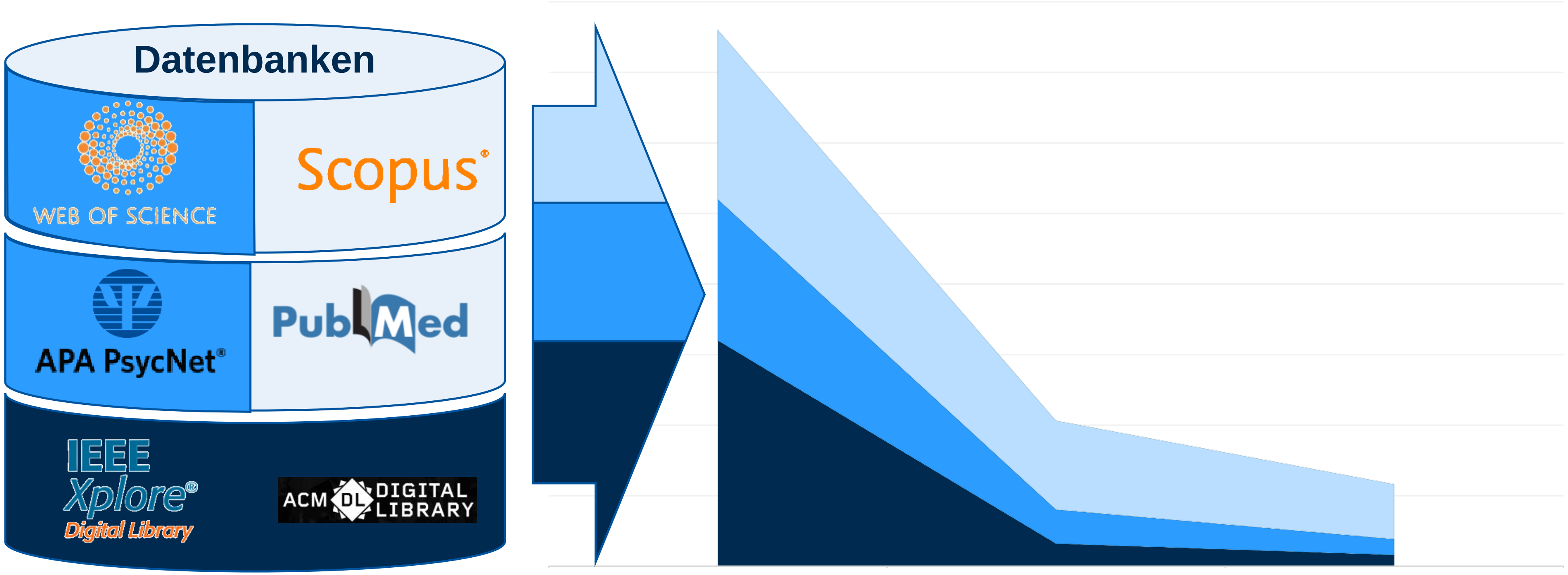
Methoden – systematische Literaturrecherche



Kriterien:

1) **Messung** von **Cognitive Load**
2) Beliebige **Medienstimuli**
3) **Zusammenhang** von Cognitive Load und Medienstimulus
4) **Stichproben**kriterium

Methoden – systematische Literaturrecherche



Kriterien:

1) **Regression**sproblem
2) **Zeitreihendaten**
3) **Vergleich** mit Drittmodellen (Baselines)
4) **keine LLM**

Methoden – Textanalyse (Features)

Textanalyse

Ad-Hoc Features

Synthese

In 2022 they conducted a study on postural sway using both the Centre of Pressure and VR-Head Tracking data. Participants stood heel-to-toe while reading text in VR. Researchers measured the centre of pressure (COP) and head position, compared path length and variance and Power Spectral Density (PSD) for frequency analysis. A randomly introduced dual task saw half the participants counting backwards in steps of 3 from a given number like 567 (564, 561, ...) to induce cognitive load. Results showed the tracked head path length increased in both horizontal axes with increased task load. Forwards variance was lower, with a significant increase for mediolateral PSD 2. Researchers hypothesize that posture correction is susceptible to changes in load, with larger, slower loops getting replaced with small, faster loops under higher load, as automatic, somatosensory based control takes over posture correction. [21]

Studienbeschreibung

Gemessene Variablen

Ergebnisse

The presented literature backs trackable movement based cognitive load correlates, which can be fed into ML models. Correlates include head-to-hand temporal synchronicity, as well as head position path length and spectral head movement analysis as actionable features for 3-point tracking in consumer grade VR. Beyond these features, they have inspired engineering angular head-to-hand synchronicity as well as acceleration based synchronicity.

Anwendung

Ad-Hoc Features

3.1.5 Section Summary

The literature that was investigated in this section identified actionable features for cognitive load tracking in adaVR:

1. Synchronicity measures have been shown to correlate with cognitive load (head-hand)
2. Frequency analysis has been shown to provide cognitive load correlates (spectral head movement analysis, speech formants)
3. Language complexity metrics have also shown to correlate with cognitive load.

Beyond evidence backed metrics, the concepts of synchronicity and frequency have been identified as patterns for further feature engineering.

Feature Klasse

Beispiele

Ad-Hoc Features

Ergebnisse - Features

metrics from literature
ad-hoc metrics

name	object	ref frame	degree	metric
user mic	microphone	window	sampling rate	audio, **volume, volume_var**
audio sources	AudioSource	scene	sampling rate	audio, **count, vol_max, vol_max_Var**
textboxes	UI.Text, TextElement, TMProTMP_text	scene/ canvas	string[](onscreen_text)	**count**, text, delta
haptic feedback	event	-	even(strength, duration)	**count, count_Var, strength, strength_Var, duration, duration_Var**
co-user count		scene		**count**

Ergebnisse

Nutzeraudio → **Formanten Analyse**

name	compute	metric
onscreen text metrics	analysis ([textboxes (text)])	**word-count, complicated word count, lexile level, Flesch-Kinkaid**, (...)
user audio formants	formant analysis ([user mic (audio)])	formants, **formants_var**

Ergebnisse - Features

metrics from literature
ad-hoc metrics

name	object	ref frame	degree	metric
user mic	microphone	window	sampling rate	audio, **volume, volume_var**
audio sources	AudioSource	scene	sampling rate	audio, **count, vol_max, vol_max_Var**
textboxes	UI.Text, TextElement, TMProTMP_text	scene/ canvas	string[](onscreen_text)	**count**, text, delta
haptic feedback	event	-	even(strength, duration)	**count, count_Var, strength, strength_Var, duration, duration_Var**
co-user count		scene		**count**

Ad-Hoc Features

Gruppenkommunikation → **Sprachkomplexität**

name	compute	metric
onscreen text metrics	analysis ([textboxes (text)])	**word-count, complicated word count, lexile level, Flesch-Kinkaid**, (…)
user audio formants	formant analysis ([user mic (audio)])	formants, **formants_var**

Methoden – Textanalyse (Machine Learning Modelle)

Textanalyse

ELO Vergleich

Synthese

3.3.7 Liquid State Machines

In a 2023 study Researchers created a cost efficient online forecasting algorithm based on a neuromorphic spiking reservoir. The algorithm, a Liquid State Machine (LMS) forecaster with spiking neurons, was more robust than stacked LSTM, SARIMA and ARIMA models while performing at lower, constant cost. It's a univariate online learner and requires prior knowledge of the feature value range. [42]

The reservoir of LSMs acts as a gradually forgetting memory in an online learner. This algorithm is characterised as cost efficient and robust, the constant performance cost may be especially valuable for longer VR applications at high repetion rate. It would have to be adapted to support multivariate time series for application in cognitive load prediction.

Model

Baselines

Eigenschaften

Anwendung

Methoden - Machine Learning Modelle

Textanalyse

ELO Vergleich

Synthese

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

- **Schach**
- **ordinal**skalierte **Paar**vergleiche
- Initialisiert um *E(x) = 1000*
- **probablistische** Methode

Prof. Arpad Elo

Methoden - Machine Learning Modelle

Textanalyse

ELO Vergleich

Synthese

3.3.7 Liquid State Machines

In a 2023 study Researchers created a cost efficient online forecasting algorithm based on a neuromorphic spiking reservoir. The algorithm, a Liquid State Machine (LMS) forecaster with spiking neurons, was more robust than stacked LSTM, SARIMA and ARIMA models while performing at lower, constant cost. It's a univariate online learner and requires prior knowledge of the feature value range. [42]

The reservoir of LSMs acts as a gradually forgetting memory in an online learner. This algorithm is characterised as cost efficient and robust, the constant performance cost may be especially valuable for longer VR applications at high repetion rate. It would have to be adapted to support multivariate time series for application in cognitive load prediction.

Model

Baselines

Eigenschaften

Anwendung

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

3.3.22 Section Summary

This section offers a structured overview of the reviewed papers which imployed a wide array of machine learning models, largely on online learning regression problems. It closely follows a set of constraints to evaluate the fit of a series of algorithms. Tree based algorithms, ARIMA models and Long-Term-Short-Term models are especially frequently evaluated in this context. Special interest is given to Infinitel Lattice Learners and an other wrapping model, which could increase accuracy of their base models while maintaining performance either by partitioning them into part of the input space or by using cumulated multistep forecasting data for learning, as well as to Mixture of Experts models, which architecturally might best reflect the different tasks and environments, that meet users in virtual environments.

After the review, the data was consolidated in accordance to its ordinal scale in an elo-rating system across all reviewed papers. Careful interpretations demonstrate that generally speaking, the provided literature provides better algorithms than k Nearest Neighbours, Decision Trees and ARIMA for the reported use cases, but Gradient Boosting Methods and Random Forrests rank high on accuracy in the tasks that see them implemented.

Textanalyse

ELO Rating

Ergebnisse – Machine Learning Modelle (Architektur)

Textanalyse

ELO Vergleich

Synthese

3.3.22 Section Summary

This section offers a structured overview of the reviewed papers which imployed a wide array of machine learning models, largely on online learning regression problems. It closely follows a set of constraints to evaluate the fit of a series of algorithms. Tree based algorithms, ARIMA models and Long-Term-Short-Term models are especially frequently evaluated in this context. Special interest is given to Infinitel Lattice Learners and an other wrapping model, which could increase accuracy of their base models while maintaining performance either by partitioning them into part of the input space or by using cumulated multistep forecasting data for learning, as well as to Mixture of Experts models, which architecturally might best reflect the different tasks and environments, that meet users in virtual environments.

After the review, the data was consolidated in accordance to its ordinal scale in an elo-rating system across all reviewed papers. Careful interpretations demonstrate that generally speaking, the provided literature provides better algorithms than k Nearest Neighbours, Decision Trees and ARIMA for the reported use cases, but Gradient Boosting Methods and Random Forrests rank high on accuracy in the tasks that see them implemented.

- **Wrapping Modelle** - Verbesserung der Leistung einfacher Modelle (online)
- **Mixture of Expert Modelle** – Experten je Task-Anforderung

Ergebnisse - Machine Learning Modelle

Textanalyse

ELO Vergleich

Synthese

Elo Ratings

2025-12-28 14:33:30 - Min Games: 5 - Ratings initialized at 1000

Rank	Games	Algorithm	Rating
1	31	GBT	1212.4
2	18	AR-LSTM	[illegible]
3	6	PAPAGEI-S	[illegible]
4	18	ResLSTM	[illegible]
5	5	SplineR	[illegible]
6	64	RF	1037.0
7	7	ESN	1029.3
8	45	SVM	1023.0
9	29	XGBoost	1011.4
10	10	LogR	1011.3
11	7	NRU	1007.8
12	9	LassoR	1002.9
13	6	GNB	1000.8
14	43	kNN	998.3
15	5	RNN	989.7
16	7	GRU	[illegible]
17	10	MLP	[illegible]
18	15	NN	970.7
19	7	naive	963.3
20	6	RidgeR	962.4
21	49	DT	958.1
22	6	average	953.3
23	17	LR	947.9
24	16	LSTM	943.5
25	7	VAR	923.0
26	19	ARIMA	844.6

- **Gradient Boosting**, **Random Forrest** und SVM überdurchschnittliches Ranking, viele Vergleiche

selten verglichene, hoch platzierte Modelle, wurden gegen viele Baselines getestet

E(X) = 1000

- **kNN DT** schwächeres Ranking, viele Vergleiche

häufig verglichene, tiefplatzierte Modelle dienen in der Literatur als zu übertreffende Baselines

Hydrology Lehr- und Forschungsgebiet Ingenieurhydrologie

RWTH AACHEN UNIVERSITY

Wozu Cognitive Load messen?

Stiftung Innovation in der Hochschullehre

Projektdaten

Projektnr.	Projektnummer 3003-1084
Hochschulname	Rheinisch-Westfälische Technische Hochschule Aachen
Projekttitel	Adaptive Digital Assistance in Virtual Reality
Akronym	adaVR

Arbeitsplan Freiraum 2025

Bitte markieren Sie die entsprechenden Monate, in denen das Arbeitspaket umgesetzt wird, in einer beliebigen Farbe.

Bei mehr als 10 Arbeitspaketen fügen Sie neue Zeilen hinzu.

AP-Nr.	Titel des Arbeitspakets (Zwischen 10 und 70 Zeichen)	Kurzschreibung des Arbeitspakets (Zwischen 30 und 200 Zeichen)
AP01	Entwicklung eines VR-Aufzeichnungstools für Nutzungs-Logs	Zur Aufzeichnung von Nutzungsdaten, die in späteren Arbeitspaketen zum Training genutzt werden.
AP02	Erstellung von Fragebögen für Studierende	Zur Selbsteinschätzung und Wahrnehmung der VR-Erfahrung der Studierenden.
AP03	Integration des Aufzeichnungstools in bestehende VR-Anwendung	Integration in eine bereits bestehende VR-Anwendung, um die praktische Anwendbarkeit zu erproben.
AP04	Datenerhebung über den Einsatz von VR in der Lehre ohne Assistenzsystem	Erste Datensammlung, die Trainingsdaten für späteren Einsatz von ML-Algorithmen liefert. Außerdem Generierung von Vergleichsdaten für Evaluation.
AP05	Training & Evaluation von Machine Learning mit den gewonnenen Daten	Erstellung eines ML-Algorithmus für das adaptive Assistenzsystem. Verschiedene ML-Verfahren werden gegeneinander getestet
AP06	Implementierung des adaptiven Assistenzsystems	In AP05 gewählter Algorithmus wird implementiert und an Hilfsmaßnahmen gekoppelt.
AP07	Datenerhebung über den Einsatz von VR in der Lehre mit Assistenzsystem	Zweite Datensammlung, die Daten zur Evaluation der Wirksamkeit der Maßnahmen liefert.
AP08	Evaluation des Adaptiven Assistenzsystems	Abschließende statistische Evaluation des adaptiven Assistenzsystems.

	2025									2026												2027		
AP-Nr.	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez	Jan	Feb	Mrz	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez	Jan	Feb	Mrz
AP01	■	■	■	■																				
AP02	■	■	■	■	■	■																		
AP03					■	■																		
AP04							■	■	■	■														
AP05											■	■	■											
AP06													■	■	■	■	■	■						
AP07																			■	■	■	■		
AP08																							■	■

Cognitive Load messen

Entwicklung
Auto-Adaptiver VLEs

Vielen Dank
für Ihre Aufmerksamkeit

Bildquellen

John Sweller - https://static.wikitide.net/psutechcommtopicswiki/9/9b/Cog_load_theory2.jpeg
Arpad Elo - https://www.historyhit.com/app/uploads/2021/11/1980-Elo2.jpg?x76761

Ergebnisse - Features

metrics from literature
ad-hoc metrics

name	object	ref frame	degree	metric
user mic	microphone	window	sampling rate	audio, **volume, volume_var**
audio sources	AudioSource	scene	sampling rate	audio, **count, vol_max, vol_max_Var**
textboxes	UI.Text, TextElement, TMProTMP_text	scene/ canvas	string[](onscreen_text)	**count**, text, delta
haptic feedback	event	-	even(strength, duration)	**count, count_Var, strength, strength_Var, duration, duration_Var**
co-user count		scene		**count**

„optimales Interface" → **Feedback Signale**

name	compute	metric
onscreen text metrics	analysis ([textboxes (text)])	**word-count, complicated word count, lexile level, Flesch-Kinkaid**, (…)
user audio formants	formant analysis ([user mic (audio)])	formants, **formants_var**