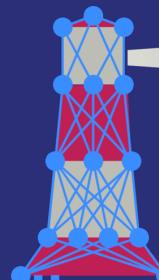


# Chatbots: Möglichkeiten & Grenzen



**illuminAld GmbH**

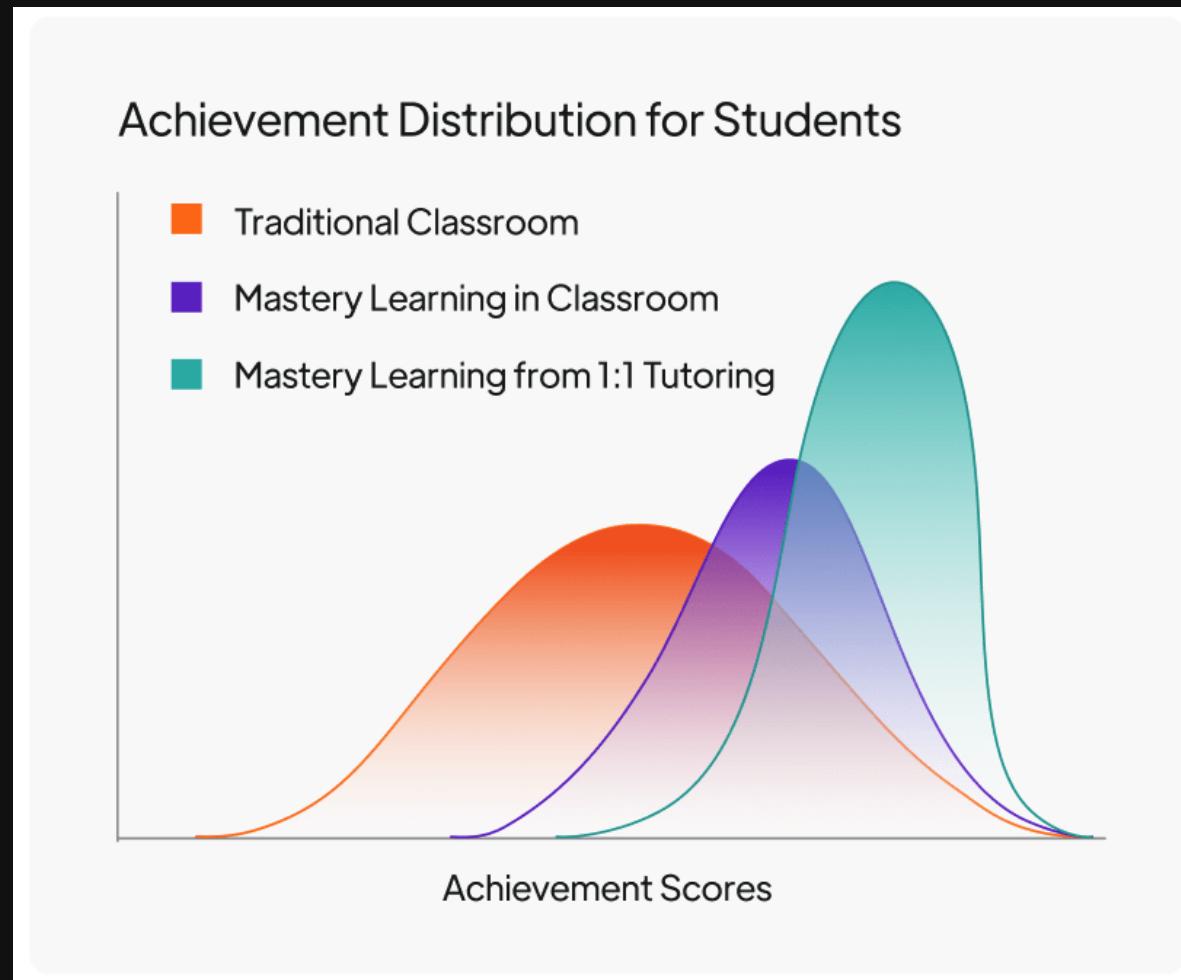
Brighter with AI

Dr. Stefan Hackstein

[www.illuminAld.de](http://www.illuminAld.de)

# Begleitetes Lernen

Bloom (1984): individuelles Tutoring  $\Rightarrow 2\sigma$  Verbesserung  
??? mit KI für jeden ???



# Education Today

## Hindernisse

- Asynchrone Geschwindigkeit  $v_{\text{Lehrer}} \neq v_{\text{Study 1}} \neq v_{\text{Study 2}} \neq \dots$
- Vorwissen unbekannt
- Viele Fragen, wenig Zeit, Angst zu Fragen

# The Future of Education

## Hindernisse

- Asynchrone Geschwindigkeit  $v_{\text{Lehrer}} \neq v_{\text{Study 1}} \neq v_{\text{Study 2}} \neq \dots$
- Vorwissen unbekannt
- Viele Fragen, wenig Zeit, Angst zu Fragen

## Asynchrones Lernen mit KI Tutor

- **individuelle** Geschwindigkeit
- Platz für **alle Fragen**
- **wertvollere Interaktion** zwischen Studys & Lehrkraft

# Asynchrones Lernen

## Collaborative Education

- Studys lernen individuell
- Lehrkraft unterstützt

Zaka, Fox, Docherty 2019:

- mehr Kollaboration
- individuelle Lernwege

<https://www.youtube.com/embed/GwpVZBcBCIM?enablejsapi=1>

<https://www.youtube.com/embed/QTEmdunMfCk?enablejsapi=1>

# Asynchrones Lernen

# Collaborative Education

- Studys lernen individuell
- Lehrkraft unterstützt

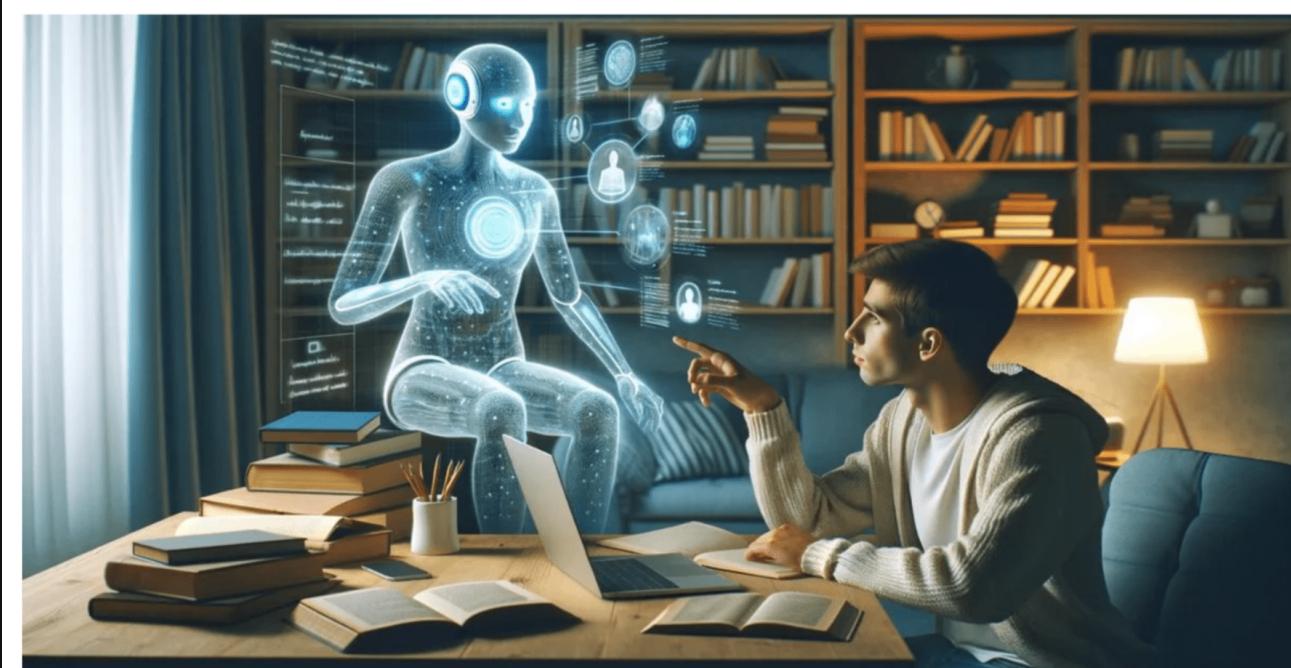


KI unterstützt  
potenziert individuelles Lernen

**KI Tutor für den eigenen Unterricht?**

**=> KI in der Lehre - Intermediate**

# Persönlicher Tutor: Vision



THE FUTURE OF LEARNING: PERSONALIZED  
AI TUTORS FOR EVERY STUDENT



Peter H. Diamandis [in](#)

Data-driven optimist inspiring entrepreneurs through research,  
investment & community to create an abundant future for humanity |...



Sir Anthony Seldon:

"Machines will replace Teachers in 10 years."

# Persönlicher Tutor: Realität

- Eingeschränkte Individualisierung
- Begrenztes Verständnis
- Mangelnde Flexibilität
- Informationsflut
- Halluzinationen
- Potemkins Mankoridis+ 2025



THE FUTURE OF LEARNING: PERSONALIZED  
AI TUTORS FOR EVERY STUDENT



Peter H. Diamandis [in](#)

Data-driven optimist inspiring entrepreneurs through research,  
investment & community to create an abundant future for humanity |...



 What is an ABAB rhyming scheme?

 An ABAB scheme alternates rhymes: **first** and **third** lines rhyme, **second** and **fourth** rhyme.

Correct

 Fill in the blank so that the resultant poem uses the ABAB rhyming scheme:

*Wondrous winter calls **out***

*Shivering under the frost*

*Lies a lonely cat, sitting **[?]***

*Alone but hardly lost*

 Lies a lonely cat, sitting **soft**

Potemkin

 Does **out** rhyme with **soft**?

 No

Correct

Potemkins Mankoridis+ 2025

Chatbot kann Vorgehen erklären, aber nicht ausführen

# Forschungsstand

## Generative AI chatbots in higher education: a review of an emerging research area

Cormac McGrath<sup>1</sup>  · Alexandra Farazouli<sup>1</sup>  · Teresa Cerratto-Pargman<sup>2</sup> 

Accepted: 9 August 2024  
© The Author(s) 2024

[link to paper](#)



- Eklektische Phase der aufkommenden Forschung
- Unheinheitliche Konzeption des Lernens
- Diskurs: Utopisch bis Dystopisch



# **ChatGPT banned in some schools, but many experts say it can improve education**

# ChatGPT banned in some schools, but many experts say it can improve education

ChatGPT & Künstliche Intelligenz

## Uni in Prag schafft Bachelorarbeiten ab

von Laura Meyer

29.12.2023 | 14:11

[Cosmos](#) » [Technology](#)

# ChatGPT banned in some schools, but many experts say it can improve education

ChatGPT & Künstliche Intelligenz

## Uni in Prag schafft Bachelorarbeiten ab

von Laura Meyer

29.12.2023 | 14:11

Ohio State University says all students will be required to train in AI

The  
Guardian

University to embed curriculum teaching undergraduates how artificial intelligence 'can be responsibly applied'

# Illuminate

3-Jahre Forschungsprojekt BeLearn



Dr. Andrew Ellis

- Bayesian Mastery Tracking
- Optimized Exercise Selection
- Targeted Multi-Agent AI-Tutoring



Dr. Stefan  
Hackstein

- Research: Unified Platform
- Education: Sophisticated AI-Tutoring System

## Translation:

- Students: Next Generation Learning Tool
- Lecturers: Insight for targeted intervention
- Institution: Learning Analytics

# Illusion of Thinking

## 7. Jun 2025

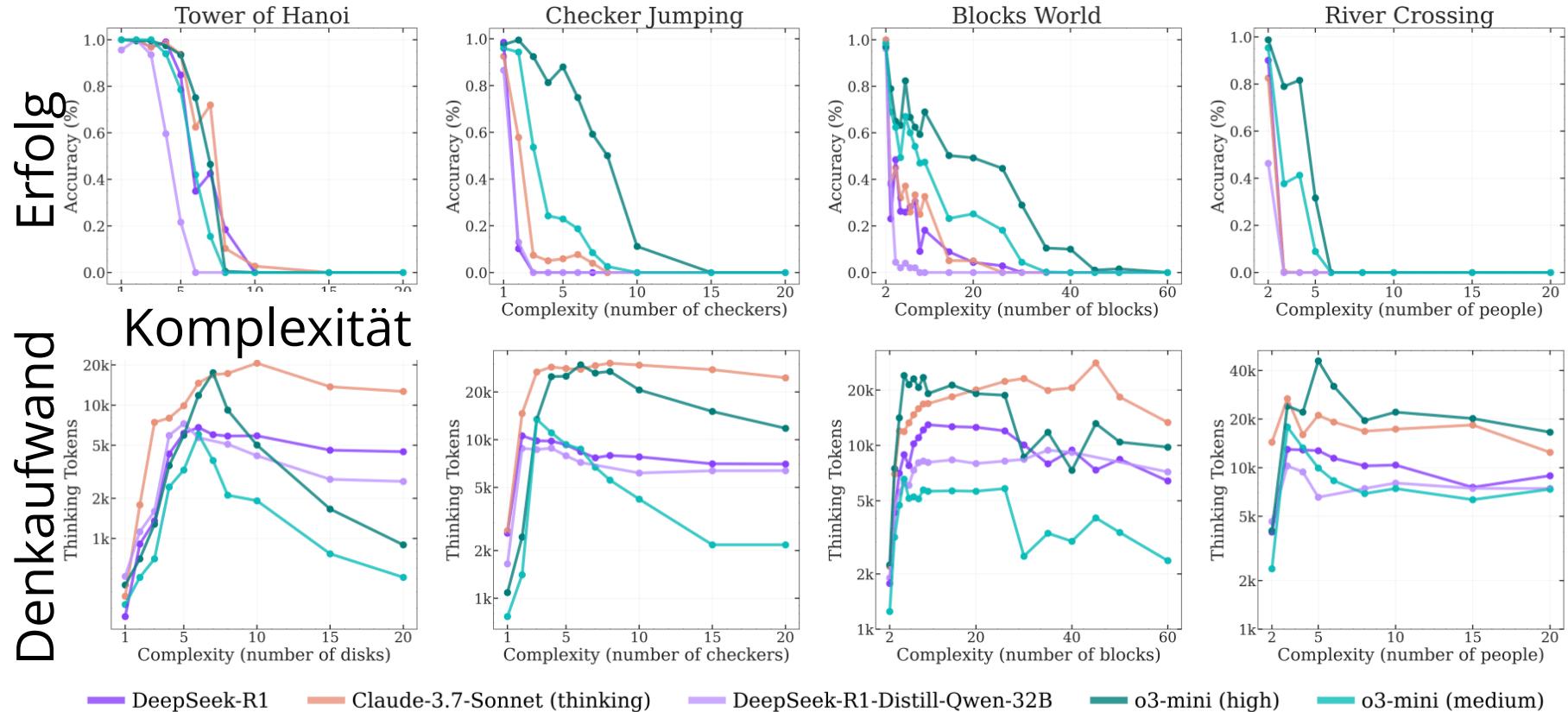
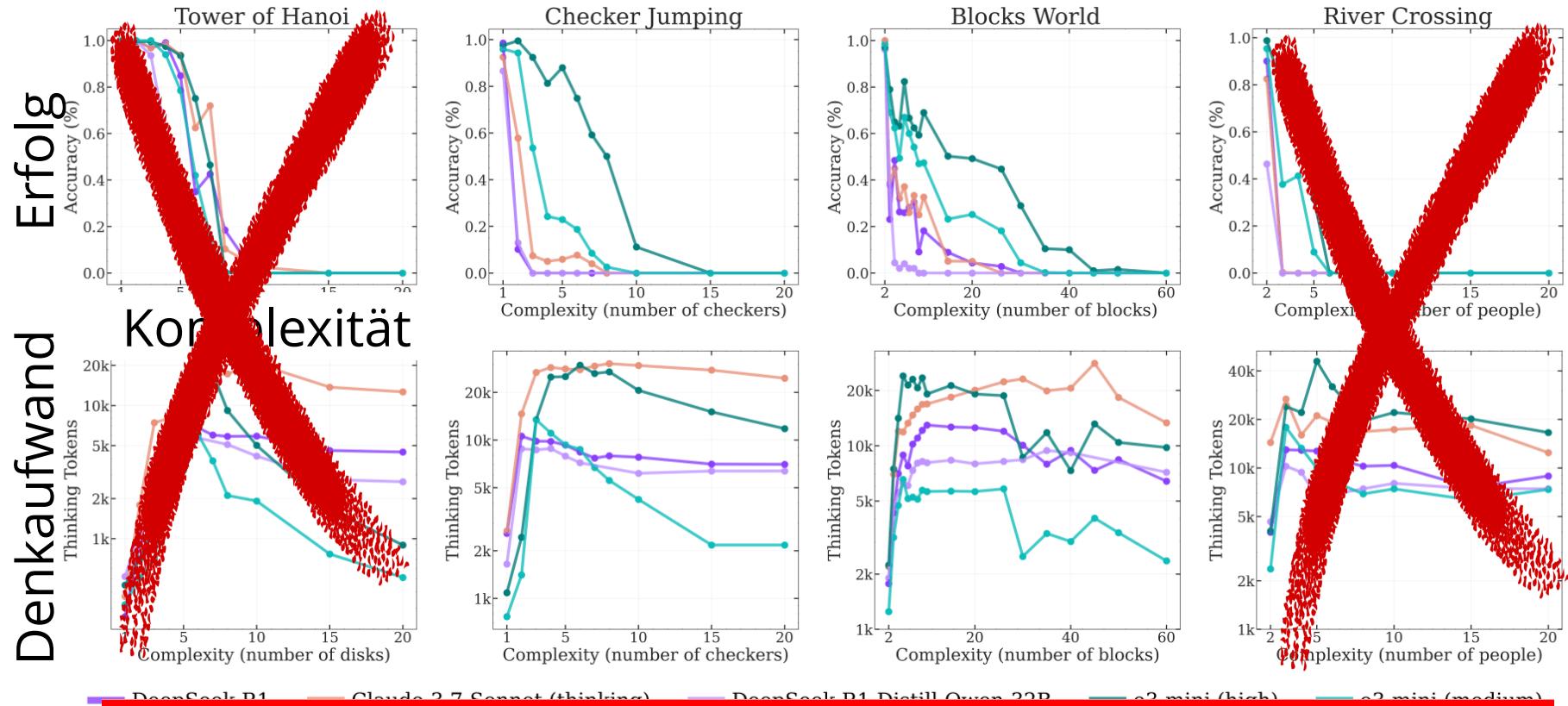


Figure 6: Accuracy and thinking tokens vs. problem complexity for reasoning models across puzzle environments. As complexity increases, reasoning models initially spend more tokens while accuracy declines gradually, until a critical point where reasoning collapses—performance drops sharply and reasoning effort decreases.

# Illusion of Thinking

## 7. Jun 2025



KI betreibt begrenzten Denkaufwand

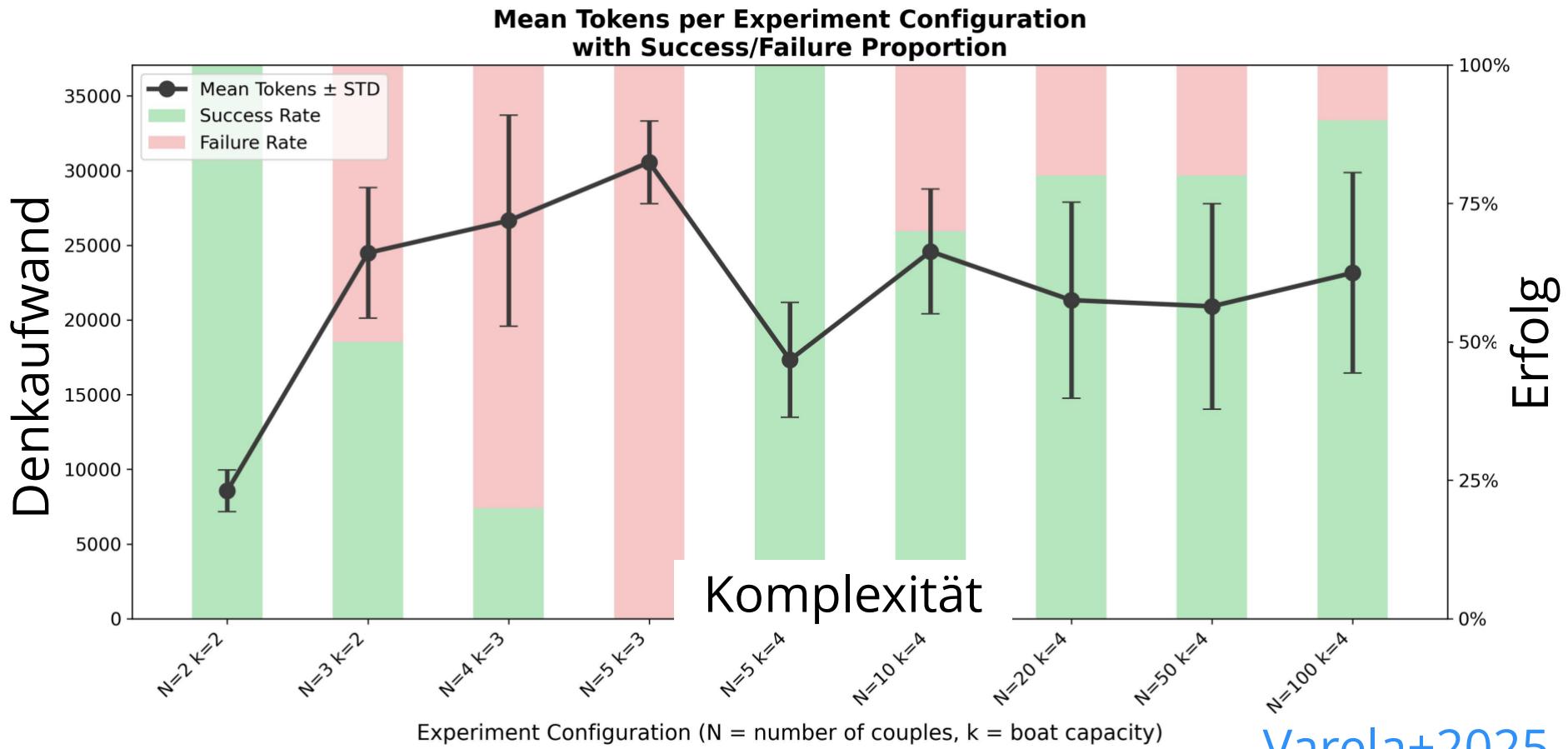
Fig

environments. As complexity increases, reasoning models initially spend more tokens while accuracy declines gradually, until a critical point where reasoning collapses—performance drops sharply and reasoning effort decreases.

Shojaee+ 2025

# Rethinking Illusion of Thinking

3. Jul 2025



Lösung durch richtiges Verständnis von Problem & KI

# How do Olympiad Medalists Judge LLMS

## in competitive Programming

Zheng+ 2025

Model	Hard	Medium	Easy	Rating	Pct.%	AvgTok	AvgCost
<i>Reasoning Models</i>							
o4-mini-high	0.0%	53.5%	83.1%	2 116	1.5%	23 819	\$0.1048
Gemini 2.5 Pro	0.0%	25.4%	70.4%	1 992	2.3%	29 879	\$0.2988
o3-mini	0.0%	16.9%	77.5%	1 777	4.9%	18 230	\$0.0802
DeepSeek R1	0.0%	9.9%	56.3%	1 442	18.0%	16 716	\$0.0366
Gemini 2.5 Flash	0.0%	12.7%	47.9%	1 334	30.3%	35 085	\$0.0116
DeepSeek R1 Distill-Llama-70B	0.0%	2.8%	33.8%	999	56.0%	12 425	\$0.0050
Claude 3.7 Sonnet (Max Reasoning)	0.0%	1.4%	36.6%	992	56.5%	19 075	\$0.2861
Gemini 2.0 Flash Reasoning	0.0%	0.0%	29.6%	893	63.1%	11 143	\$0.0390
<i>Non-Reasoning Models</i>							
GPT-4.1 mini	0.0%	5.6%	28.2%	1 006	55.5%	2 662	\$0.0043
DeepSeek V3 0324	0.0%	5.6%	32.4%	984	57.1%	2 712	\$0.0030
GPT-4.1	0.0%	0.0%	23.9%	889	64.2%	2 131	\$0.0170
GPT-4.5	0.0%	0.0%	26.8%	881	64.8%	968	\$0.1452
Qwen-Max	0.0%	0.0%	14.1%	821	69.4%	1 244	\$0.0080
Claude 3.7 Sonnet (No Reasoning)	0.0%	1.4%	16.9%	804	70.7%	3 554	\$0.0533
Llama 4 Maverick	0.0%	0.0%	15.5%	634	80.4%	1 160	\$0.0007
Claude 3.5 Sonnet	0.0%	0.0%	14.1%	617	81.4%	810	\$0.0122
Gemma 3 27B	0.0%	0.0%	8.5%	601	82.5%	668	\$0.0001
GPT-4o	0.0%	0.0%	9.9%	592	83.1%	1 133	\$0.0227
Meta Llama 3.1 405B Instruct	0.0%	0.0%	9.9%	574	84.3%	568	\$0.0005
DeepSeek V3	0.0%	0.0%	12.7%	557	84.9%	1 020	\$0.0011

# How do Olympiad Medalists Judge LLMS in competitive Programming

Zheng+ 2025

Model	Hard	Medium	Easy	Rating	Pct.%	AvgTok	AvgCost
<i>Reasoning Models</i>							
o4-mini-high	0.0%	53.5%	83.1%	2 116	1.5%	23 819	\$0.1048
Gemini 2.5 Pro	0.0%	25.4%	70.4%	1 992	2.3%	29 879	\$0.2988
o3-mini	0.0%	16.9%	77.5%	1 777	4.9%	18 230	\$0.0802
DeepSeek R1	0.0%	9.9%	56.3%	1 442	18.0%	16 716	\$0.0366
Gemini 2.5 Flash	0.0%	12.7%	47.9%	1 334	30.3%	35 085	\$0.0116
DeepSeek R1 Distill-Llama-70B	0.0%	2.8%	33.8%	999	56.0%	12 425	\$0.0050
Claude 3.7 Sonnet (Max Reasoning)	0.0%	1.4%	36.6%	992	56.5%	19 075	\$0.2861
Gemini 2.0 Flash Reasoning	0.0%	0.0%	29.6%	893	63.1%	11 143	\$0.0390

KI kann viele Einfache & Mittlere Aufgaben lösen

KI scheitert völlig an wirklich schwierigen Aufgaben

Studies: Versuchung ist Gross / Übung ist wichtig

H

MS

2025

---

## Model

---

Reasoning  
o4-mini  
Gemini  
o3-mini  
DeepSee  
Gemini  
DeepSee  
Claude  
Gemini

---

## AvgCost

---

\$0.1048  
\$0.2988  
\$0.0802  
\$0.0366  
\$0.0116  
\$0.0050  
\$0.2861  
\$0.0390

K  
KI  
S

n  
en  
g



**SOLUTIONS**

# Chatbots: Möglichkeiten & Grenzen

- können Lehrende entlasten, Lehre bereichern
- können individuelles Lernen potenzieren
- aber starker Anreiz, essentielle Übung zu umgehen
- Forschung & Hochschulentwicklung sehr heterogen
- BFH in Vorreiterrolle