

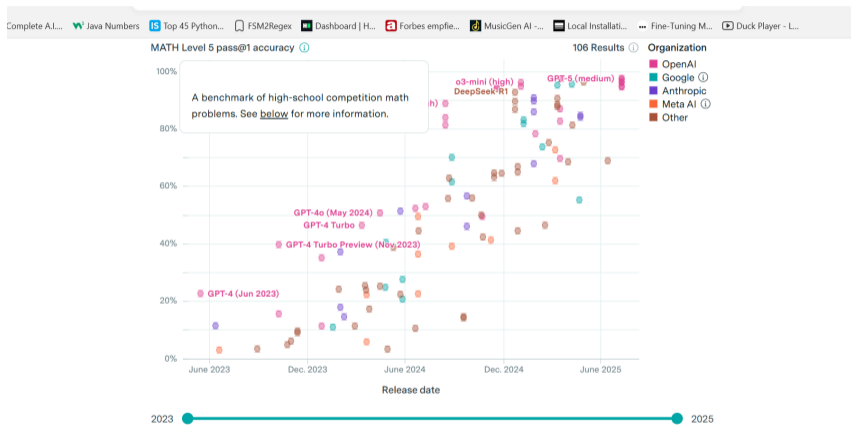
# Künstliche Intelligenz und Mathematik

Merlin Carl und Hinrich Lorenzen

Institut für Mathematik  
Europa-Universität Flensburg

May 18, 2026

# Mathematische Benchmarks 1: Schulmathematik



Quelle: <https://epoch.ai/benchmarks>

## FrontierMath Tier 4

*FrontierMath is a benchmark of hundreds of original, exceptionally challenging mathematics problems crafted and vetted by expert mathematicians. The questions cover most major branches of modern mathematics – from computationally intensive problems in number theory and real analysis to abstract questions in algebraic geometry and category theory. Solving a typical problem requires multiple hours of effort from a researcher in the relevant branch of mathematics, and for the upper end questions, multiple days.*

Epoch AI, [https:](https://epoch.ai/benchmarks/frontiermath-tier-4?view=graph&tab=release-date)

[//epoch.ai/benchmarks/frontiermath-tier-4?view=graph&tab=release-date](https://epoch.ai/benchmarks/frontiermath-tier-4?view=graph&tab=release-date)

*“I came up with a problem which experts in my field would recognize as an open question in number theory—a good Ph.D.-level problem,” (...)*

# Forschungsmathematik

*“I came up with a problem which experts in my field would recognize as an open question in number theory—a good Ph.D.-level problem,” (...)  
The bot spent the first two minutes finding and mastering the related literature in the field.*

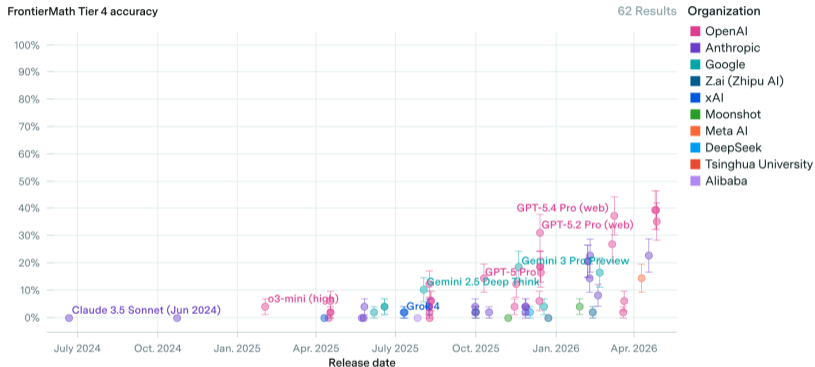
# Forschungsmathematik

*“I came up with a problem which experts in my field would recognize as an open question in number theory—a good Ph.D.-level problem,” (...)  
The bot spent the first two minutes finding and mastering the related literature in the field. Then it wrote on the screen that it wanted to try solving a simpler “toy” version of the question first in order to learn. A few minutes later, it wrote that it was finally prepared to solve the more difficult problem. Five minutes after that, o4-mini presented a correct (...) solution.*

(L. Chiou, At Secret Math Meeting, Researchers Struggle to Outsmart AI, Scientific American (2025))

# Mathematische Benchmarks 2: Forschungsmathematik

## AI performance on the most challenging expert-level mathematics problems



EPOCH AI | CC-BY

epoch.ai

Quelle: <https://epoch.ai/benchmarks/frontiermath-tier-4?view=graph&tab=release-date>

<https://epoch.ai/benchmarks/frontiermath-tier-4?view=graph&tab=release-date>

## Ist Mathematik obsolet? (Jeremy Avigad, 2025)

*One can, therefore, argue that symbolic methods are no longer important because mathematical reasoning, as we know it, is no longer important. Technology has given us something better. (...)*

## Ist Mathematik obsolet? (Jeremy Avigad, 2025)

*One can, therefore, argue that symbolic methods are no longer important because mathematical reasoning, as we know it, is no longer important. Technology has given us something better. (...)*

*It is therefore important to consider what role mathematics should still play in our lives, if any, and why.*

*(Jeremy Avigad, Is Mathematics Obsolete? (2025))*

## Woher diese Stärke?

- ▶ Generelle Erfahrung: Generative KI-Systeme sind bisweilen erstaunlich stark, aber unzuverlässig

## Woher diese Stärke?

- ▶ Generelle Erfahrung: Generative KI-Systeme sind bisweilen erstaunlich stark, aber unzuverlässig
- ▶ Halluzinationen; kein Selbstwissen; falsche Informationen in überzeugendem Stil dargelegt.

## Woher diese Stärke?

- ▶ Generelle Erfahrung: Generative KI-Systeme sind bisweilen erstaunlich stark, aber unzuverlässig
- ▶ Halluzinationen; kein Selbstwissen; falsche Informationen in überzeugendem Stil dargelegt.
- ▶ Kein Kontakt zur “wirklichen Welt”, um das zu kompensieren (ein LLM kann nicht “nachsehen”).

## Woher diese Stärke?

- ▶ Generelle Erfahrung: Generative KI-Systeme sind bisweilen erstaunlich stark, aber unzuverlässig
- ▶ Halluzinationen; kein Selbstwissen; falsche Informationen in überzeugendem Stil dargelegt.
- ▶ Kein Kontakt zur “wirklichen Welt”, um das zu kompensieren (ein LLM kann nicht “nachsehen”).
- ▶ Erste LLMs haben so auch auf Mathematik reagiert: Weitgehend unsinnige Pseudo-Beweistexte.

## Woher diese Stärke?

- ▶ Generelle Erfahrung: Generative KI-Systeme sind bisweilen erstaunlich stark, aber unzuverlässig
- ▶ Halluzinationen; kein Selbstwissen; falsche Informationen in überzeugendem Stil dargelegt.
- ▶ Kein Kontakt zur “wirklichen Welt”, um das zu kompensieren (ein LLM kann nicht “nachsehen”).
- ▶ Erste LLMs haben so auch auf Mathematik reagiert: Weitgehend unsinnige Pseudo-Beweistexte.
- ▶ Was ist passiert?

# Klassische automatische Theorembeweiser

- ▶ Automated Theorem Prover (ATP)

# Klassische automatische Theorembeweiser

- ▶ Automated Theorem Prover (ATP)
- ▶ Seit den frühen 60er Jahren

# Klassische automatische Theorembeweiser

- ▶ Automated Theorem Prover (ATP)
- ▶ Seit den frühen 60er Jahren
- ▶ Formallogische Systeme: Formalisierte Axiome, formale Ableitungsregeln

# Klassische automatische Theorembeweiser

- ▶ Automated Theorem Prover (ATP)
- ▶ Seit den frühen 60er Jahren
- ▶ Formallogische Systeme: Formalisierte Axiome, formale Ableitungsregeln
- ▶ Damit: Suche nach Beweis ist systematisches Durchsuchen von Ableitungsbäumen

# Klassische automatische Theorembeweiser

- ▶ Automated Theorem Prover (ATP)
- ▶ Seit den frühen 60er Jahren
- ▶ Formallogische Systeme: Formalisierte Axiome, formale Ableitungsregeln
- ▶ Damit: Suche nach Beweis ist systematisches Durchsuchen von Ableitungsbäumen
- ▶ Problem: Kombinatorische Explosion; keine “großen Ideen”.

# Stärke neuerer Systeme: Beweisassistenten als “ground truth”

- ▶ Sprachmodell “spielt” gegen Beweisassistenten

## Stärke neuerer Systeme: Beweisassistenten als “ground truth”

- ▶ Sprachmodell “spielt” gegen Beweisassistenten
- ▶ Beweis erst akzeptiert, wenn Beweisassistent zustimmt

## Stärke neuerer Systeme: Beweisassistenten als “ground truth”

- ▶ Sprachmodell “spielt” gegen Beweisassistenten
- ▶ Beweis erst akzeptiert, wenn Beweisassistent zustimmt
- ▶ Halluzinationen scheitern am Assistenten

## Stärke neuerer Systeme: Beweisassistenten als “ground truth”

- ▶ Sprachmodell “spielt” gegen Beweisassistenten
- ▶ Beweis erst akzeptiert, wenn Beweisassistent zustimmt
- ▶ Halluzinationen scheitern am Assistenten
- ▶ Möglich durch deutlichen Zuwachs in der Fähigkeit zur Autoformalisierung

## Stärke neuerer Systeme: Beweisassistenten als “ground truth”

- ▶ Sprachmodell “spielt” gegen Beweisassistenten
- ▶ Beweis erst akzeptiert, wenn Beweisassistent zustimmt
- ▶ Halluzinationen scheitern am Assistenten
- ▶ Möglich durch deutlichen Zuwachs in der Fähigkeit zur Autoformalisierung
- ▶ Damit wird sogar die menschliche *Überprüfung* des Beweises obsolet (anders als bei den meisten anderen Anwendungen generativer KI).

## Stärke neuerer Systeme: Beweisassistenten als “ground truth”

- ▶ Sprachmodell “spielt” gegen Beweisassistenten
- ▶ Beweis erst akzeptiert, wenn Beweisassistent zustimmt
- ▶ Halluzinationen scheitern am Assistenten
- ▶ Möglich durch deutlichen Zuwachs in der Fähigkeit zur Autoformalisierung
- ▶ Damit wird sogar die menschliche *Überprüfung* des Beweises obsolet (anders als bei den meisten anderen Anwendungen generativer KI).

## Fazit

- ▶ Mathematik, wie sie die meisten MathematikerInnen beherrschen, wird vermutlich bald von Maschinen besser beherrscht.

## Fazit

- ▶ Mathematik, wie sie die meisten MathematikerInnen beherrschen, wird vermutlich bald von Maschinen besser beherrscht.
- ▶ Ist Mathematik dann “obsolet” (wie Avigad sagt)?

## Fazit

- ▶ Mathematik, wie sie die meisten MathematikerInnen beherrschen, wird vermutlich bald von Maschinen besser beherrscht.
- ▶ Ist Mathematik dann “obsolet” (wie Avigad sagt)?
- ▶ Ja, wenn... “a mathematician is a machine for turning coffee into theorems.” (Alfred Renyi)

## Fazit

- ▶ Mathematik, wie sie die meisten MathematikerInnen beherrschen, wird vermutlich bald von Maschinen besser beherrscht.
- ▶ Ist Mathematik dann “obsolet” (wie Avigad sagt)?
- ▶ Ja, wenn... “a mathematician is a machine for turning coffee into theorems.” (Alfred Renyi)
- ▶ Nicht unbedingt, wenn “In re mathematica ars proponendi quaestionem pluris facienda est quam solvendi.” (Georg Cantor)

# Fazit

- ▶ Mathematik, wie sie die meisten MathematikerInnen beherrschen, wird vermutlich bald von Maschinen besser beherrscht.
- ▶ Ist Mathematik dann “obsolet” (wie Avigad sagt)?
- ▶ Ja, wenn... “a mathematician is a machine for turning coffee into theorems.” (Alfred Renyi)
- ▶ Nicht unbedingt, wenn “In re mathematica ars proponendi quaestionem pluris facienda est quam solvendi.” (Georg Cantor)
- ▶ Vermittlung zwischen Maschine und Gesellschaft

# Fazit

- ▶ Mathematik, wie sie die meisten MathematikerInnen beherrschen, wird vermutlich bald von Maschinen besser beherrscht.
- ▶ Ist Mathematik dann “obsolet” (wie Avigad sagt)?
- ▶ Ja, wenn... “a mathematician is a machine for turning coffee into theorems.” (Alfred Renyi)
- ▶ Nicht unbedingt, wenn “In re mathematica ars proponendi quaestionem pluris facienda est quam solvendi.” (Georg Cantor)
- ▶ Vermittlung zwischen Maschine und Gesellschaft
- ▶ Woher weiß man, ob die Maschine das beantwortet hat, was man gefragt hat?

# Fazit

- ▶ Mathematik, wie sie die meisten MathematikerInnen beherrschen, wird vermutlich bald von Maschinen besser beherrscht.
- ▶ Ist Mathematik dann “obsolet” (wie Avigad sagt)?
- ▶ Ja, wenn... “a mathematician is a machine for turning coffee into theorems.” (Alfred Renyi)
- ▶ Nicht unbedingt, wenn “In re mathematica ars proponendi quaestionem pluris facienda est quam solvendi.” (Georg Cantor)
- ▶ Vermittlung zwischen Maschine und Gesellschaft
- ▶ Woher weiß man, ob die Maschine das beantwortet hat, was man gefragt hat?
- ▶ Mathematik oft Ausdruck weiterer (programmatischer, philosophischer, ästhetischer, praktischer) Vorstellungen; diesen Ausdruck kann nur das Individuum selbst finden.

# Fazit

- ▶ Mathematik, wie sie die meisten MathematikerInnen beherrschen, wird vermutlich bald von Maschinen besser beherrscht.
- ▶ Ist Mathematik dann “obsolet” (wie Avigad sagt)?
- ▶ Ja, wenn... “a mathematician is a machine for turning coffee into theorems.” (Alfred Renyi)
- ▶ Nicht unbedingt, wenn “In re mathematica ars proponendi quaestionem pluris facienda est quam solvendi.” (Georg Cantor)
- ▶ Vermittlung zwischen Maschine und Gesellschaft
- ▶ Woher weiß man, ob die Maschine das beantwortet hat, was man gefragt hat?
- ▶ Mathematik oft Ausdruck weiterer (programmatischer, philosophischer, ästhetischer, praktischer) Vorstellungen; diesen Ausdruck kann nur das Individuum selbst finden.
- ▶ Verständnis weiterhin nötig, wenn sich Entwicklung nicht ganz vom Menschen ablösen soll.

# Fazit

- ▶ Mathematik, wie sie die meisten MathematikerInnen beherrschen, wird vermutlich bald von Maschinen besser beherrscht.
- ▶ Ist Mathematik dann “obsolet” (wie Avigad sagt)?
- ▶ Ja, wenn... “a mathematician is a machine for turning coffee into theorems.” (Alfred Renyi)
- ▶ Nicht unbedingt, wenn “In re mathematica ars proponendi quaestionem pluris facienda est quam solvendi.” (Georg Cantor)
- ▶ Vermittlung zwischen Maschine und Gesellschaft
- ▶ Woher weiß man, ob die Maschine das beantwortet hat, was man gefragt hat?
- ▶ Mathematik oft Ausdruck weiterer (programmatischer, philosophischer, ästhetischer, praktischer) Vorstellungen; diesen Ausdruck kann nur das Individuum selbst finden.
- ▶ Verständnis weiterhin nötig, wenn sich Entwicklung nicht ganz vom Menschen ablösen soll.
- ▶ KI wird Mathematik vermutlich in sehr absehbarer Zeit stark verändern.