



Graduate-level scientific questions across biology, physics, and chemistry.

EDUCATION LE DEBOIL

Model	Diamond Set (Zero-shot CoT)
Claude 3.5 Sonnet	59.4%
GPT-4o	53.6%
Gemini 1.5 Pro	46.2%
GPT-4	35.7%
GPT-3.5	28.1%

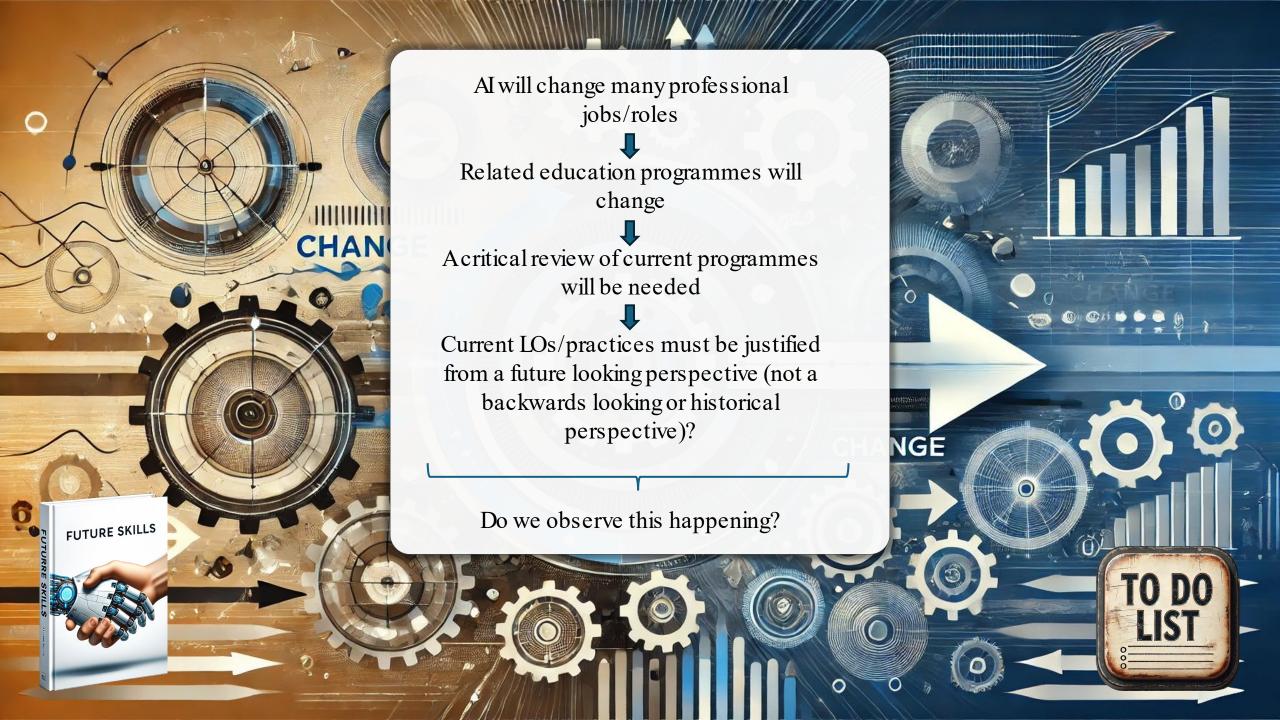
- PhD level experts 65%
- Highly-skilled non-expert 34% (Google + 30mins/question)

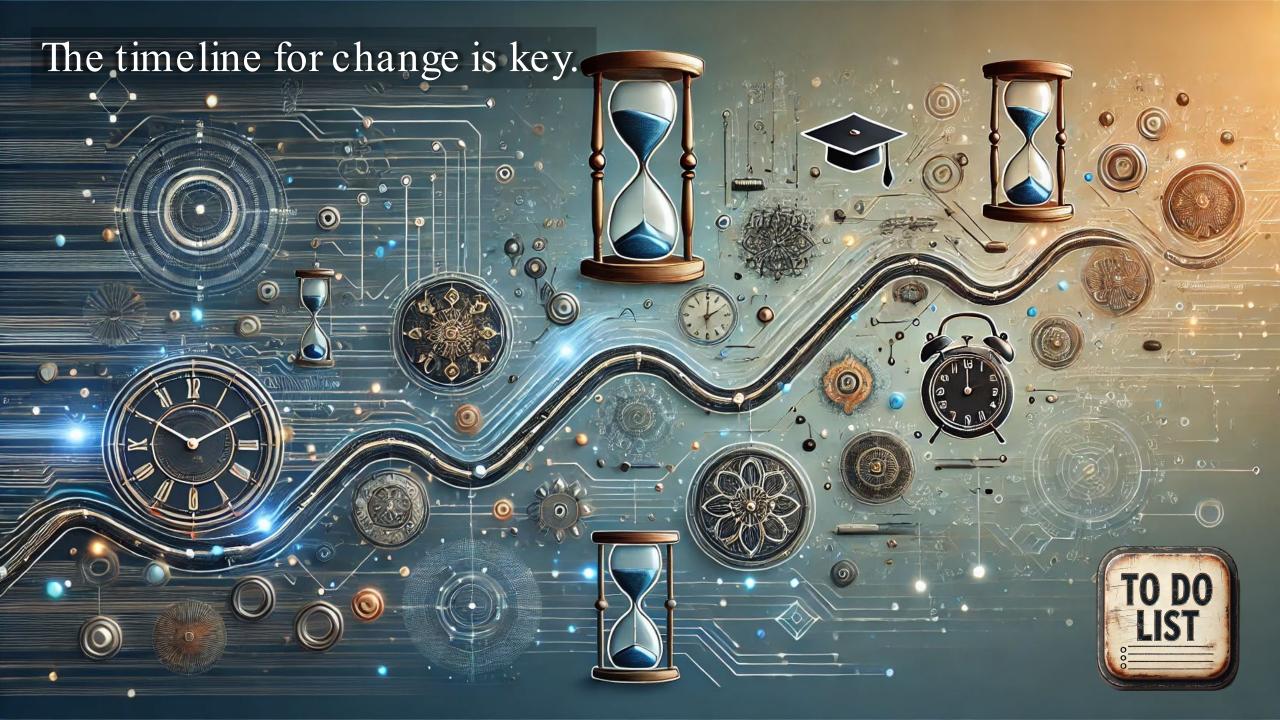
## **MATH**

Arithmetic reasoning benchmark of competition mathematics problems.

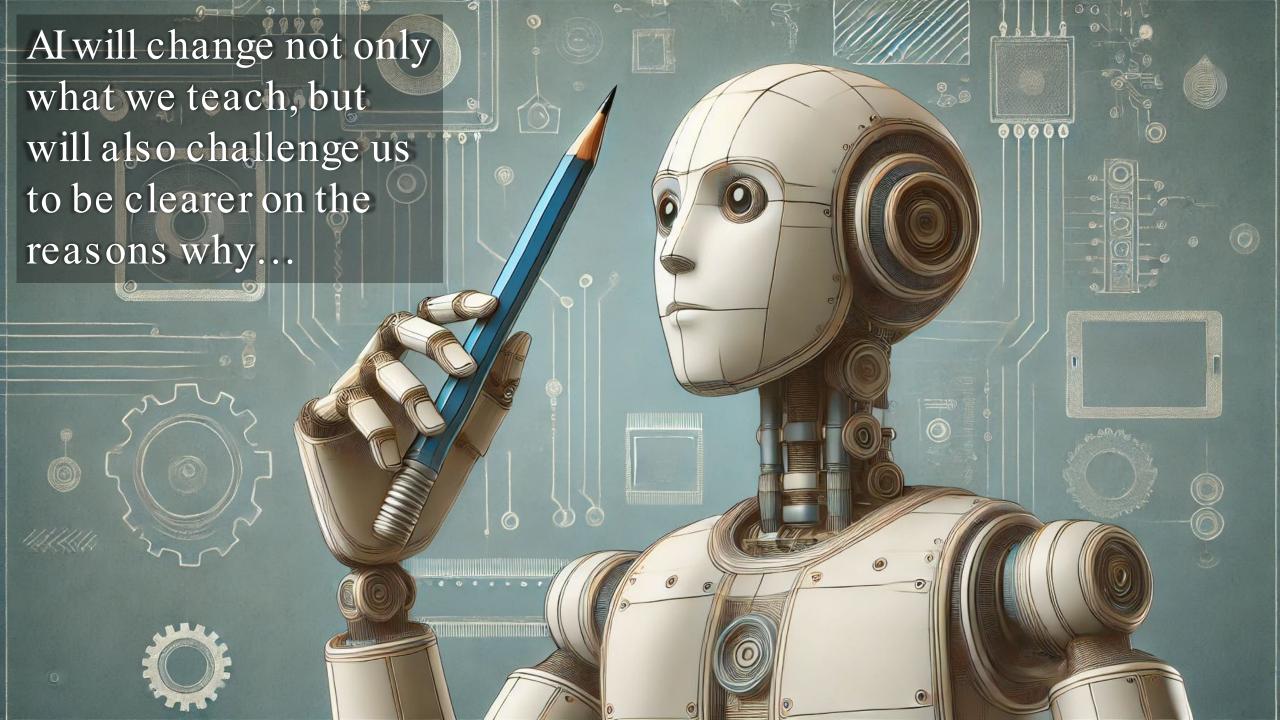
Model	Score
GPT-4o	76.6%
Claude 3.5 Sonnet	71.1%
Gemini 1.5 Flash	67.7%
GPT-4	52.9%
GPT-3.5	34.1%

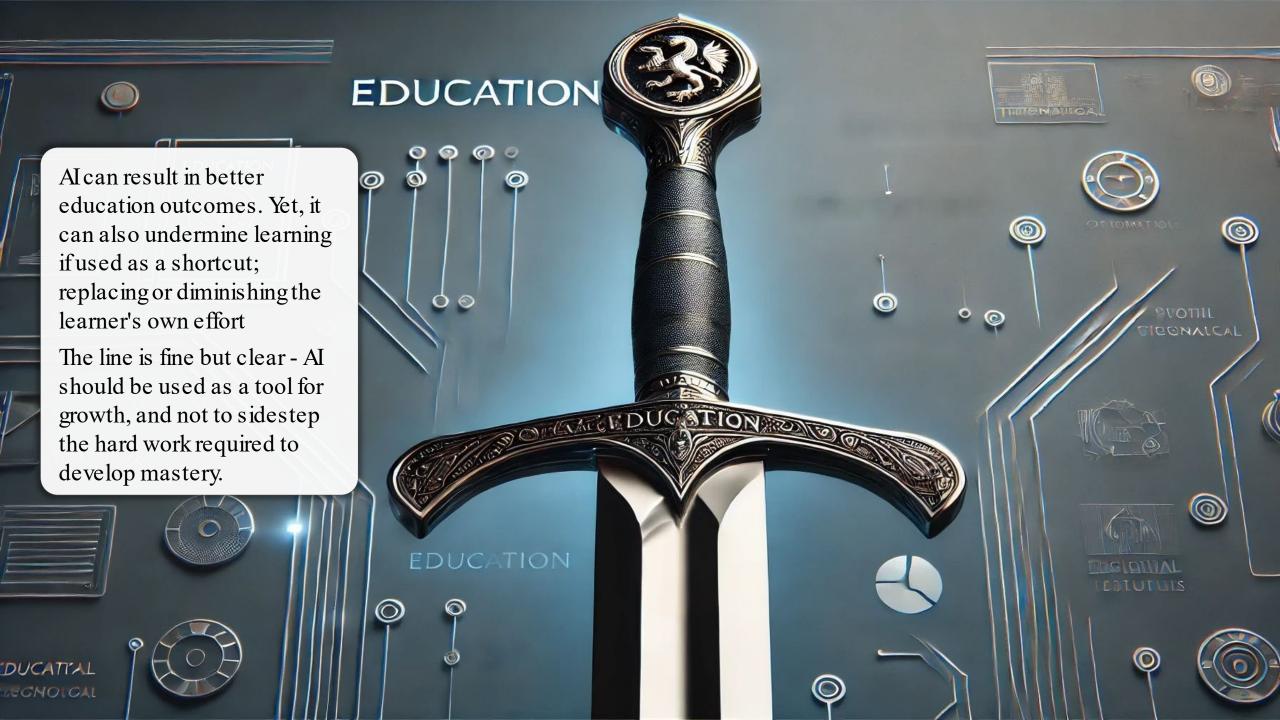
## AI will change many roles... Time spent using various types of skills by wage quintile in the United States, 2030, % 35 -35 40 Midpoint automation scenario, with generative Al acceleration Health professionals Basic Higher Physical and Social and Technological 30 cognitive cognitive emotional manual skills skills skills skills skills Health aides, technicians. Highest 25 and wellness 24 40 26 wage quintile STEM professionals 2nd 33 10 31 17 20 -Increasing labor demand Increasing labor demand and modest change of and high change of work 40 3rd 21 14 14 work activities activities 35 4th 21 18 13 13 Builders Managers Creatives and arts management Lowest wage 52 12 quintile Transportation Property services maintenance Business and legal professionals References Mechanical Community installation Agriculture Education and · McKinsey & Company (2023), The state of Al in services and repair workforce 2023: Generative Al's breakout year training • OECD (2023), OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market Production ANOWLED'CE TRAU &







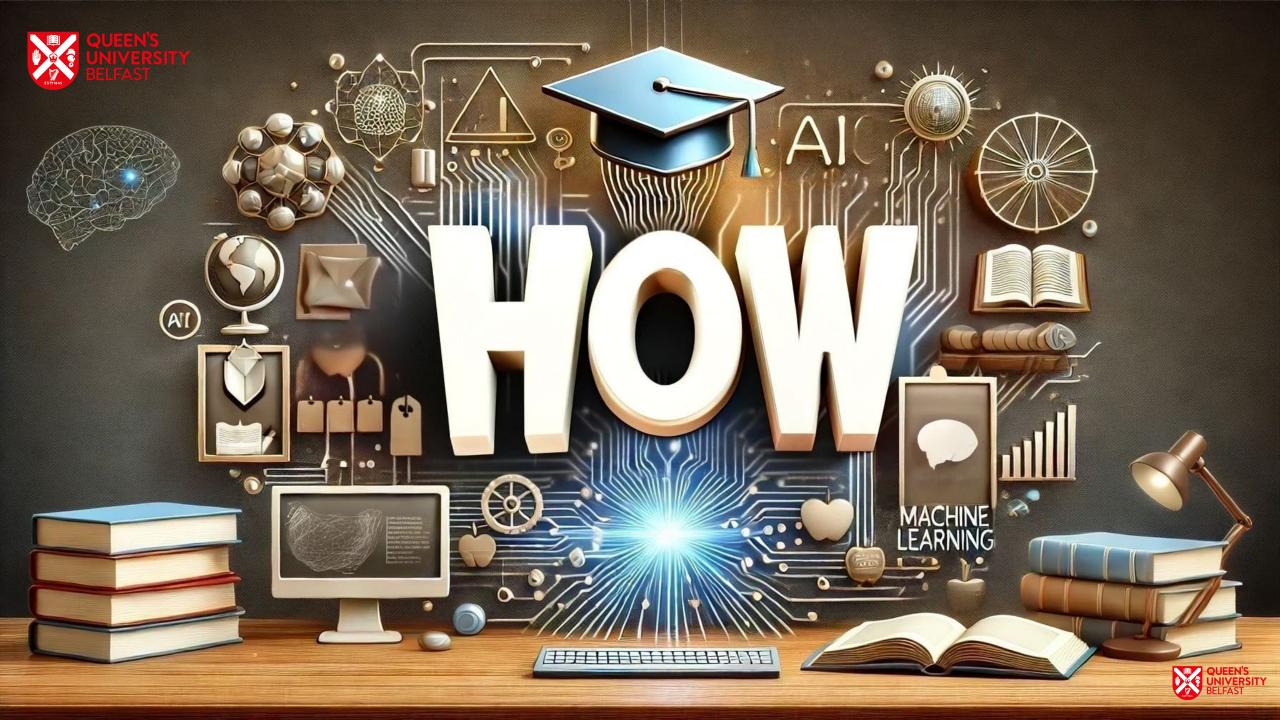




We need to actively engage our students in meaningful discussions about the responsibilities of being a learner and the purpose behind the learning tasks we assign.

Alongside this, we will need a small number of robust checkpoints or progress gates.





This is a shared problem, with a joint solution required.

- Institutional: Policies, Systems, Capabilities, Support.
- Programme:
  Skills and
  Competencies,
  Course design,
  Benchmarking
  (PSRBs).
- Individual:
  Learner
  engagement,
  activities and
  support.





## Institutional:

- Policies: Values, Equity, Inclusion, Accountability, Safety.
- Horizon Scanning: Al Capabilities, Jobs & Skills.
- Innovation Support: Resourcing, Scaling.
- Knowledge Base: Case studies, Reports, Guides.
- Training and Support: Staff, Students.
- Systems: Staff, Student, Services.









## Individual:

- Virtual course assistant: Interactive course guides, provide real-time Q&A.
- Build practice content: Produce quizzes and tests, create practice exams.
- Enhance feedback: Expand feedback, generate suggested improvements.
- Personal tutor: Explain course concepts, offer learner guidance.
- Improve accessibility: Reformat course materials, multilingual translations.





