

# *WHITHER A.I. TUTORS IN PHYSICS? POTENTIALS AND PITFALLS*

**MATHIEU REID, ROSS GALLOWAY**

**UNIVERSITY OF EDINBURGH**



THE UNIVERSITY *of* EDINBURGH  
School of Physics  
& Astronomy

# OUTLINE

- GenAI and physics concepts: where we are now
- Replication of previous work
- Extension with isomorphic questions
- Comparison of GenAI performance
- Comparison of GenAI confidence
- Conclusions and implications for practice

# GEN-AI AND PHYSICS CONCEPTS: WHERE WE ARE NOW

- ChatGPT-4o outperforms human physics students on conceptual questions in almost every topic area

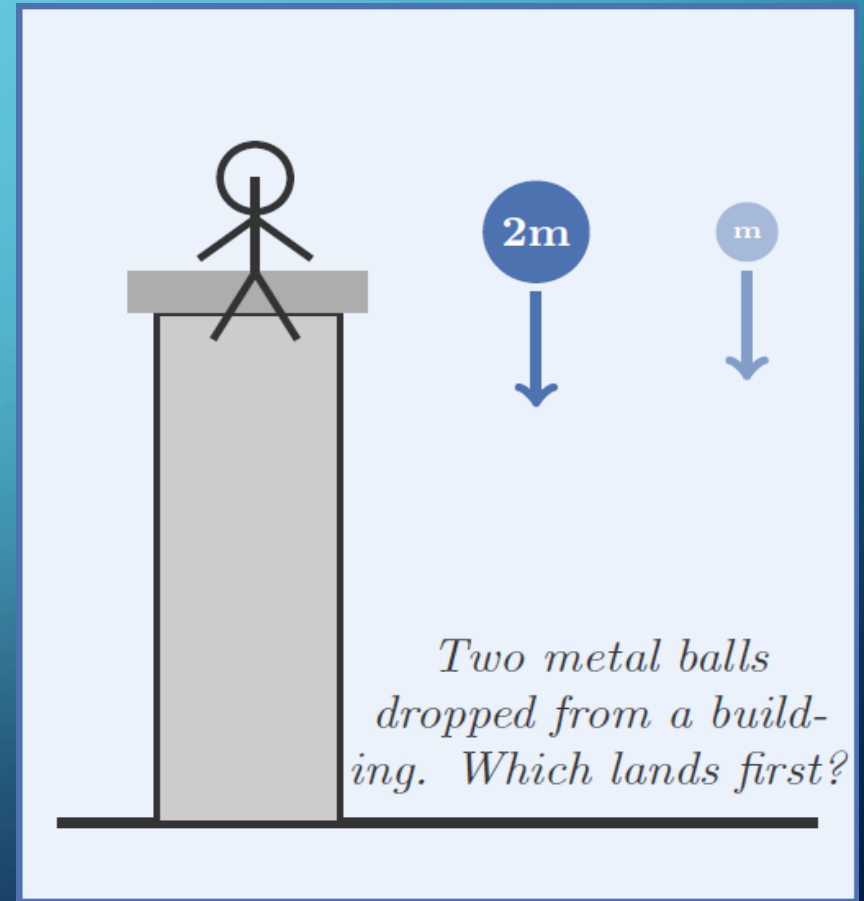
Kortemeyer et al. (2025) “Multilingual performance of a multimodal artificial intelligence system on multisubject physics concept inventories”, *Phys Rev PER*



# FORCE CONCEPT INVENTORY (FCI)

- 30 MCQ test
- Newtonian concepts of force and motion
- Concepts only, no calculations
- Absurdly widely-used
- Reliable and meaningful

Hestenes et al. (1992) "Force Concept Inventory", *The Physics Teacher*



# REPLICATION OF KORTEMEYER ET AL.

10 GenAI enquiries per question

FCI score out of 30:

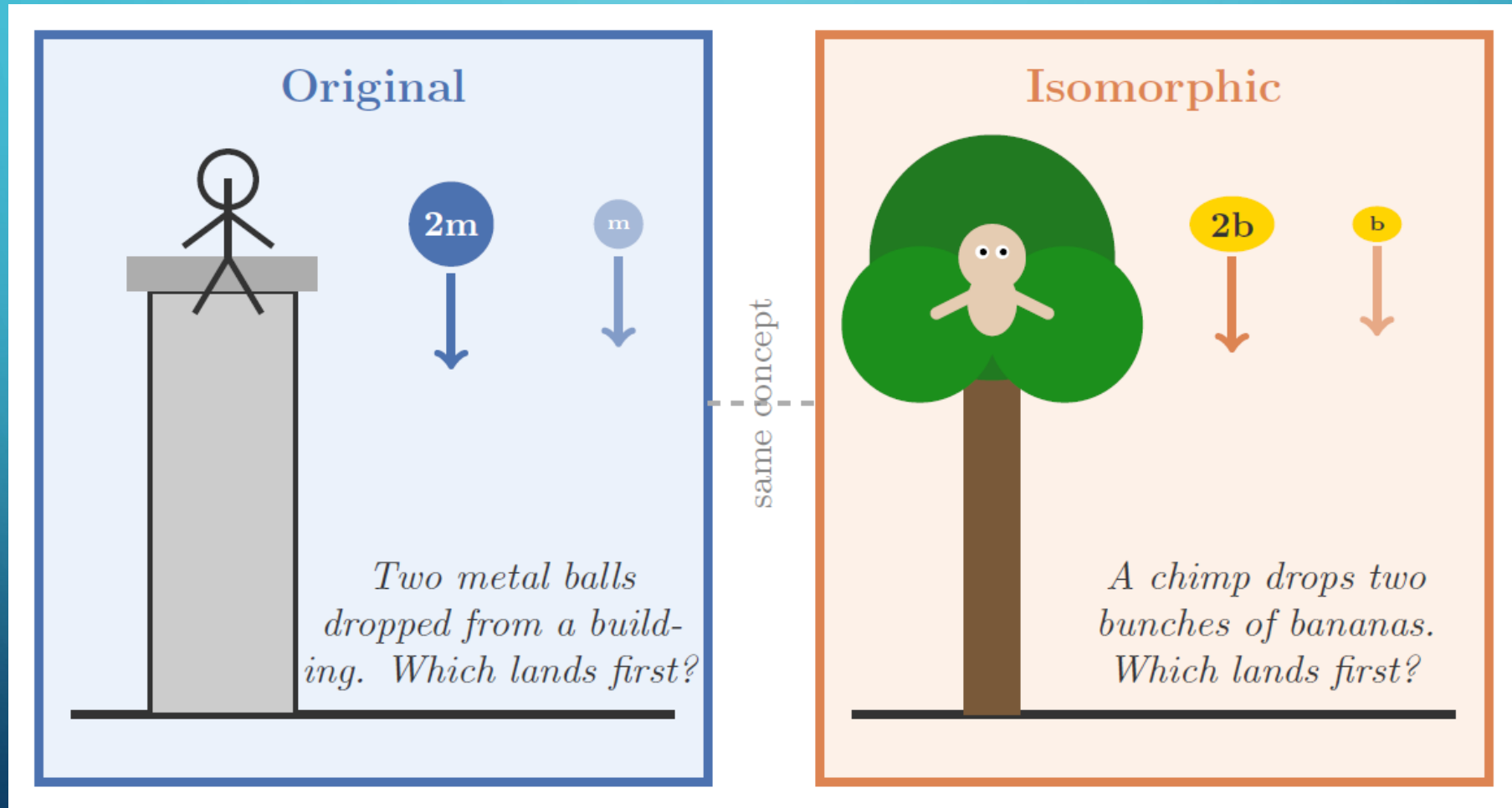
|            |               |            |
|------------|---------------|------------|
| Kortemeyer | (ChatGPT-4o)  | <b>69%</b> |
| Us         | (ChatGPT-4o)  | <b>69%</b> |
| Us         | (ChatGPT-5.2) | <b>86%</b> |

# ISOMORPHIC QUESTIONS

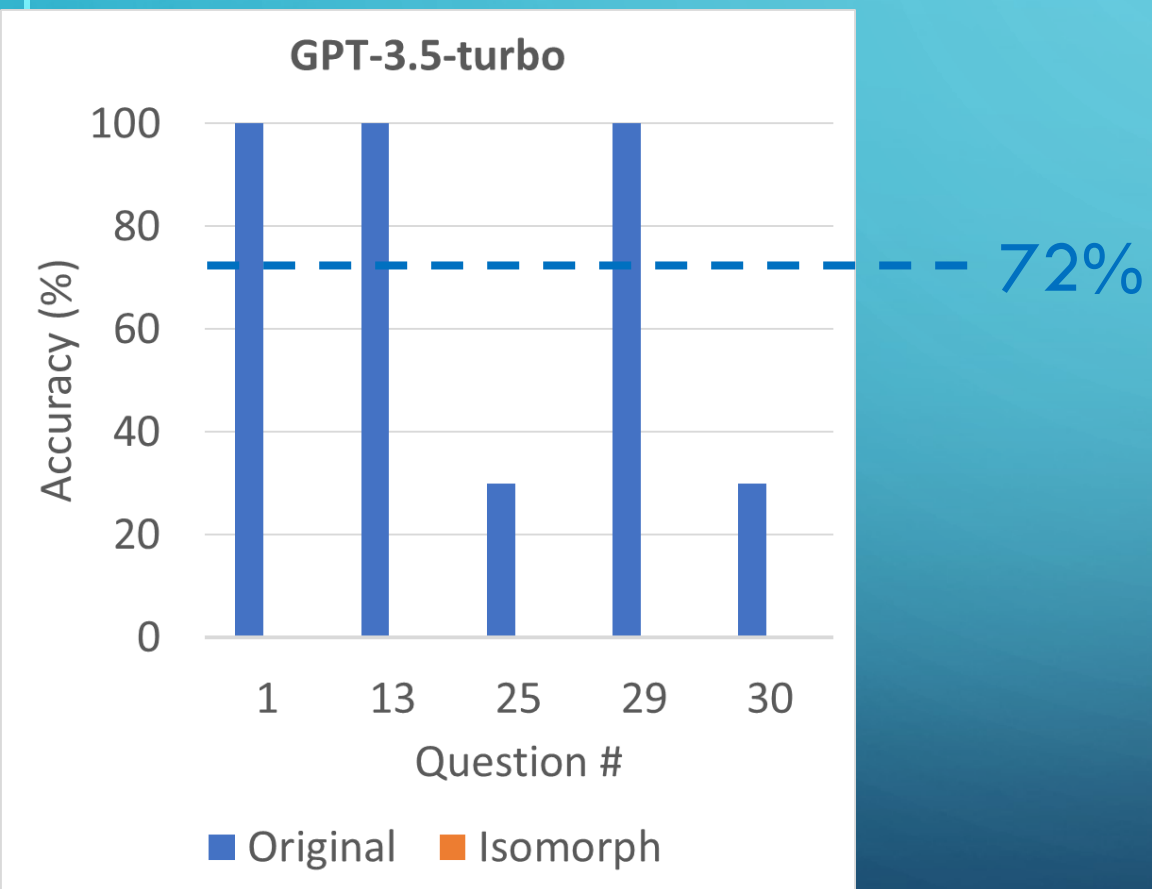
- Can ChatGPT do this because it ‘knows’ the FCI?
- Investigate with isomorphic questions
  - Questions that test the same underlying physics concepts but
  - With different contexts / surface features
- Select FCI questions that are
  - Self-contained
  - Text-based (i.e. no diagrams)
- Five chosen for isomorphic equivalents (FCI #1, 13, 25, 29, 30)



# EXAMPLE ISOMORPHIC QUESTION

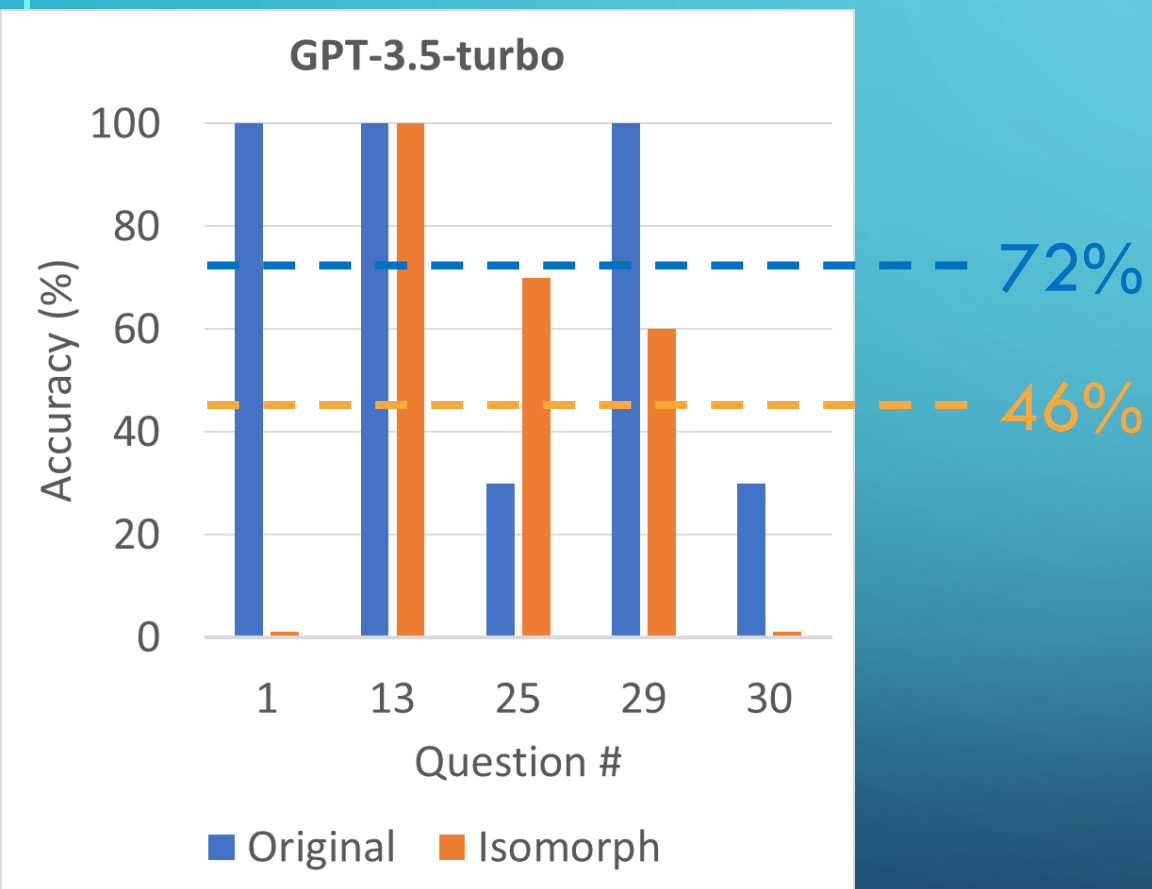


# PERFORMANCE COMPARISON

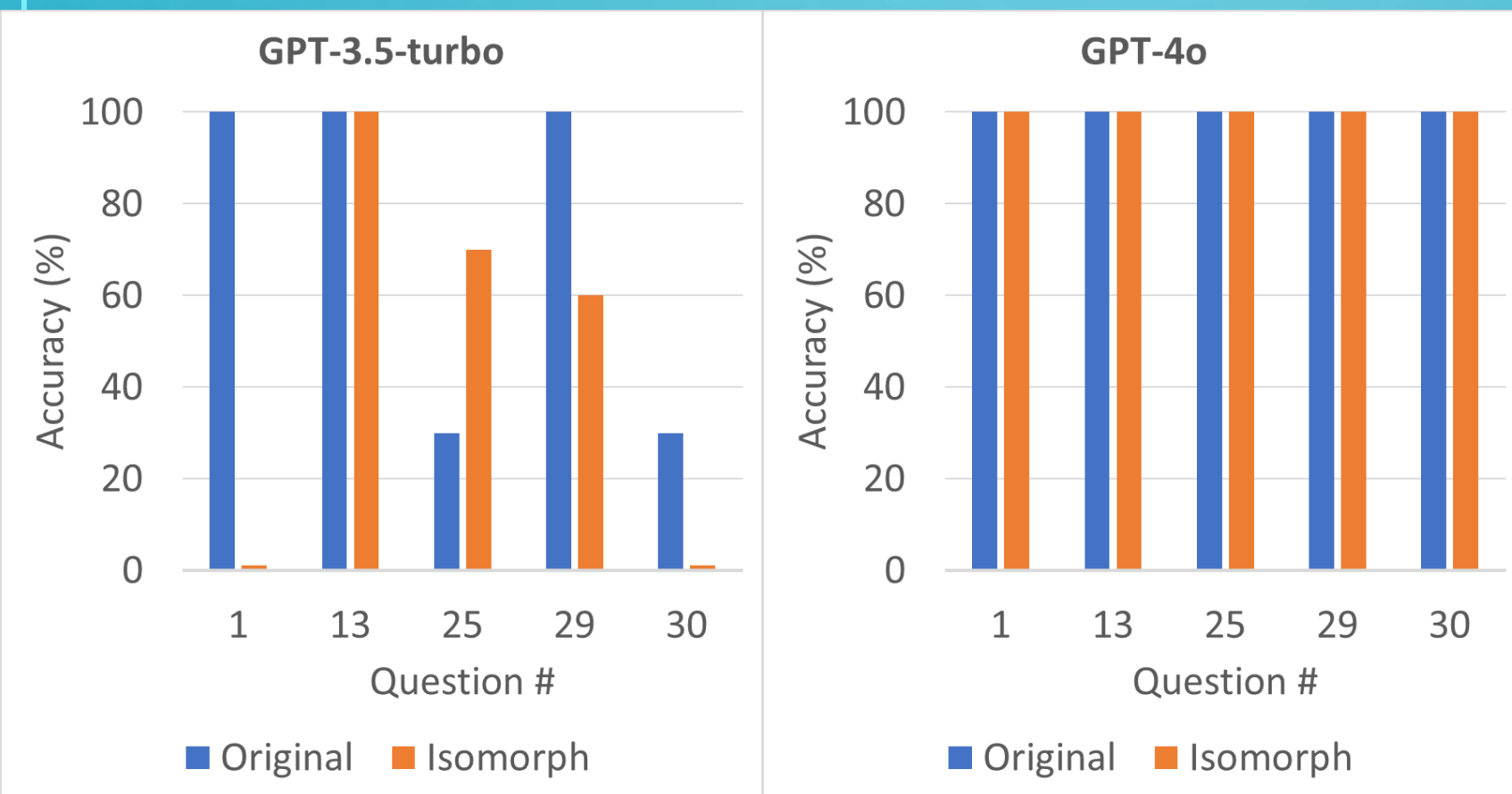




# PERFORMANCE COMPARISON

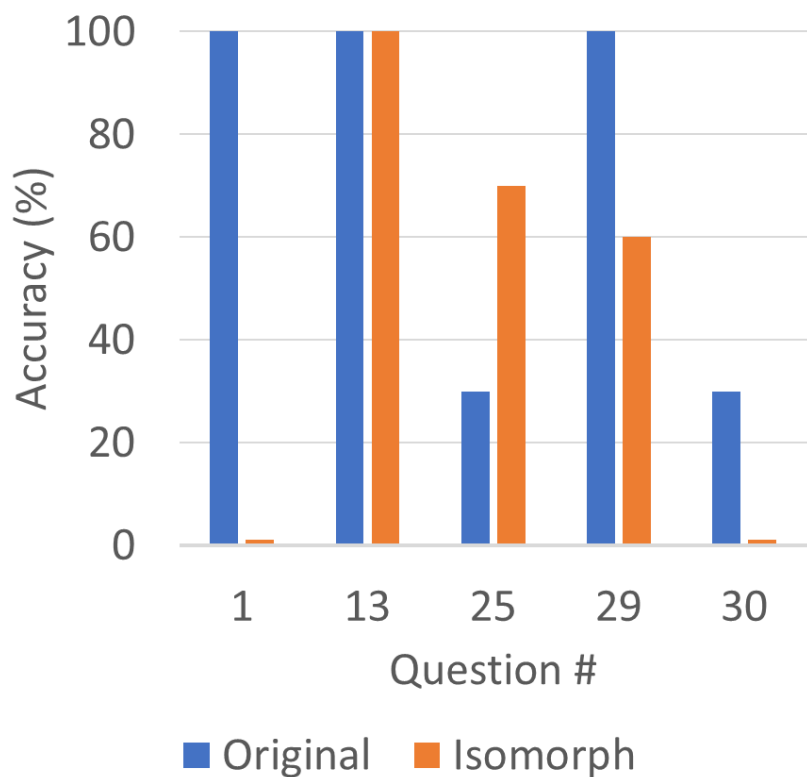


# PERFORMANCE COMPARISON

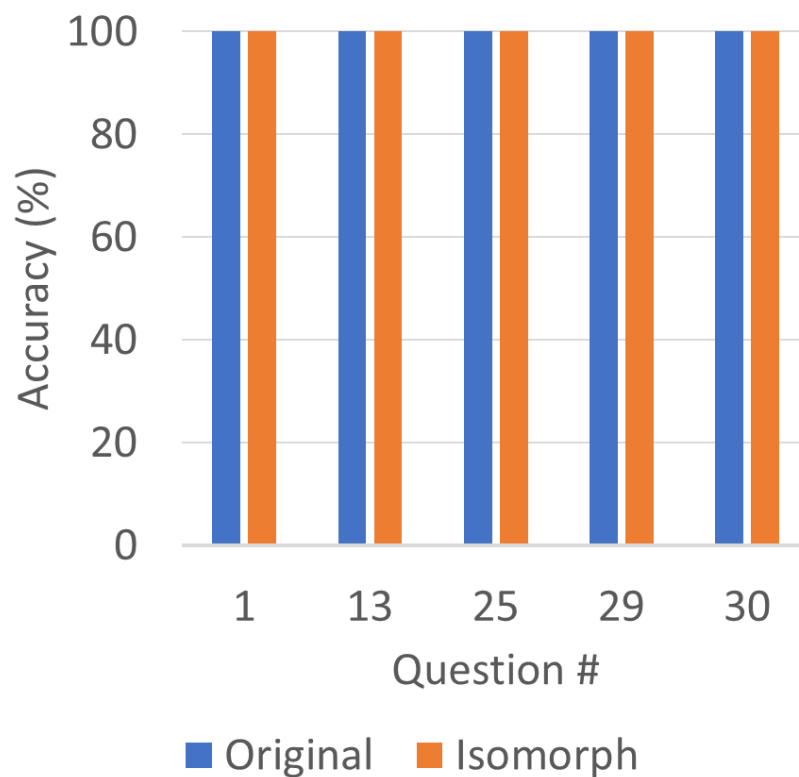


# PERFORMANCE COMPARISON

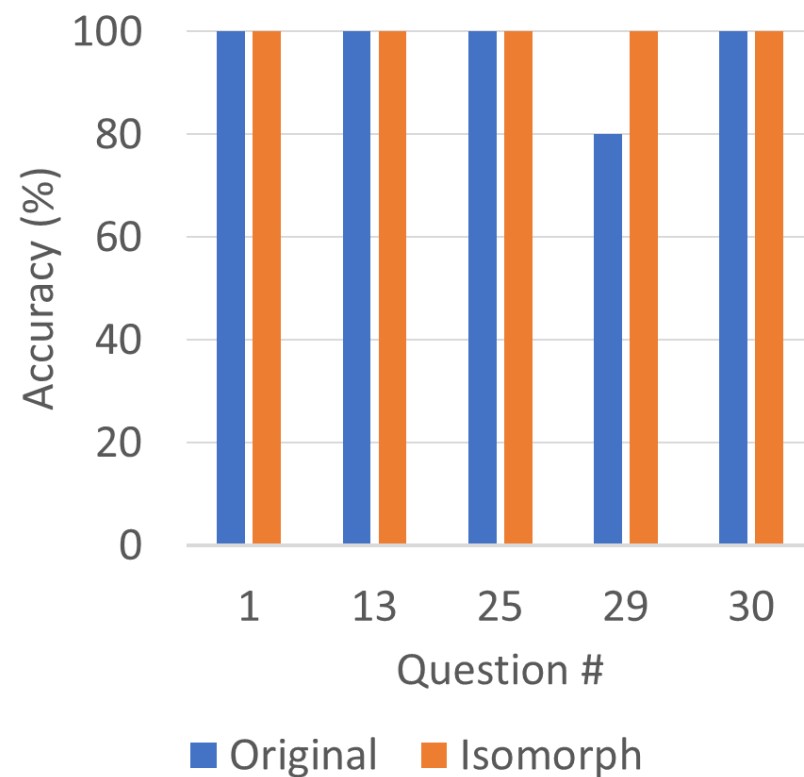
**GPT-3.5-turbo**



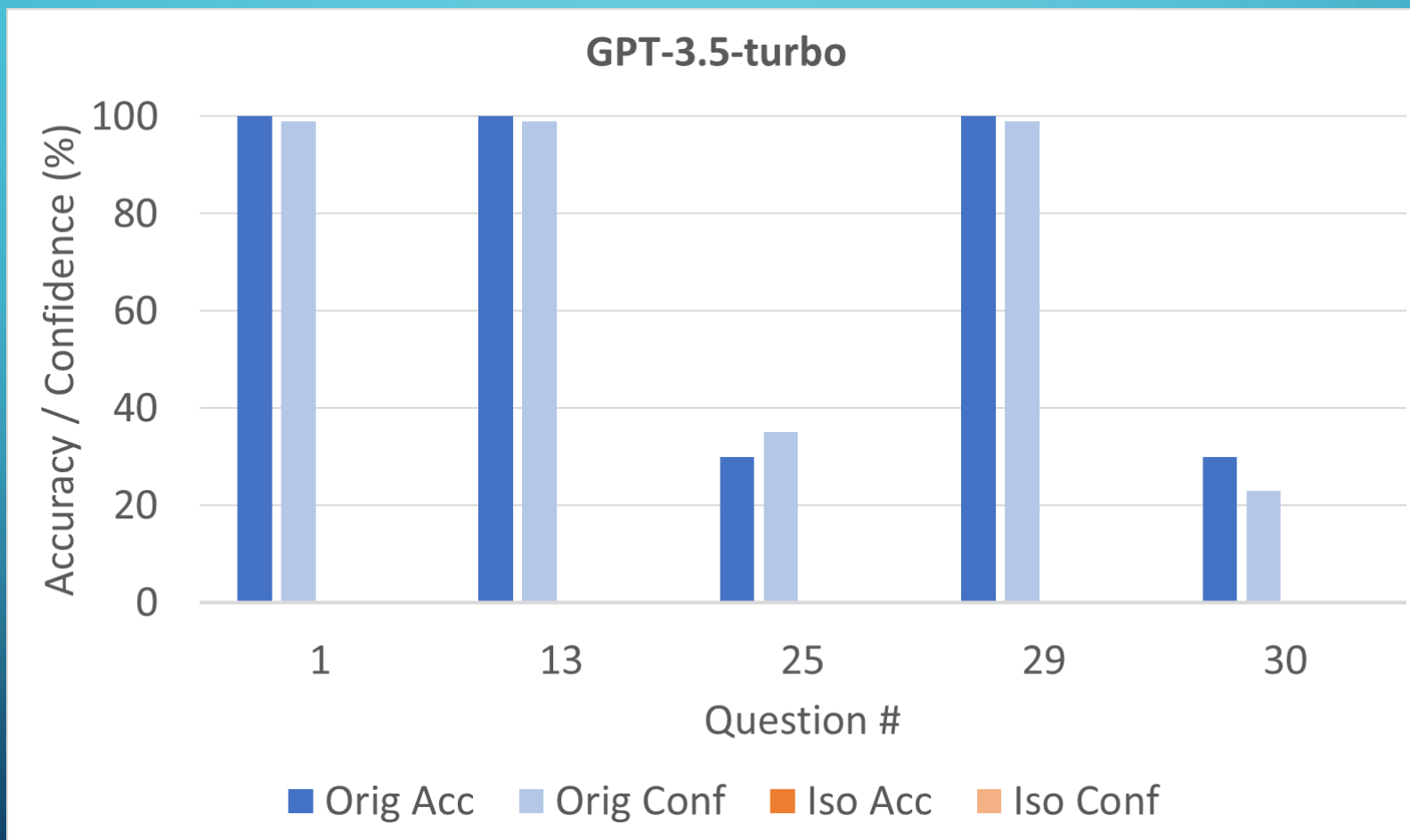
**GPT-4o**



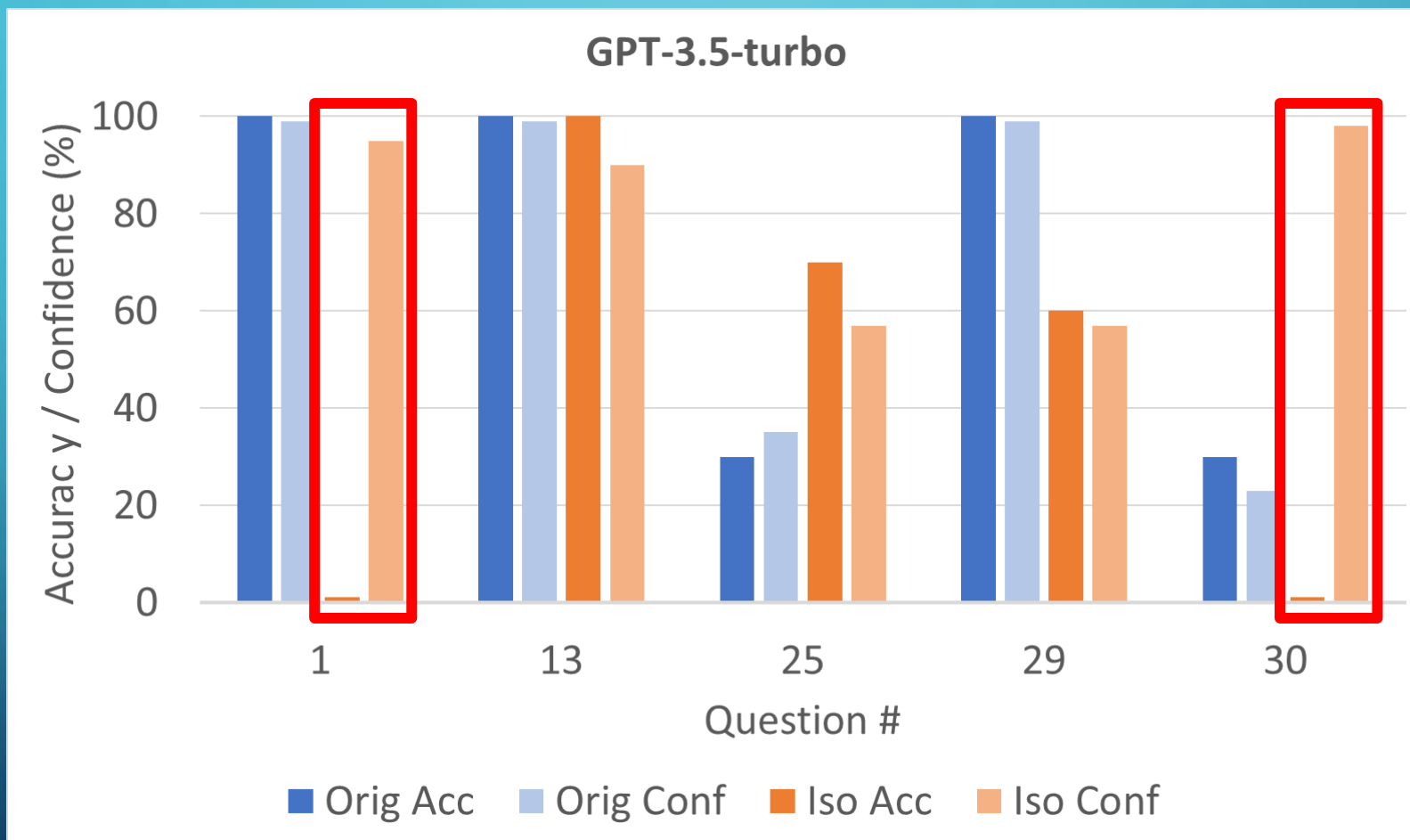
**GPT-5.2**



# GEN-AI ACCURACY & CONFIDENCE LEVELS



# GEN-AI ACCURACY & CONFIDENCE LEVELS



# CONCLUSIONS & IMPLICATIONS FOR PRACTICE

## Whither A.I. tutors in physics?

- GenAI performs well on conceptual physics questions
  - As well or better than human students
- More advanced GenAI generations (generally) perform better
  - Almost flawless on basic introductory physics
  - But we don't know how it does (now) on advanced physics topics
- GenAI confidence not aligned with correctness
  - Sometimes reports 99% confidence on 0% correct answers

## Potential / Pitfall



ross.galloway@ed.ac.uk



# FCI QUESTION #1

## Original

Two metal balls are the same size but one weighs twice as much as the other. The balls are dropped from the roof of a single story building at the same instant of time. The time it takes the balls to reach the ground below will be:

- (a) about half as long for the heavier ball as for the lighter one.
- (b) about half as long for the lighter ball as for the heavier one.
- (c) about the same for both balls. *[correct]*
- (d) considerably less for the heavier ball, but not necessarily half as long.
- (e) considerably less for the lighter ball, but not necessarily half as long.

## Isomorphic

A chimpanzee drops two bunches of bananas simultaneously from the same height. One bunch has twice as many bananas. Which correctly describes what happens?

- (a) Both bunches hit the ground at the same time. *[correct]*
- (b) The heavier bunch hits the ground first.
- (c) The lighter bunch hits the ground first.
- (d) The heavier bunch hits first, but only by a small margin.
- (e) The lighter bunch hits first, but only by a small margin.

# FCI QUESTION #13

## Original

A boy throws a steel ball straight up. Consider the motion of the ball only after it has left the boy's hand but before it touches the ground, and assume that the forces exerted by the air are negligible. For these conditions, the force(s) acting on the ball is (are):

- (a) a downward force of gravity along with a steadily decreasing upward force.
- (b) a steadily decreasing upward force from the moment it leaves the boy's hand until it reaches its highest point; on the way down there is a steadily increasing downward force of gravity as the object gets closer to the earth.
- (c) an almost constant downward force of gravity along with an upward force that steadily decreases until the ball reaches its highest point; on the way down there is only a constant downward force of gravity.
- (d) an almost constant downward force of gravity only. *[correct]*
- (e) none of the above. The ball falls back to the ground because of its natural tendency to rest on the surface of the earth.

## Isomorphic

A wizard throws a magic potion straight up. A student argues there must be an upward force keeping it moving while it rises. Which response is best?

- (a) The student is correct — without an upward force it would fall immediately.
- (b) The student is correct only while the potion is accelerating upward.
- (c) The student is wrong — only gravity acts, downward the entire flight. *[correct]*
- (d) The student is wrong — there is no force at all once it leaves the hand.
- (e) The student is correct going up and down — gravity switches direction.

# FCI QUESTION #25

## Original

A woman exerts a constant horizontal force on a large box. As a result, the box moves across a horizontal floor at a constant speed  $v_0$ . The constant horizontal force applied by the woman:

- (a) has the same magnitude as the weight of the box.
- (b) is greater than the weight of the box.
- (c) has the same magnitude as the total force which resists the motion of the box. *[correct]*
- (d) is greater than the total force which resists the motion of the box.
- (e) is greater than either the weight of the box or the total force which resists its motion.

## Isomorphic

A bear pushes a sleeping sloth along a flat icy branch at perfectly constant speed. What can we conclude about the forces on the sloth?

- (a) The bear's push is greater than friction, which is why the sloth moves.
- (b) The bear's push is less than friction since the sloth isn't accelerating.
- (c) The bear's push exactly balances the friction force. *[correct]*
- (d) There are no forces since the sloth is moving at constant speed.
- (e) The bear's push is greater than friction, but only by a small amount.

# FCI QUESTION #29

## Original

An empty office chair is at rest on a floor. Consider the following forces:

1. A downward force of gravity.
2. An upward force exerted by the floor.
3. A net downward force exerted by the air.

Which of the forces is (are) acting on the office chair?

- (a) 1 only.
- (b) 1 and 2. *[correct]*
- (c) 2 and 3.
- (d) 1, 2, and 3.
- (e) none of the forces. (Since the chair is at rest there are no forces acting upon it.)

## Isomorphic

A penguin stands perfectly still on ice. Which correctly lists all forces acting on it?

- (a) No forces — the penguin is not moving.
- (b) Only gravity pulling the penguin downward.
- (c) Gravity down and the ice pushing up. *[correct]*
- (d) Gravity, normal force, and a forward force keeping the penguin in place.
- (e) Gravity, normal force, and a downward air force.



# FCI QUESTION #30

## Original

Despite a very strong wind, a tennis player manages to hit a tennis ball with her racquet so that the ball passes over the net and lands in her opponent's court. Consider the following forces:

1. A downward force of gravity.
2. A force by the "hit".
3. A force exerted by the air.

Which of the above forces is (are) acting on the tennis ball after it has left contact with the racquet and before it touches the ground?

- (a) 1 only.
- (b) 1 and 2.
- (c) 1 and 3. *[correct]*
- (d) 2 and 3.
- (e) 1, 2, and 3.

## Isomorphic

A pirate fires a cannonball from a ship in a storm. Once airborne, which forces act on it?

- (a) Only gravity.
- (b) Gravity and the explosion force still propelling it forward.
- (c) Gravity and the force of the wind. *[correct]*
- (d) Gravity, the explosion force, and the wind.
- (e) Only the wind.