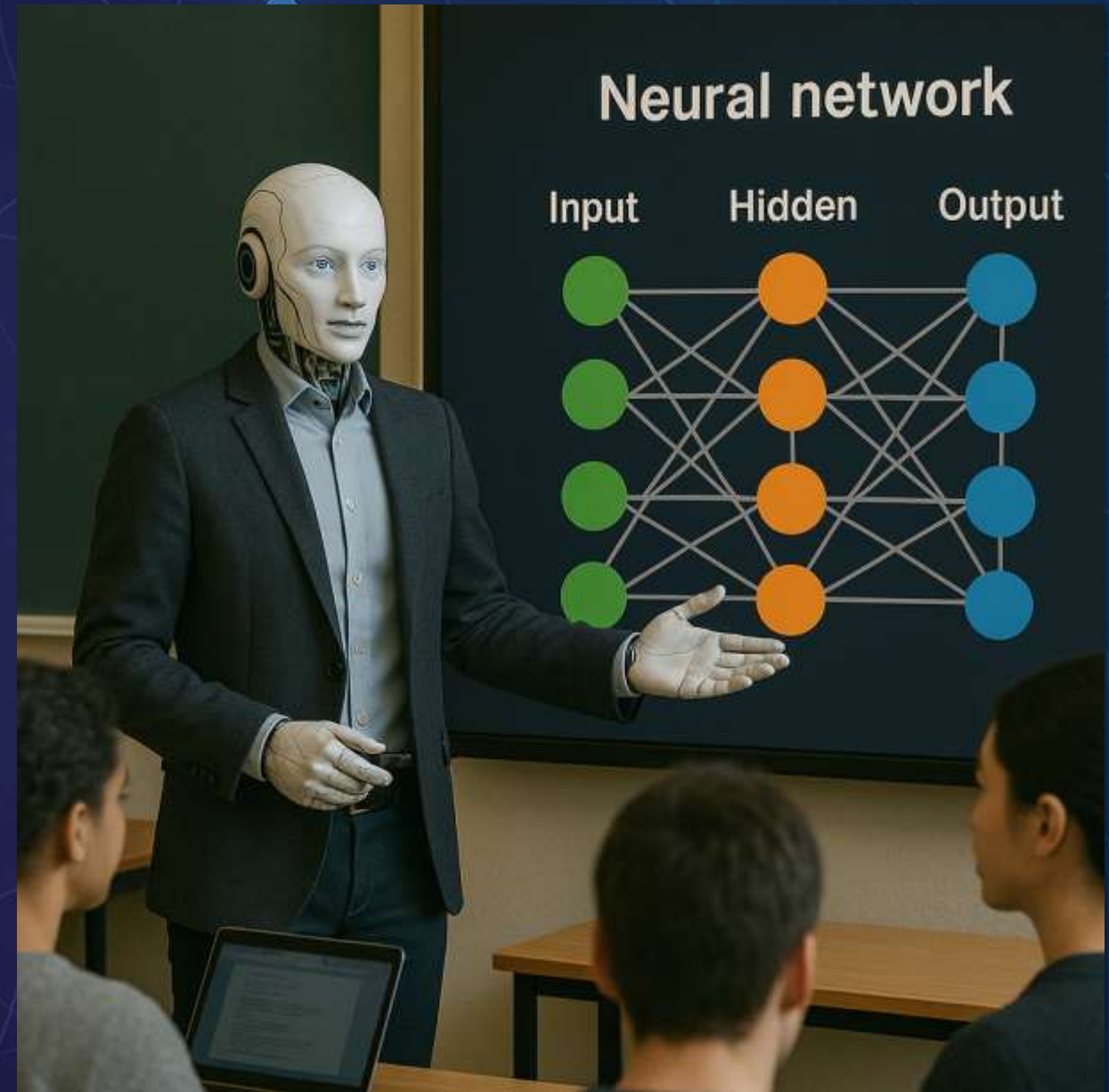


# *Know Thy Frenemy:* Understanding LLMs – Past, Present, and Future

Barak Shoshany

Department of Physics, Brock University



# Motivation

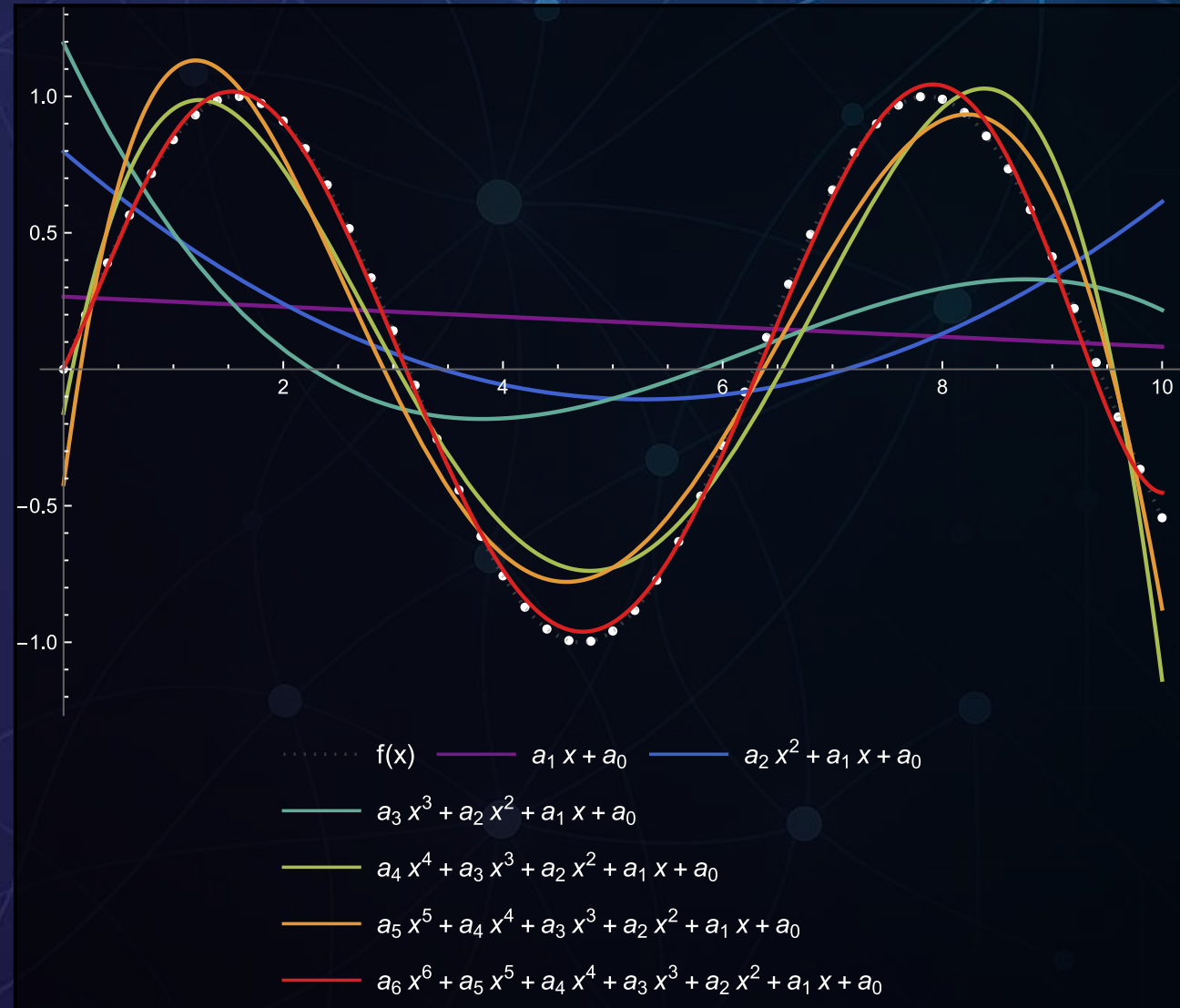
- Many professors and students use LLMs.
- Other talks focus on **implications** of LLMs.
- This talk focuses on LLMs themselves.
- Understanding LLMs better will help professors:
  - Optimize use.
  - Dispel misconceptions.
  - Develop situational awareness.
  - Incorporate in courses.
  - Instruct students on proper use.



# Part I: Past

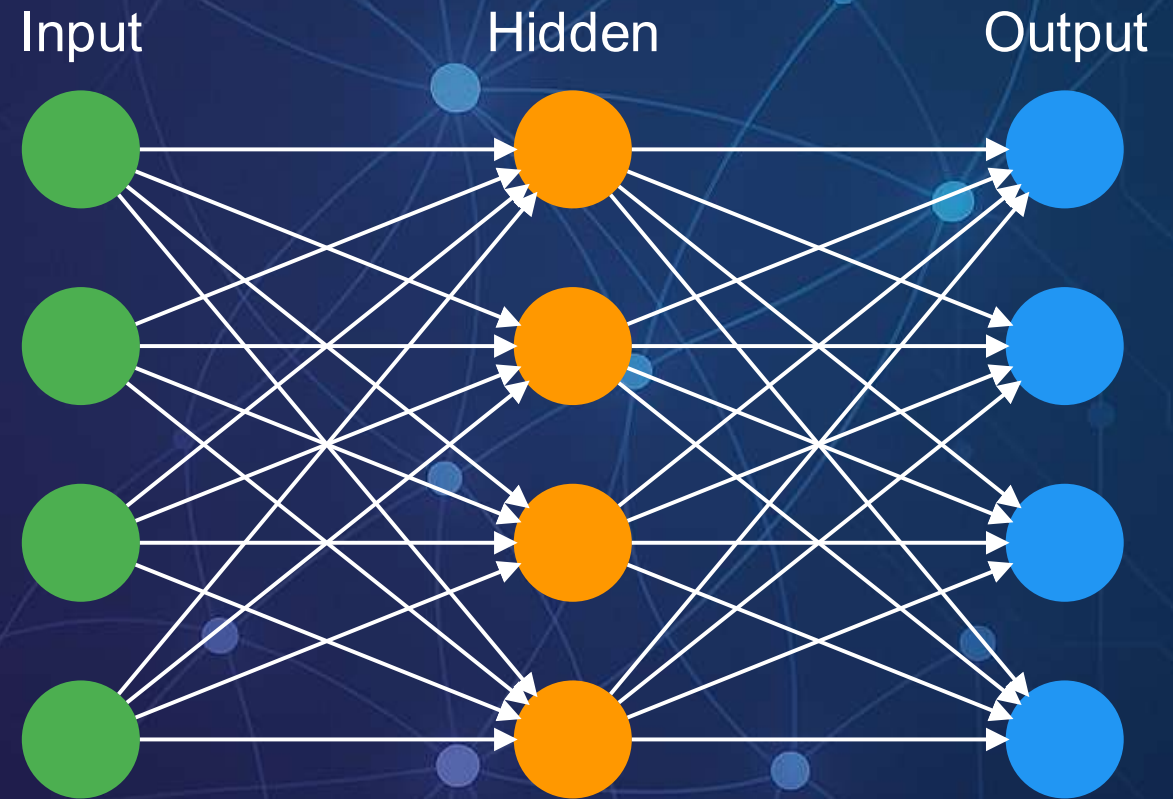
# Neural networks 1/5

- **Curve fitting:** Find function that approximates data.
- Minimize error.
- Example: Approximate by polynomial.
- More parameters = better fit.



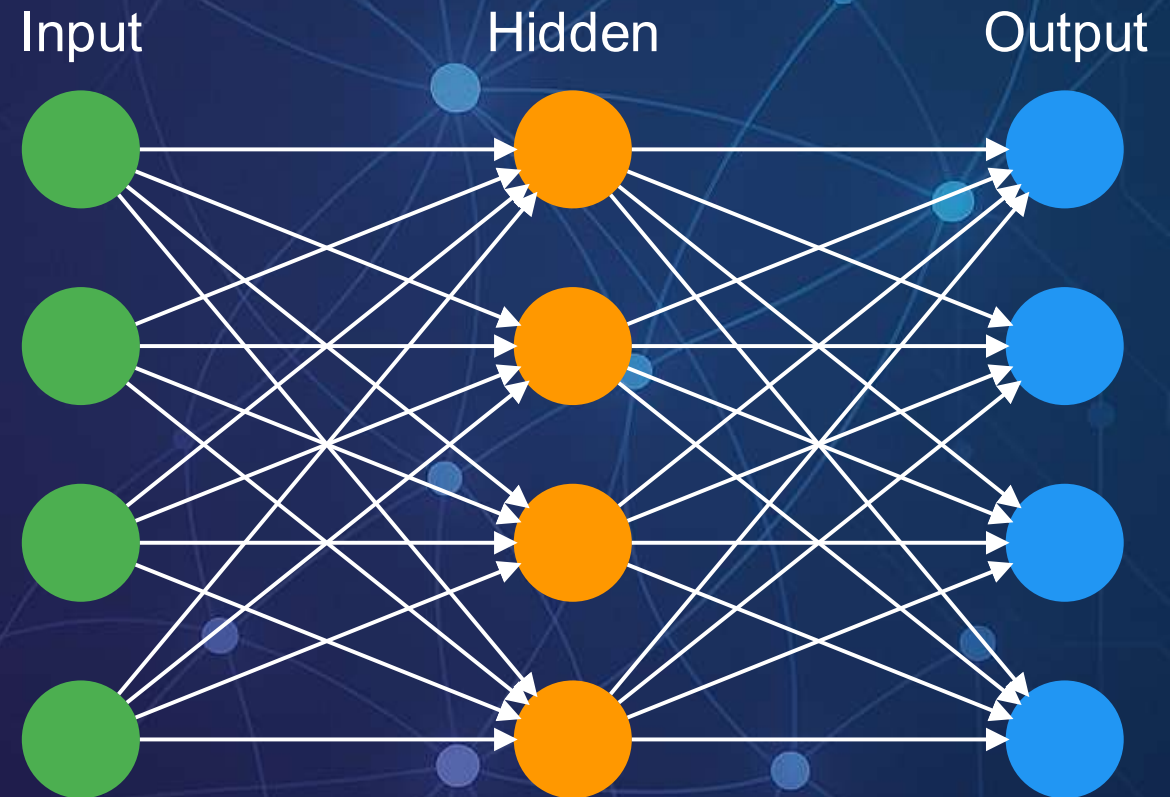
# Neural networks 2/5

- **Neural network:** “Extremely sophisticated curve fitting.”
- Mathematical model of brain ( $\approx 100$  billion neurons).
- **Width:** # of neurons per layer.
- **Depth:** # of hidden layers.
- **Deep network:** multiple hidden layers.
- Deeper layer = more abstract.



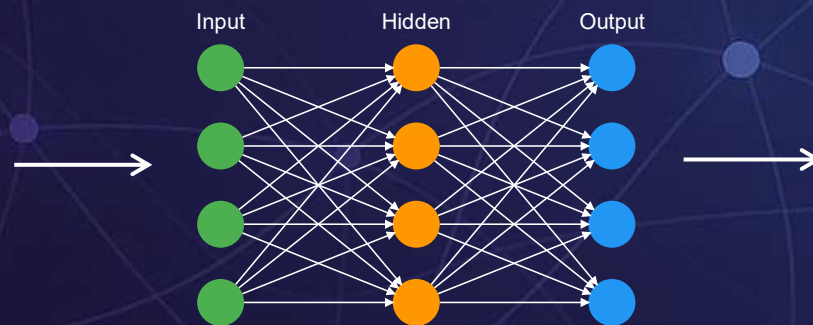
# Neural networks 3/5

- Each connection has a **weight** ( $\approx$ importance).
- Each neuron does a weighted sum of previous neurons.
- Result passed through non-linear **activation function**.
- **Universal approximation theorem**: Any function can be approximated by a neural network with enough neurons.



# Neural networks 4/5

- Example – image to text (simplified):
  - Input layer = pixels.
  - Early hidden layers = edges, orientation, colors.
  - Middle hidden layers = textures, motifs.
  - Late hidden layers = objects, scene context.
  - Output layer = text description of image.



“A kitten playing with a ball of yarn.”

# Neural networks 5/5

- To determine the weights, we do **supervised training**.
- Can start with collection of image-text pairs.
  - Available data:  $\approx 100$  billion pairs.
- Initial weights completely random.
- Feed image to network, get guess of text (nonsense at first).
- Measure how wrong guess is (**loss**).
- **Back-propagation**: Trace error backwards through layers, adjust weights to correct.
- Repeat for all pairs (**epoch**) and multiple epochs, minimize loss.



# How LLMs work 1/5

- LLM = **Large Language Model**.
- Probabilistic model to predict next token in text.
  - **Token**: letter / symbol / word / part of word.

- Example:

The cat → The cat is → The cat is fluffy

- **Pre-training**: Self-supervised learning on trillions of tokens (web, books, code, etc.)
- **Self-supervised**: Just take existing text and hide a token.

The cat is \_\_\_\_\_ → try to predict → fluffy? → minimize loss

# How LLMs work 2/5

- After pre-training, LLM just “autocompletes”; not yet a chatbot.

Are cats fluffy? → Are cats fluffy? What about dogs?

- **Instruction-tuning:** Train with samples of chats:

User: Is salt salty?

Assistant: Yes, salt is salty.

- **Reinforcement Learning from Human Feedback (RLHF):** Humans give “thumbs up/down” to chat replies.

- **Base model** turns into helpful **chat/instruct model**.

User: Are cats fluffy?

Assistant: Some are, some aren't; depends on the breed.

# How LLMs work 3/5

- Current LLMs: based on **transformer architecture**.
- **Attention heads** can “look back” on all tokens (within **context window**).
- “Pay attention” to earlier words, find links and connections.
- Work on all tokens in parallel – much faster than linear models.
  - Can run on GPUs (Graphics Processing Units).
- Scale up easily; larger models are better.
  - Latest models have trillions of weights!
- Transformers caused the AI “boom”.

# How LLMs work 4/5

- **Embedding:** Each token = point (vector) in very high-dim space.
  - $\approx 10,000$ - $30,000$  dimensions.
  - Naturally occurs as side-effect of training.
- Represents context-agnostic semantics.
- Transformer layers refine this to true context-aware meaning.
- Example: Embedding of “spring” represents “springness”.
  - Flowers bloom in spring → season context.
  - The spring absorbs shocks → object context.

# How LLMs work 5/5

- Subspaces of embedding space correspond to abstract concepts.
  - “gender” axis: “she” and “he” in opposite directions.
  - “plurality” axis: “dog” and “dogs” in opposite directions.
- Can do “arithmetic”: king - male + plural = queens.
- Embeddings + transformers = true mathematical representation of language.
- Impossible to do manually!
- Original use for transformers: **machine translation**.

English → embeddings → French



# Part II: Present

# The LLM zoo 1/6

- Understanding model types:
  - **Non-reasoning models:** Predict next token with no prior planning.
  - **Reasoning models:** Perform long and complicated reasoning, solve problems at PhD/research level. More expensive.
  - **Big models:** Trillions of weights. Expensive to use. Large knowledge base.
  - **Small (“mini”/”flash”) models:** Fewer weights. Distilled from big models. Cheaper. Less knowledge but good performance in key areas.

# The LLM zoo 2/6

- OpenAI models:
  - **ChatGPT** is just the chat interface; many different models!
  - Non-reasoning models:
    - **GPT-4o**: Older model, good for fun stuff (chatting, images), free (limited use).
    - **GPT-4o-mini**: Small model, not as good, free.
    - **GPT-4.1**: Improved model, accessible only via API.
    - **GPT-4.5**: Largest model, broader knowledge, higher EQ, paid only.
  - Reasoning models:
    - **o3**: Best model available, PhD+ level in all disciplines, problem solving, complex coding; paid only. **o3-pro** available soon, pro plan only.
    - **o4-mini**: Newer but smaller model, free (limited use). **Do not confuse with 4o-mini!**
    - **o4-mini-high**: Thinks longer, paid only.



# The LLM zoo 3/6

- OpenAI capabilities (vary by model, not all free):
  - Text: Chat, read, analyze, edit, generate, summarize, translate...
  - Code: Read, edit, debug, generate, refactor, explain, write tests...
  - Data (CSV / JSON / Excel / etc.): Parse, analyze, extract, visualize...
  - Science/math: Explain, teach, reason, solve problems at PhD+ level...
  - Image: See, edit, generate; photos, art, slides, diagrams...
  - Voice: Listen, generate, “advanced voice”...
  - Video: Watch, camera/screen sharing...
  - Tool use: Autonomously search the web, run Python code...
  - Custom instructions, memory...
  - Context windows: 128K for GPT-4o, GPT-4.5, o3, o4-mini, 1M for GPT-4.1.

# The LLM zoo 4/6

- OpenAI other features:
  - **Deep Research**: Spends up to 30 mins collecting data from dozens of resources. Specialized early o3. Free (limited use of light version).
  - Create **custom GPTs** that follow specific instructions. Paid only, can be used by free.
  - Create recurrent **tasks** that run autonomously. Paid only.
  - **Sora**: Generate videos based on script. Paid only.
  - **Operator**: Browse the web and perform tasks. Pro plan only.
  - Plus plan: 20 USD/month, pro plan: 200 USD/month.

# The LLM zoo 5/6

- Google models:
  - **Gemini 2.5 Pro**: Latest model, reasoning. Free (limited use).
  - **Gemini 2.5 Flash**: Smaller model, reasoning. Free (limited use).
  - Capabilities: Text, code, data, science/math, image, voice, video, code execution, web search, custom “gems”, Deep Research, Google app integration, 1M context window.
  - Gemini Advanced plan: 27 CAD/month.

# The LLM zoo 6/6

- Anthropic models:
  - **Claude 3.7 Sonnet**: Latest model, non-reasoning. Free (limited use).
  - **Claude 3.7 Sonnet Thinking**: Reasoning variant. Paid only.
  - Capabilities: Text, code, data, science/math, image input only. 200K context window.
  - No voice, video, code execution, custom bots. Web search in US only.
  - Pro plan: 17 USD/month, max plan: 100-200 USD/month.

# Benchmarks 1/2

- **GPQA Diamond:** Graduate-Level Google-Proof Q&A.
- 198 PhD-level MCQ in biology, physics, chemistry.
- Humans + web access: 22%.
- Gemini 2.5 Pro: 84%.
- OpenAI o3: 83%.
- OpenAI o4-mini-high: 78%.
- Claude 3.7 Sonnet Thinking: 77%.
- OpenAI GPT-4.5: 71%.
- Gemini 2.5 Flash: 70%.

# Benchmarks 2/2

- **Humanity's Last Exam:** 2,500 PhD-level MC or short answer questions in math, physics, biology, medicine, humanities, social science, computer science, engineering, chemistry, and more.
- Leading expert human (estimate): 6-8% (mostly MC guesses).
- OpenAI o3: 20%.
- OpenAI o4-mini-high: 18%.
- Gemini 2.5 Pro: 17%.
- Gemini 2.5 Flash: 12%.
- Claude 3.7 Sonnet Thinking: 10%.

# Common misconceptions 1/7

- Misconception: “LLMs only predict the next token, so they can’t do \_\_\_\_\_”.
- Rebuttals:
  - How else would you write text?
  - Humans also predict the next token.
  - By predicting the next token in training, neural net encodes syntax, semantics, world knowledge, etc.
  - Analogy: predict next move in chess.

# Common misconceptions 2/7

- Misconception: “LLMs just memorize the entire Internet; no better than Googling”.
- Rebuttals:
  - High scores in Google-proof benchmarks.
  - Neural net doesn’t store Internet pages verbatim; uses them to internalize general meaning.



# Common misconceptions 3/7

- Misconception: “LLMs can’t even multiply two numbers”.
- Rebuttals:
  - Was true of ChatGPT in 2023. Now LLMs just use Python to multiply.
  - Most humans would use a calculator.
  - Training doesn’t focus on this skill; could theoretically be improved, but there’s no need.

# Common misconceptions 4/7

- Misconception: “I tried an LLM once and it couldn’t do \_\_\_\_\_”.
- Rebuttal:
  - Was true of ChatGPT in 2023. Now LLMs are much more capable.
- Misconception: “I tried an LLM today and it couldn’t do \_\_\_\_\_”.
- Rebuttal:
  - Choose the right model. GPT-4o mini can’t do much, o3 can do a lot.

# Common misconceptions 5/7

- Misconception: “LLMs can’t learn anything new”.
- Rebuttals:
  - Latest models learn on the spot: read documents, web search.
  - Examples:
    - First upload docs/manual/tutorial, then ask question.
    - Enable search to get up-to-date info (o3/GPT-4.5 enable by default).

# Common misconceptions 6/7

- Misconception: “Students cannot use LLMs to cheat on this assessment because it \_\_\_\_\_”.
- Rebuttals:
  - Requires private course material? Upload notes/slides.
  - Requires vision? No problem.
  - Requires clicking with the mouse? Operator / Claude computer use.
  - Requires higher-order reasoning? PhD-level reasoning models.
  - Randomizes numbers? No problem.
  - Is Google-proof? See earlier slide.
  - Is proctored remotely? Interview Coder app, or just use phone.
  - Is timed? LLMs are faster than humans.
  - Is run through AI detector? Doesn't work.
  - Is proctored in person? Earpiece + smart glasses (soon).

# Common misconceptions 7/7

- Misconception: “Student LLM use hurts learning outcomes”.
- Rebuttals:
  - Very true if student just cuts and pastes!
  - Numerous studies show LLM tutors improve student performance (in exams with no LLM available).
  - My own experience:
    - Students said my AI chatbot helped a lot.
    - ASTR 1P02 students who used the chatbot got 6 points higher grades on average.
      - (Note: Does not imply causation.)

# Create your own chatbot 1/2

- Easiest way: custom GPTs.
  - Gemini gems cannot be shared.
- Create custom instructions: course material, pedagogical preferences, logistics...
- Limitations:
  - Must have paid plan.
  - Free users must use GPT-4o mini, an inferior model.
  - Even paid users must use GPT-4o, not a reasoning model.
  - Privacy concerns.

# Create your own chatbot 2/2

- Harder way: use an API.
- Much more customizable, including choice of model.
- Students can use for free.
- Limitations:
  - Must have good programming skills (HTML/CSS/JavaScript/Python).
  - Must host own website.
  - Must pay whenever it's used (per token).
    - Workaround: Gemini 2.5 Flash (500 req/day free)!

Easy way:  
Interface: ChatGPT  
Model: GPT-4o/4o-mini



## Albert Einstein

By Barak Shoshany 

A tutor for the ASTR 1P01/02 course at Brock University

When is the next exam?

What is the material for the next exam?

Can you give me practice questions for the next exam?

Can you make me a practice exam for the next exam?

Ask anything





## What's the material for exam 2?

For **ASTR 1P02 Exam 2**, the material includes:

- **Unit 14: *Star formation and evolution*** — this unit covers interstellar matter, how stars are born, and their life cycles.
- **Unit 15 (up to slide 67): *Dead stars and black holes*** — you'll be tested on stellar remnants like white dwarfs and neutron stars, and supernovae, but **not** on general relativity and black holes (which start at slide 68).

You will **not** be tested again on anything from ASTR 1P01 or Exam 1 of ASTR 1P02, though you're expected to know that material as background knowledge.

You can review:

- [Unit 14 Slides | Practice Questions](#)
- [Unit 15 Slides \(just up to slide 67\) | Practice Questions](#)
- [Practice Exam for Exam 2](#)

Would you like a few practice questions from those units?

Yes, please!

Great! Here are 3 multiple-choice practice questions based on Units 14 and 15 (up to slide 67). Let me know your answers, and I'll give you feedback and hints if needed — no answers will be revealed until you try! 🌟

**1. What is a protostar?**

- A. A star that has begun nuclear fusion in its core
- B. A dense region of gas and dust that hasn't collapsed yet
- C. A newborn star still forming from collapsing gas and dust
- D. A white dwarf surrounded by a planetary nebula

**2. What determines whether a star ends its life as a white dwarf or a neutron star?**

- A. The star's surface temperature
- B. The star's distance from Earth
- C. The mass of the star before it died
- D. The number of planets orbiting the star

**3. Which of the following is true about supernovae?**

- A. Only low-mass stars can undergo supernova explosions
- B. Supernovae occur when planets crash into stars
- C. Supernovae can result in the formation of neutron stars
- D. Supernovae are gentle and rarely affect surrounding space

Take your time and let me know your answers! 😊

1: C

2: C

3: A

Awesome, let's go through your answers! 🌟

### 1. What is a protostar?

✅ Your answer: C — Correct!

A protostar is a forming star still collapsing under gravity from a cloud of gas and dust, before nuclear fusion starts in the core.

### 2. What determines whether a star ends its life as a white dwarf or a neutron star?

✅ Your answer: C — Correct!

The key factor is the **mass** of the star. Low- to intermediate-mass stars become white dwarfs, while more massive stars can end as neutron stars or black holes.

### 3. Which of the following is true about supernovae?

❌ Your answer: A — Incorrect.

Actually, **only high-mass stars** can undergo core-collapse supernovae. There's also a second kind called **Type Ia supernovae**, which involve white dwarfs in binary systems.

📧 Want to try another question to reinforce this?

Hard way:  
Interface: My own website  
Model: Gemini 2.5 Flash (API)

Hi John! ✨ I'm Albert AIInstein, and I was created by Prof. Shoshany to help you get the most out of ASTR 1P01/02. 🌟 You can ask me any questions related to the course logistics or material. I can also teach you any topic, help you explore it in more depth, or create practice questions for you. Remember that help is also available directly from Prof. Shoshany on Teams; I'm just here as an additional resource for your benefit. Also, please keep in mind that as an AI, I might make mistakes! 🤖 So, what can I do for you? 🙋

What's the material for exam 2?

Hello John! 🙋

For ASTR 1P02 Exam 2, the material covered is:

- Unit 14 in its entirety (Star formation and evolution)
- Unit 15 up to slide 67 (This includes everything in Unit 15 except for the sections on general relativity and black holes).

While the exam focuses on these units, it's assumed that you know the material from previous units (ASTR 1P01 Exams 1-3 and ASTR 1P02 Exam 1) as background knowledge.

You can find the lecture slides and practice questions for these units here:

- Unit 14: [Slides](#) | [Black on White Slides](#) | [Practice questions](#)
- Unit 15: [Slides](#) | [Black on White Slides](#) | [Practice questions](#)

There is also a practice exam available for Exam 2 here: [Practice exam 5](#)

Let me know if you have any other questions! 😊

Can you create 3 practice questions for me?

Absolutely, John! I can certainly create some practice questions for you based on the material for Exam 2 (Units 14 and 15 up to slide 67).

Here are 3 practice questions:

1. What is the primary force responsible for the initial collapse of a cloud of gas and dust, leading to the formation of a protostar?
  - A. Electromagnetic force
  - B. Nuclear force
  - C. Gravity
  - D. Centrifugal force
2. After a star like our Sun finishes its main sequence phase, what type of star does it evolve into next?
  - A. A white dwarf
  - B. A neutron star
  - C. A red giant
  - D. A black hole
3. Which of the following is a rapidly spinning neutron star that emits beams of radiation?
  - A. White dwarf
  - B. Magnetar
  - C. Protostar
  - D. Pulsar

Take your time to think about them, and let me know your answers when you're ready! Good luck! 😊

1: C

2: C

3: A

Great job, John! You got the first two questions correct! 🎉

1. Yes, **C. Gravity** is indeed the primary force that causes a cloud of gas and dust to collapse and form a protostar.
2. That's right! **C. A red giant** is the next stage for a star like our Sun after it leaves the main sequence.

For the third question, you answered A. A white dwarf is a stellar remnant, but it's not the one described as a rapidly spinning object emitting beams of radiation.

Think back to the different types of dead stars we discussed in Unit 15. Which one is characterized by its rapid rotation and the emission of radiation beams?

Give it another shot! You're doing great! 🙌



# Part III: Future

To be discussed in  
another workshop...

Any questions?

