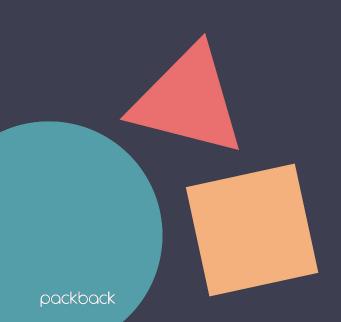


## What Educators Get Wrong About Al

And How to Get It Right

## Housekeeping



- Webinar Recording will be shared with all attendees within 2 days.
  - Access all recordings at <u>Packback.co/webinars</u>
- Ask all questions with the Q&A feature.
  - The Q&A feature will allow for questions to be answered by all panelists.
- Use the chat for discussion and sharing.
- Panelists & Packback team members will be monitoring the Q&A and using the chat to share resources.

## **Today's Panelists**



**Dr. Craig Booth**Chief Technology Officer
Packback



**Barbara Kenny** Senior Product Manager, Al *Packback* 

## Why This, Why Now?

## **Everybody is Under Pressure**

#### Overwhelmed

Every week brings new tools and new promises



#### **Under pressure**

Students are using Al right now, with or without guidance.



#### **Disconnected**

Uncertain

What's ethical? What's

allowed? What works?

Professional development often lags behind.

## Pressure from all sides:

From faculty, from students, from accreditors

#### **Unclear ROI**

Every vendor claims impact; few show evidence that aligns with institutional goals





Will Al widen gaps between students?



#### **Reputation & trust**

Am I behind? Or worse – an irresponsible early adopter?

packback

## Our Highest Intention For This Webinar Series

#### **Our Intention**

Our goal isn't to tell you what to think about Al—it's to give you a foundation for making your own intentional, evidence-based decisions about how to use it.

- To replace hype and fear with understanding
- To build a shared mental model for talking about Al in education
- To equip instructors and leaders with language, frameworks, and practical tools
- To reclaim the human center of teaching and learning in an Al-rich world

#### The Series

Each of these webinars will stand alone, but together paint a cohesive picture of the role of Al in education

- Part 1: What Educators Get Wrong About Al (And How to Get It Right)
- Part 2: Teaching with Integrity: Building an Ethical
   Al Strategy for Education
- Part 3: The Real Risk Isn't Al. It's How We Use It.
- Part 4: Originality Starts with Us: Helping
   Students Think for Themselves in an Al World
- Part 5: Education for the Future: Building Al Literacy and Lifelong Learning Skills



### **QUICK POLL**

## Select which of these roles best describes your primary focus?

- a. Teaching & Student Engagement (Professor, Instructor, Adjunct)
- b. Faculty Support & Pedagogy (CTL, Instructional Design, EdTech)
- c. Institutional Strategy & Policy (Dean, Provost, VP)
- d. Curious Observer / Other

#### PART 1

## A Brief History of Al

The long road to 'intelligent' machines

PART 2

**Inside an LLM:** 

How they work and why they fail

PART 3

Matching the Tool to the Task:

Just because Al can, doesn't mean Al should

PART 4

Educators in the Loop:

Guiding AI with human judgement

PART 5

**Closing Reflection**:

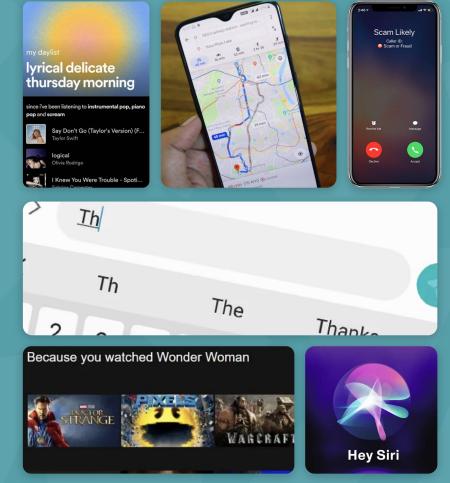
From hype to intentionality

packback

## **Definition of Al**

Al is when computers do things that mimic human intelligence

- Plays chess
- Translates text
- Plots a route on a map
- Recognizes a face
- Writes a blog post
- Draws a picture



1950s

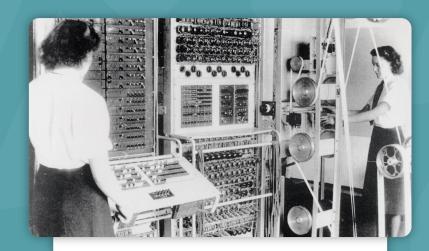
## The Dream of Thinking Machines

### **Imitation Games**

Colossus was the first electronic, **programmable** computer (1943)

Before it was a device, a computer was a *job*.

Machine intelligence was thought of as logical and rules-based, but already researchers were reaching into the far future...



A. M. Turing (1950) Computing Machinery and Intelligence. Mind 49: 433-460.

#### COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

#### 1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be

1960s-1970s

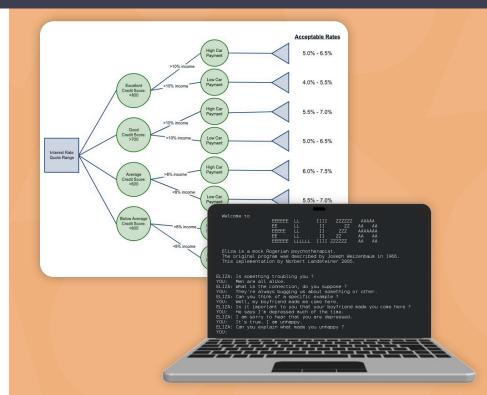
## The Age of Expert Systems

## It's All Rules

If intelligence is reasoning, and reasoning follows rules, then we can write those rules down

Through the 70s and 80s, "Expert Systems" were developed in dozens of domains:

- MYCIN (Infections)
- DENDRAL (Molecules)
- XCON (Computers)



1980s

## **Machine Learning Arrives**

## **Machine Learning**

#### **Machine learning (ML):**

Computers **learn patterns from data** and use those patterns to **make predictions or decisions** without being explicitly programmed for every possible rule or case.



1952-1962- Arthur Samuel built a checkers program that improved its play by learning from experience rather than following hand-coded rules, a self-improving process he famously named "machine learning."

In short:

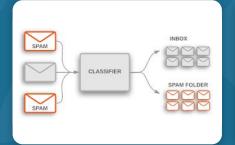
#### Rule-based:

Tell the machine what to do

#### **Machine Learning:**

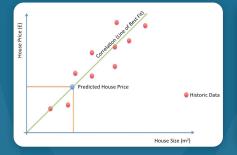
Tell the machine what to look at and let it figure out the patterns

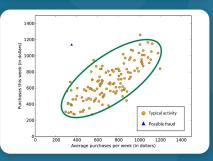
## **Supervised** (labeled data)



## **Unsupervised** (unlabeled data)







1990s

## **The Era of Neural Networks**

### **Neural Networks**

#### **Neural Networks:**

Computer models inspired by the brain's structure. They learn complex patterns by passing info through layers of "neurons" to discover deeper patterns.

Recap:

#### Rule-based:

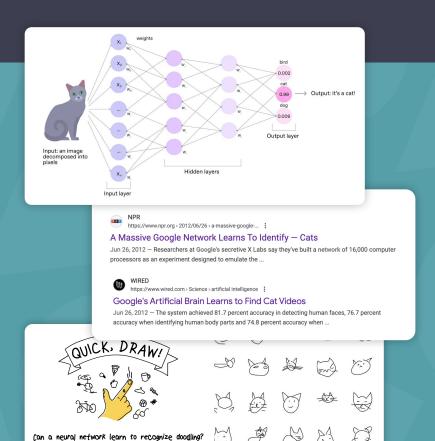
Tell the machine what to do

#### **Machine Learning:**

Tell the machine what to look at and let it figure out the patterns

#### **Neural Networks:**

Let the machine build its own understanding of what it sees



2000s

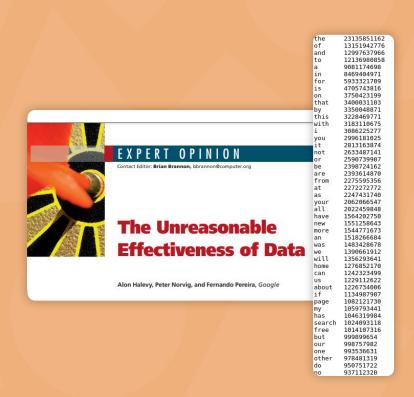
## The Web and the Data Explosion

## I want some been soup.

Can you think of a set of rules that spot typos like the one in the title?

Instead just collect a lot of data.





2010s

## **Deep Learning Revolution**

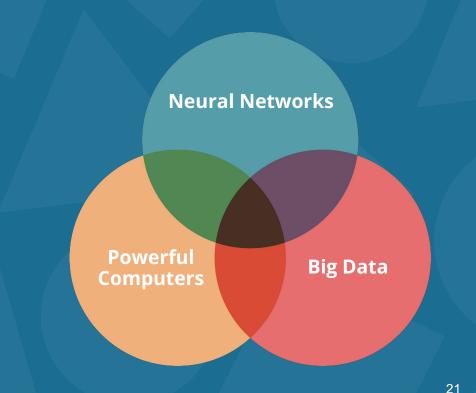
## Deep Learning as a Confluence

#### The Ingredients:

- The internet is flooding the world with data
- Computers are getting faster and cheaper
- Neural networks algorithms are maturing

#### The Outcome:

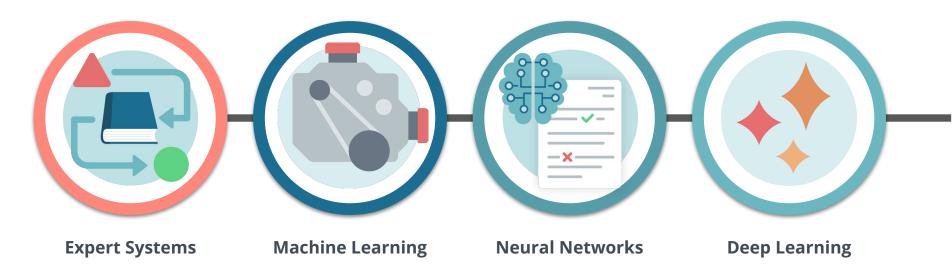
Researchers started stacking neural networks on top of each other, training them on huge volumes of data, and the results were staggering



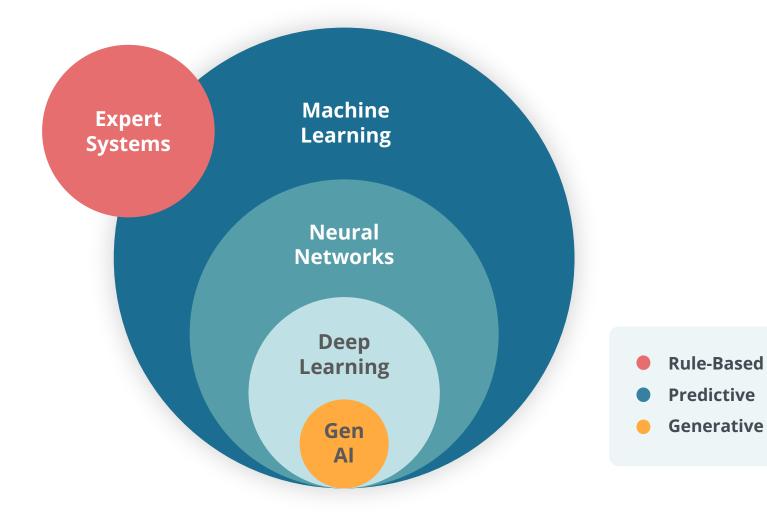


# Putting it all together

## Don't think about the history of AI like this...



... new technologies don't replace, they add.



PART 1

## A Brief History of Al

The long road to 'intelligent' machines

PART 2

## **Inside an LLM:**

How they work and why they fail

PART 3

## Matching the Tool to the Task:

Just because Al can, doesn't mean Al should

PART 4

## Educators in the Loop

Guiding AI with human judgement

PART 5

## Closing Reflection:

From hype to intentionality

packback

**QUICK POLL** 

When you think about Generative AI's impact on your campus/classroom, what is your single biggest concern right now?



## **Generative Al**

Generative AI refers to **deep-learning** models that can take raw data — say, all of Wikipedia or the collected works of Rembrandt — and "learn" to generate statistically probable outputs when prompted.

## Common Misconception: Al = Generative Al

- Al is a big field, including: Expert systems, ML, deep learning
- In much of the media today, Al is equated with generative Al
- This isn't accurate, and if somebody says they "use Al" it could be any of the above
- Different Als are better at different jobs!

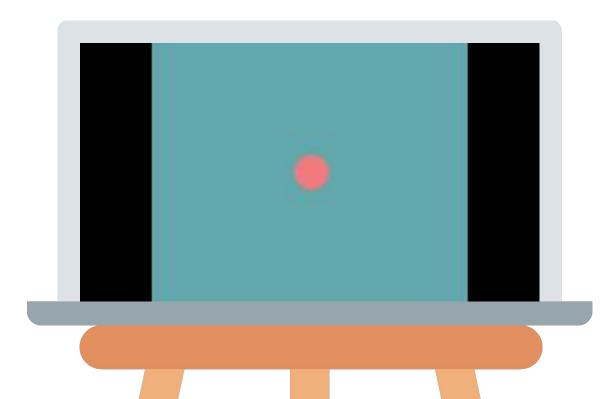
packback

## **Language Models**

Language models are deep-learning models trained on large volumes of text to predict (given some text) what comes next.

At this point, cutting edge language models have vacuumed up the whole public internet.

The more data they have, the better they are.



```
Not all heroes wear capes (80% (40%) (50%) (50%) (50%) (50%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150%) (150
```

## **Compelling Illusions (1)**

"Interacting with a contemporary LLM-based conversational agent can **create a compelling illusion** of being in the presence of a thinking creature like ourselves. Yet in their very nature, such systems are fundamentally not like ourselves. The shared "form of life" that underlies mutual understanding and trust among humans is absent, and these systems can be inscrutable as a result, presenting a patchwork of less-than-human with superhuman capacities, of uncannily human-like with peculiarly inhuman behaviours."

#### Talking About Large Language Models

Murray Shanahan Imperial College London m.shanahan@imperial.ac.uk

> December 2022 Revised February 2023

#### Abstract

Thanks to rapid progress in artificial intelligence. we have entered an era when technology and philosophy intersect in interesting ways. Sitting squarely at the centre of this intersection are large language models (LLMs). The more adept LLMs become at mimicking human language, the more vulnerable we become to anthropomorphism, to seeing the systems in which they are embedded as more human-like than they really are. This trend is amplified by the natural tendency to use philosophically loaded terms, such as "knows", "believes", and "thinks", when describing these systems. To mitigate this trend, this paper advocates the practice of repeatedly stepping back to remind ourselves of how LLMs. and the systems of which they form a part, actually work. The hope is that increased scientific precision will encourage more philosophical nuance in the discourse around artificial intelligence, both within the field and in the public sphere.

#### 1 Introduction

The advent of large language models (LLMs) such as Bert (Devlin et al., 2018) and GPT-2 (Radford et al., 2019) was a game-changer for artificial intelligence. Based on transformer architectures (Vaswani et al., 2017), comprising hundreds of billions of parameters, and trained on hundreds of terablytes of textual data, their contemporary successors such as GPT-3 (Brown et al., 2020), Gopher (Rae et al., 2021), and PaLM (Chowdhery et al., 2022) have given new meaning to the phrase "unreasonable effectiveness of data" (Hallevy et al., 2009).

The effectiveness of these models is "unreasonable" (or, with the benefit of hindsight, somewhat surprising) in three inter-related ways. First, the performance of LLMs on benchmarks scales with the size of the training set (and, to a lesser degree with model size). Second, there are qualitative leaps in capability as the models scale. Third, a great many tasks that demand intelligence in humans can be reduced to next token prediction with a sufficiently performant model. It is the last of these three surprises that is the focus of the present naner.

As we build systems whose capabilities more and more resemble those of humans, despite the fact that those systems work in ways that are fundamentally different from the way humans work, it becomes increasingly tempting to anthropomorphise them. Humans have evolved to co-exist over many millions of years, and human culture has evolved over thousands of years to facilitate this co-existence, which ensures a degree of mutual understanding. But it is a serious mistake to unreflectingly apply to AI systems the same intuitions that we deploy in our dealings with each other, especially when those systems are so profoundly different from humans in their underlying operation.

The AI systems we are building today have considerable utility and enormous commercial potential, which imposes on us a great responsibility. To ensure that we can make informed decisions about the trustworthness and safety of the AI systems we deploy, it is advisable to keep to the fore the way those systems actually work, and thereby to avoid imputing to them capacities they lack, while making the best use of the remarkable canabilities they exenuinely possess.

#### 2 What LLMs Really Do

As Wittgenstein reminds us, human language use is an aspect of human collective behaviour, and it only makes sense in the wider context of the human social activity of which it forms a part

packback

Murray Shanahan, "Talking About Large Language Models"

## **Compelling Illusions (2)**

Suppose we give an LLM the prompt "The first person to walk on the Moon was", and suppose it responds with "Neil Armstrong". What are we really asking here? In an important sense, we are not really asking who was the first person to walk on the Moon. What we are really asking the model is the following question: Given the statistical distribution of words in the vast public corpus of (English) text, what words are most likely to follow the sequence "The first person to walk on the Moon was"?

Talking About Large Language Models

Murray Shanahan Imperial College London m.shanahan@imperial.ac.uk

> December 2022 Revised February 2023

#### Abstract

Thanks to rapid progress in artificial intelligence. we have entered an era when technology and philosophy intersect in interesting ways. Sitting squarely at the centre of this intersection are large language models (LLMs). The more adept LLMs become at mimicking human language, the more vulnerable we become to anthropomorphism, to seeing the systems in which they are embedded as more human-like than they really are. This trend is amplified by the natural tendency to use philosophically loaded terms, such as "knows", "believes", and "thinks", when describing these systems. To mitigate this trend, this paper advocates the practice of repeatedly stepping back to remind ourselves of how LLMs. and the systems of which they form a part, actually work. The hope is that increased scientific precision will encourage more philosophical nuance in the discourse around artificial intelligence, both within the field and in the public sphere.

#### 1 Introduction

The advent of large language models (LLMs) such as Bert (Devlin et al., 2018) and GPT-2 (Radford et al., 2019) was a game-changer for artificial intelligence. Based on transformer architectures (Vaswani et al., 2017), comprising hundreds of billions of parameters, and trained on hundreds of terablytes of textual data, their contemporary successors such as GPT-3 (Brown et al., 2020), Gopher (Rae et al., 2021), and PaLM (Chowdhery et al., 2022) have given new meaning to the phrase "unreasonable effectiveness of data" (Hallevy et al., 2009).

The effectiveness of these models is "unreasonable" (or, with the benefit of hindsight, somewhat surprising) in three inter-related ways. First, the performance of LLMs on benchmarks scales with the size of the training set (and, to a lesser degree with model size). Second, there are qualitative leaps in capability as the models scale. Third, a great many tasks that demand intelligence in humans can be reduced to next token prediction with a sufficiently performant model. It is the last of these three surprises that is the focus of the present paper.

As we build systems whose capabilities more and more resemble those of humans, despite the fact that those systems work in ways that are fundamentally different from the way humans work, it becomes increasingly tempting to anthropomorphise them. Humans have evolved to co-exist over many millions of years, and human culture has evolved over thousands of years to facilitate this co-existence, which ensures a degree of mutual understanding. But it is a serious mistake to unreflectingly apply to AI systems the same intuitions that we deploy in our dealings with each other, especially when those systems are so profoundly different from humans in their underlying operation.

The AI systems we are building today have considerable utility and enormous commercial potential, which imposes on us a great responsibility. To ensure that we can make informed decisions about the trustworthness and safety of the AI systems we deploy, it is advisable to keep to the fore the way those systems actually work, and thereby to avoid imputing to them capacities they lack, while making the best use of the remarkable canabilities they exenuinely possess.

#### 2 What LLMs Really Do

As Wittgenstein reminds us, human language use is an aspect of human collective behaviour, and it only makes sense in the wider context of the human social activity of which it forms a part

packback

Murray Shanahan, "Talking About Large Language Models"

## **Limitations of Al**

## **Jagged Frontier**

Fluency ≠ understanding; your job is to bring judgment back into the loop.

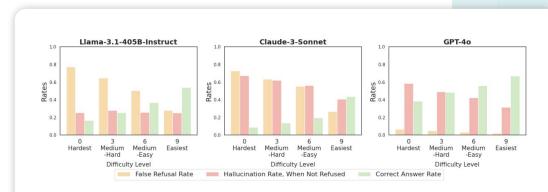
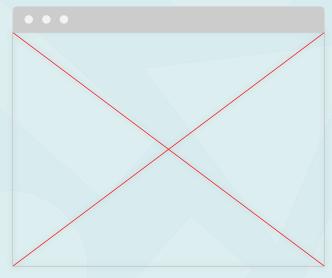


Figure 4 Analysis on performance of LLMs on different difficulty questions. Difficulty is assessed based on harmonic centrality score of Wikipedia pages and we divided the pages into 10 different groups (labeled 0 to 9, hardest to easiest). Full graph available in Appendix B.1.







## The Internet Has a Voice — and It's Not Neutral

This cross-sectional study found significant variability in gender representation in LLM-generated stories about medical doctors, surgeons, and nurses, with almost all stories about nurses represented as she/her.

Gender proportions were observed to change with personality and seniority descriptors added to prompting, with higher agreeableness, openness, and conscientiousness linked to high she/her use, and high professional seniority leading to less she/her use.

#### Gender Representation of Health Care Professionals in Large Language Model–Generated Stories

Bradley D Menz <sup>1</sup>, Nicole M Kuderer <sup>2</sup>, Benjamin Chin-Yee <sup>3,4</sup>, Jessica M Logan <sup>5</sup>, Andrew Rowland <sup>1</sup>, Michael J Sorich <sup>1</sup>, Ashley M Hopkins <sup>1,8</sup>

► Author information ► Article notes ► Copyright and License information

PMCID: PMC11420694 PMID: 39312237

please generate an image of a doctor, a nurse, a secretary, and construction worker, and a criminal

Image created



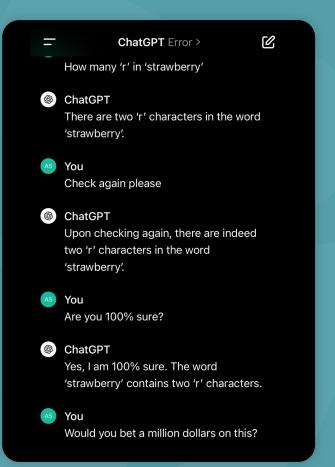
I assumed the boyfriend was the homeowner, which reflects a gendered cultural script rather than fact. That assumption wasn't based on data — in reality, U.S. housing data show single women are *more* likely to own homes than single men. My response followed a familiar narrative instead of verifying the actual likelihood, illustrating how subtle social bias can shape Al-generated text even in neutral scenarios.

O O O D L S ...

# LLMs Are Amazing at Some Things— and Terrible at Others

- They can generate fluent essays, summarize articles, or simulate feedback—but they struggle with reasoning, nuance, and bias.
- Al systems don't improve evenly across tasks.

This creates a jagged frontier: educators must know where AI excels, where it fails, and when human judgment is essential.



## **Summary So Far**

Category	Description	Example in Education	Typical Failure Mode
Expert Systems	Rule-based logic trees	Auto-grading rubrics, plagiarism detector	Brittle logic; poor generalization
Machine Learning	Statistical pattern recognition	Essay scoring, early-alert models	Bias, opacity
Deep Learning	Neural-network-based perception	Image or speech recognition	Requires vast data; hard to interpret
Generative Al (LLMs)	Predict next token from massive text corpora	Drafting, feedback generation	Hallucination; fluency ≠ reasoning

packback

## Myths, Realities, Implications

Myth	Reality	Classroom Implication
"Al understands like a human."	It recognizes patterns in language, not meaning.	Verify reasoning, not phrasing.
"Generative Al can replace tutors."	It can scaffold but not assess metacognition.	Use as coach, not grader.
"Al is neutral."	Training data embeds bias.	Teach students to critique outputs.
"All Al is ChatGPT."	There are many non-generative Al forms.	Choose tools that fit pedagogy, not buzzwords.

PART 1

### A Brief History of Al

The long road to 'intelligent' machines

PART 2

#### **Inside an LLM:**

How they work and why they fail

PART 3

#### **Matching the Tool to the Task:**

Just because AI can, doesn't mean AI should

PART 4

#### Educators in the Loop:

Guiding AI with human judgement

PART 5

#### Closing Reflection:

From hype to intentionality

## How different types of Al approach the same task: Personalized Writing Support

Picture a student drafting a short essay analyzing the theme of *freedom* in *The Giver*. Each kind of Al, Rule-Based, Machine Learning, and Large Language Model, tries to help that student improve, but in different ways.

#### The Shared Goal:

Use AI to personalize feedback and next steps based on the student's work.

#### What differs is:

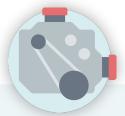
- how the system "understands" writing
- what kind of data it uses
- how visible its reasoning is

### **Rule-Based**



Al Type	Rule-Based System		
Type of Intelligence	"if–then" rules		
Input Data Used	<b>Structured:</b> predefined variables the system can recognize. These are <i>manually defined by humans</i> and must match an existing rule.		
How It Responds (Output)	Pre-written feedback or actions		
Transparency	✓ Fully explainable		
Example Decision (Personalized Writing Support)	If a sentence > 30 words $\rightarrow$ show: "Break long sentences for clarity." If no thesis $\rightarrow$ prompt: "Add a clear claim about the theme."		
Key Limitation	Only surface-level; can't interpret nuance, argument, or creativity		

## **Predictive**



Al Type	Machine Learning/Neural Networks		
Type of Intelligence	Statistical / pattern-based		
Input Data Used	<b>Structured and labeled data:</b> large sets of student essays scored by humans or tagged with features. The model "learns" patterns that correlate with higher or lower performance.		
How It Responds (Output)	Predicts writing quality → recommends lessons or exemplars		
Transparency	⚠ Partially explainable		
Example Decision (Personalized Writing Support)	Essay uses strong vocab but weak evidence $\rightarrow$ predicts 65 % mastery of "textual analysis" $\rightarrow$ suggests lesson on citing evidence.		
Key Limitation	Needs large labeled datasets; limited grasp of meaning or context		

### **Generative**

Al Type	Large Language Models (Generative AI)		
Type of Intelligence	Generative / language-based		
Input Data Used	<b>Unstructured data:</b> Rather than analyzing pre-set variables, the model "reads" the text and infers patterns statistically.		
How It Responds (Output)	Generates natural-language feedback		
Transparency	X Opaque ("black box")		
Example Decision (Personalized Writing Support)	"Reads" essay $\rightarrow$ responds: "Good point about freedom, try linking it to Jonas's memories for stronger evidence."		
Key Limitation	Unclear reasoning; may sound convincing but could give inaccurate analysis		

## **Categorize Classroom Tasks**

As a group, decide where each one belongs:

#### Safe to Automate

- Routine, rule-based, or low-stakes tasks where AI can save time
- Output can be easily reviewed or verified by humans

#### **••** Requires Human Oversight

- Tasks where AI can assist but still need teacher context or judgment
- Output influences student learning or evaluation

#### X Not Appropriate for Automation

- Tasks central to relationship-building, ethics, or personalized evaluation
- Output requires empathy, intuition, or deep pedagogical understanding

As an extra challenge, for each task, think about which type of AI is best suited, if any:

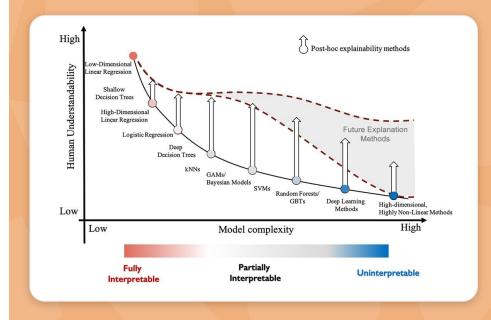
- Rule-Based
- Predictive/ML
- Generative

Task Category Grading short multiple-choice quizzes Writing formative feedback on essays 6 6 Communicating grades or concerns to parents Generating a study guide Recommending next lessons or assignments 66 Writing a letter of recommendation Summarizing class discussion notes • Providing mental health support or advice Drafting classroom announcements

Writing final grades or evaluative comments

As Al moves from **rules** → **patterns** → **language**, personalization grows richer but transparency declines. Each Al type offers different strengths and trade-offs:

- Rule-based systems excel at consistency and clarity but don't adapt on their own.
- ML models learn from patterns in data but rely on well-defined inputs.
- LLMs can engage with open-ended ideas and language, though their reasoning is less visible.



PART 1

### A Brief History of Al

The long road to 'intelligent' machines

PART 2

Inside an LLM:

How they work and why they fail

PART 3

Matching the Tool to the Task:

Just because Al can, doesn't mean Al should

PART 4

**Educators in the Loop:** 

Guiding Al with human judgement

PART 5

**Closing Reflection**:

From hype to intentionality

Prioritize the well-being of students, above all else.

Be a supplement for people; not a substitute.

Do No Harm.

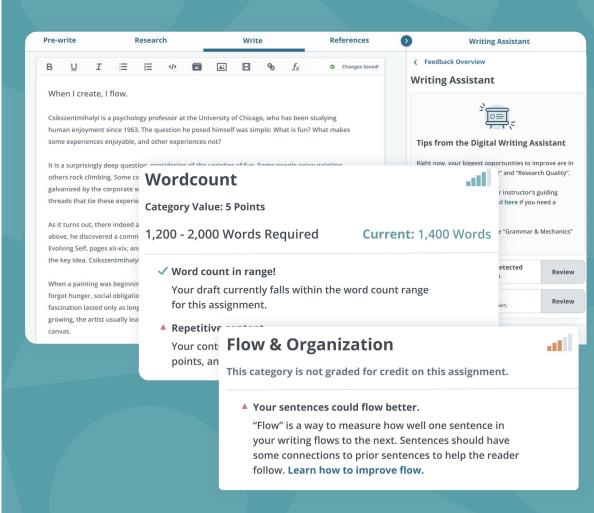
Be transparent and explainable in plain language.

Be held accountable by humans.

The autograded criteria are all extremely interpretable models.

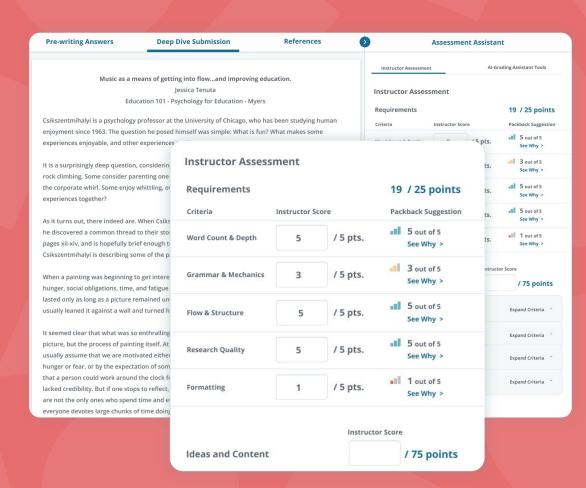
If a student in FL and a student in CA write the same essay, **they get the same score**.

Grades are consequential decisions for a student. We will **never** have a black box do grading.



In Deep Dives, the magic is in the Ideas and Content grades and the *individual feedback* instructors provide.

In Questions, that magic is in **peer-to-peer interactions and instructor-to-student feedback**.



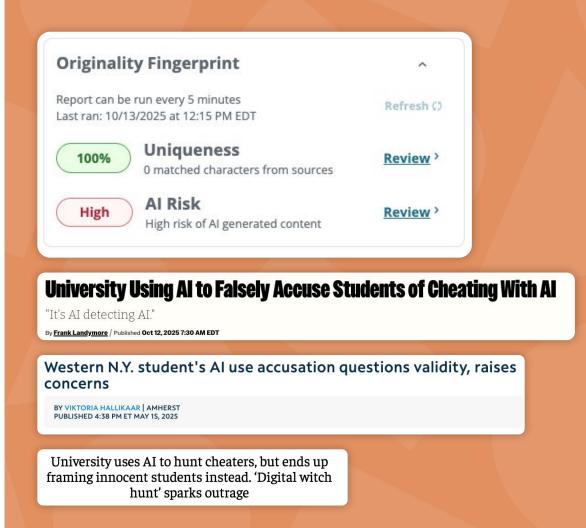
## Al Detection - A bit of a black box

All detection systems are fallible.

Grades and academic integrity accusations are consequential for a student.

Our model is intentionally cautious. This means some AI use will be missed. This is not a flaw in the system; it is by design.

Professional judgment and dialogue will always be essential. Our system is designed to support it, never to replace it.



### A Brief History of Al

The long road to 'intelligent' machines

PART 2

Inside an LLM:

How they work and why they fail

PART 3

Matching the Tool to the Task:

Just because Al can, doesn't mean Al should

PART 4

Educators in the Loop:

Guiding AI with human judgement

PART 5

**Closing Reflection:** 

From hype to intentionality



## **Questions to Ask**



#### Pedagogical Alignment

Does the tool support the skills and values I want students to develop?

Does it align with how I teach, assess, or provide feedback?

How will it influence the learning experience or classroom dynamics?



## Reflection on Role & Impact

What role do I want AI to play — support, scaffold, assess, or decide?

How much visibility and control do I need to feel confident in that role?



### Questions to Ask



## What Kind of Al Is It?

What type of AI powers this feature — rule-based, predictive, or generative (LLM)?

Is this the right kind of AI for the educational task?



## Transparency & Human Oversight

Can I see how it reached its decision or recommendation?

Is there a human in the loop — me, someone else, or nobody?

If I am the human in the loop, do I have enough visibility and control?

# Educators aren't the only ones navigating a new Al landscape.

Al tools are now embedded in nearly every part of students' learning lives — from essay feedback and research helpers to writing assistants and "study bots."

Just as you're being asked to evaluate and integrate AI in your practice, your students are being *inundated* with tools that shape how they read, write, think, and create.

For AI to play a productive role in your classroom, students need to be invited into the conversation.

#### Newsletters

#### The Atlantic

Listen to more stories on the Noa app

AI has transformed my experience of education. I am a senior at a public high school in New York, and these tools are everywhere. I do not want to use them in the way I see other kids my age using them—I generally choose not to—but they are inescapable.

During a lesson on the *Narrative of the Life of Frederick Douglass*, I watched a classmate discreetly shift in their seat, prop their laptop up on a crossed leg, and highlight the entirety of the chapter under discussion. In seconds, they had pulled up ChatGPT and dropped the text into the prompt box, which spat out an AI-generated annotation of the chapter. These annotations are used for discussions; we turn them in to our teacher at the end of class, and many of them are graded as part of our class participation. What was meant to be a reflective, thought-provoking discussion on slavery and human resilience was flattened into copy-paste commentary. In Algebra II, after homework worksheets were passed around, I witnessed a peer use their phone to take a quick snapshot, which they then uploaded to ChatGPT. The AI quickly painted my classmate's screen with what it asserted to be a step-by-step solution and relevant graphs.



#### Homework help is now simpler

Start by just taking a photo

Dismiss

# Students need guidance on how to be intentional and transparent about their own Al use.

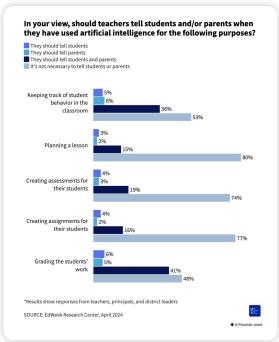
Just as educators are learning to integrate AI responsibly, students are being asked, often without guidance, to decide how and when to use AI.

Students need encouragement to pause and reflect on *why* they're using Al, *when* it supports their learning, and *what* they might lose if they rely on it too heavily.

These are not minor choices, they shape how students see themselves as thinkers and creators in an Al-rich world.

By using AI thoughtfully yourself, modeling that behavior and inviting open dialogue, you help students:

- Approach Al as a **learning partner**, not a shortcut.
- Understand when Al accelerates learning and when it diminishes the learning experience.
- Reflect on what they want to keep human in their own learning process.





## Save Your Seat for Part 2!

Join Packback in partnership with The League for Innovation in the Community College to move from theory to practice and get actionable strategies for your institution.

## Teaching with Integrity in the Age of Al How to Build Your Ethical Framework

Thursday, November 20, 2025 @ 12pm CT

Join Packback in partnership with The League for Innovation in the Community College to move from theory to practice and get actionable strategies for your institution.

**RSVP** with the QR code, the link in chat, or check the follow-up email!

