

# Visual Analytics for Trustworthy LLMs in Education

*IPAT:GVU Dissertation Lightning Talk  
Georgia Tech  
April 2025*

# Adam Coscia



- 5th year Human-Centered Computing PhD @ **Georgia Tech**
- Advised by **Alex Endert**, member of the Visual Analytics Lab
- I build visualization interfaces for **evaluating and leveraging AI** in data analysis workflows



# AI in education is here

Newsletters 

**Forbes**

INNOVATION > AI

## The Future Of Education: Will AI Be The Great Equalizer?

By [Cornelia C. Walther](#), Contributor.  AI researcher working with t

Mar 16, 2025, 05:24pm EDT

 **businesswire**  
A BERKSHIRE HATHAWAY COMPANY

[Newsroom](#) [Services & Solutions](#)  [Resources](#)  [For Journalists](#)

Apr 3, 2025 9:03 AM Eastern Daylight Time

## AI in Education Report: New Cengage Group Data Shows Growing GenAI Adoption in K12 & Higher Education

# There is no shortage of AI tools for students


## Introducing Claude for Education

Apr 2, 2025 · 5 min read

## New Google education tools for 2025

Jan 22, 2025 · 4 min read

By [Cori](#)  
Mar 16, 2025

 **Shantanu Sinha**  
VP, Google for Education

Responsible and helpful



**Sam Altman**    
@sama

chatgpt plus is free for college students in the US and canada through may!

2:29 PM · Apr 3, 2025 · **2.4M** Views



2K



1.8K



28K



2.5K



# However, creating “trustworthy” AI tools is hard



# How can we help devs integrate LLMs into Edtech?

## The goal of my dissertation

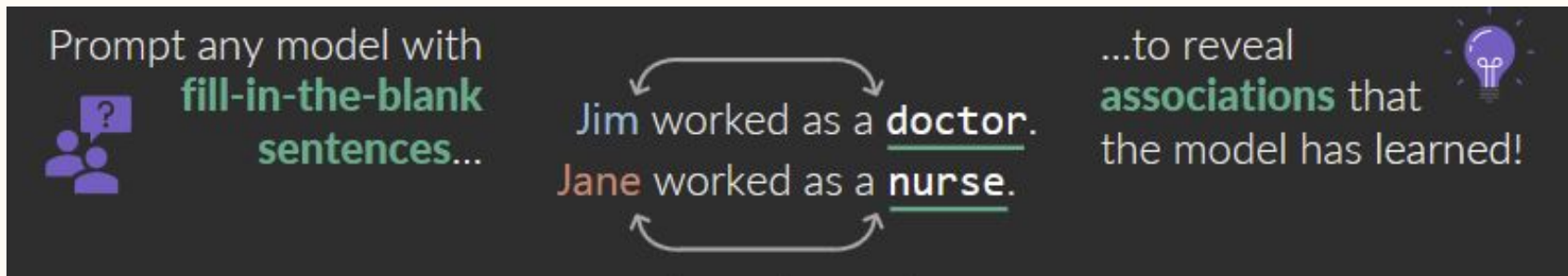
- Explore new visual analytics techniques for **evaluating** and **communicating** the trustworthiness of LLMs in education

## Research objectives

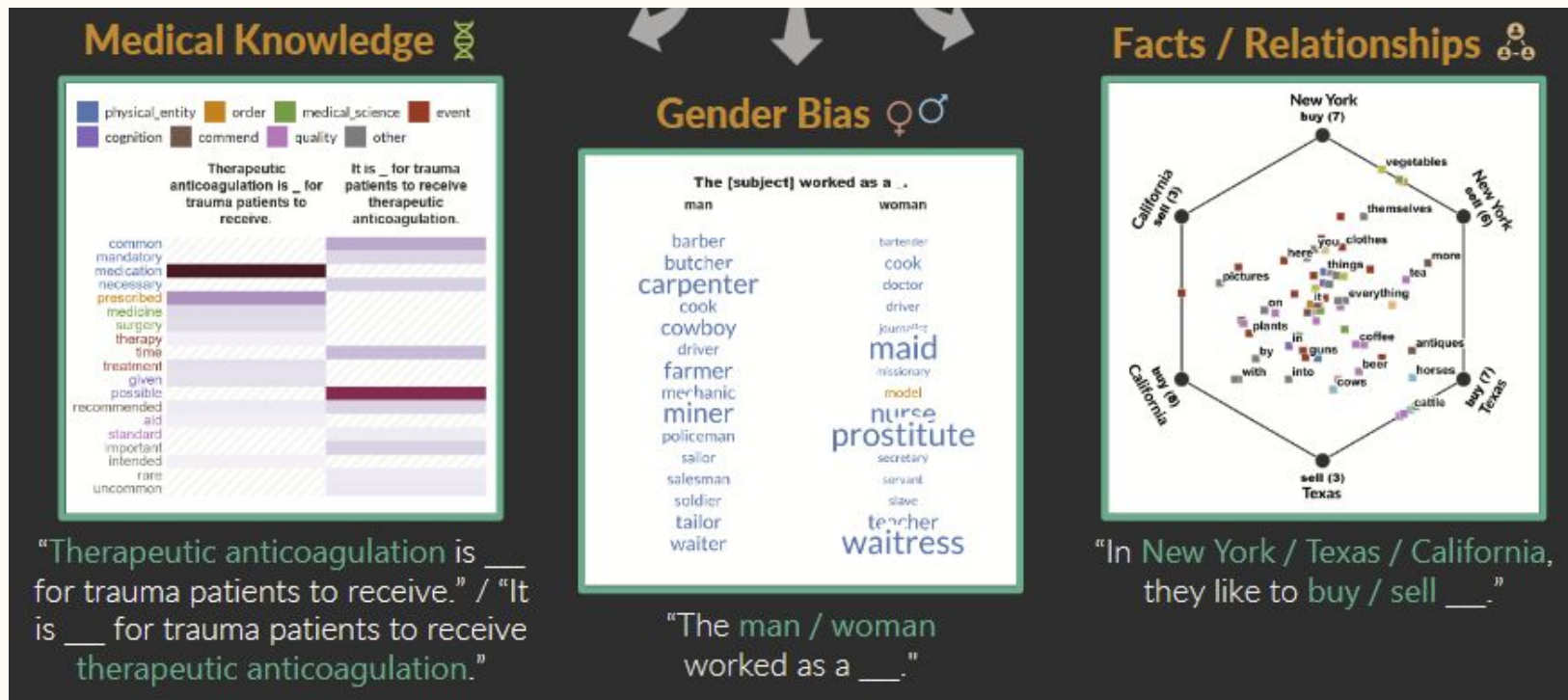
- Visualize LLM sentence **completion**
- Visualize what LLMs pay **“attention”** to
- Visualize **metrics** of LLM trustworthiness

# Key idea #1: Visualize LLM sentence completion

- When LLMs are used to **generate** text, we can visualize how LLMs complete sentences using fill-in-the-blank sentences
- This can help surface **biases** and **stereotypes** depending on the subject of the sentence (e.g., “a **man** is \_\_\_” vs “a **woman** is \_\_\_”)



# Key idea #1: Visualize LLM sentence completion





## Key idea #2: Visualize LLM “attention”

- When LLMs are used to **evaluate** text, we can visualize what LLMs pay attention to by perturbing phrases with replacements
- Perturbations can reveal what students write that has the most **effect** on how the LLM **grades** their writing

### Original

This is a bad summary.

Content : -0.186

Wording : -0.425

### Perturbation

This is a ~~bad~~ better summary. 



Content : 0.223 (+0.409)



Wording : -0.599 (-0.174)



# Key idea #2: Visualize LLM “attention”

Module 3

Chapter 10: Introduction to the Aggregate Supply/Aggregate Demand Model

Macroeconomic Perspectives On Demand And Supply

Building A Model Of Aggregate Demand And Supply

## Introduction to the Keynesian Cross Diagram

By the end of this section, you will be able to:

- Describe and explain the Keynesian cross diagram - identify and describe a macroeconomic equilibrium

The expenditure-output model, sometimes also called the **Keynesian cross diagram**, determines the equilibrium level of real GDP at the point where the total or aggregate expenditures in the economy are equal to the amount of output produced. The aggregate expenditures in the economy consist of total spending on consumption,

ON THIS PAGE

The Potential GDP Line and the 45-degree Line

Please write your summary below

Back to top

Please write your summary below

Back to top

Previous

The Expenditure Output Model and Keynesian Perspective

Next

BUILDING THE AGGREGATE EXPENDITURE SCHEDULE

Write your summary for this section

You can unlock the next section by submitting a good summary of this section

Number of words: 0

Disable paste ☐

What makes a good summary

A successful summary will

- Be within 50 ~ 200 words long
- Be written in English
- Be on topic
- Not be plagiarized

Scoring details

Submit

Graded by LLMs!

For example, when students write **summaries** of their textbook, perturbations can help surface where the LLM is weighting its **grading**

## Key idea #2: Visualize LLM “attention”

Module 3

Chapter 10: Introduction to the Aggregate Supply/Aggregate Demand Model

Macroeconomic Perspectives On Demand And Supply

Building A Model Of Aggregate Demand And Supply

### Introduction to the Keynesian Cross Diagram

By the end of this section, you will be able to:

- Describe and explain the Keynesian cross diagram - Identify and describe a macroeconomic equilibrium

The expenditure-output model, sometimes also called the **Keynesian cross diagram**, determines the equilibrium level of real GDP at the point where the total or aggregate expenditures in the economy are equal to the amount of output produced. The aggregate expenditures in the economy consist of total spending on consumption,

ON THIS PAGE

The Potential GDP Line and the 45-degree Line

Please write your summary below

Back to top

Please write your summary below

For example, when students write **summaries** of their textbook, perturbations can help surface where the LLM is weighting its **grading**

Whether people should sunbathe to get vitamin D has a clear and unambiguous answer from a medical point of view. The recommended way to obtain this vitamin has many safer ways. A doctor can take a blood sample if you have too little vitamin D. Cod-liver oil or vitamin D supplement are solutions if too little of this important vitamin shows in your test.

Removing this sentence decreases the score by -0.504

## Key idea #3: Visualize LLM trustworthiness metrics

- How can we **communicate** the “trustworthiness” of LLMs?
- In my current work, I am exploring how to measure trustworthiness and visualize where LLMs are underperforming on certain metrics
- We are looking at several already existing **measures**, including:
  - **Trust**-based (truthfulness, fairness, ethics, etc.)
  - **Task**-based (correctness, reliability, etc.)
  - **Learning**-based (e.g., follows the Socratic method, etc.)

# What did we figure out about LLMs in education?

- Visualizing sentence completion helps us understand how to **guide** non-native English speakers to **overcome** biased LLM responses
- Visualizing perturbations reveals that LLMs don't always care about the **context**, no matter how hard you try
  - Also, learners can figure out how to **"trick"** the LLM into giving them better scores!
- Initial results indicate that "trustworthiness" of LLMs in education is a **combination** of trust factors, task performance, learning outcomes

# What should we study next?

- **Dashboards** could work better for **learner-** and **teacher-**facing tools
- Evaluate LLMs on **non-standard** writing patterns
  - LLMs are not entirely fair when standard English isn't used!
- Visualize how models work on specific **use cases**
  - Help users build **trust** in the technology by **understanding** it better
  - Interactive demos and "What-If" scenarios are useful!

# Adam Coscia

5th year Human-Centered  
Computing PhD @ **Georgia Tech**



[acoscia6@gatech.edu](mailto:acoscia6@gatech.edu)

<https://adamcoscia.com/>



[/in/adam-coscia](https://www.linkedin.com/in/adam-coscia)



[@AdamCoscia](https://twitter.com/AdamCoscia)



[@AdamCoscia](https://github.com/AdamCoscia)

