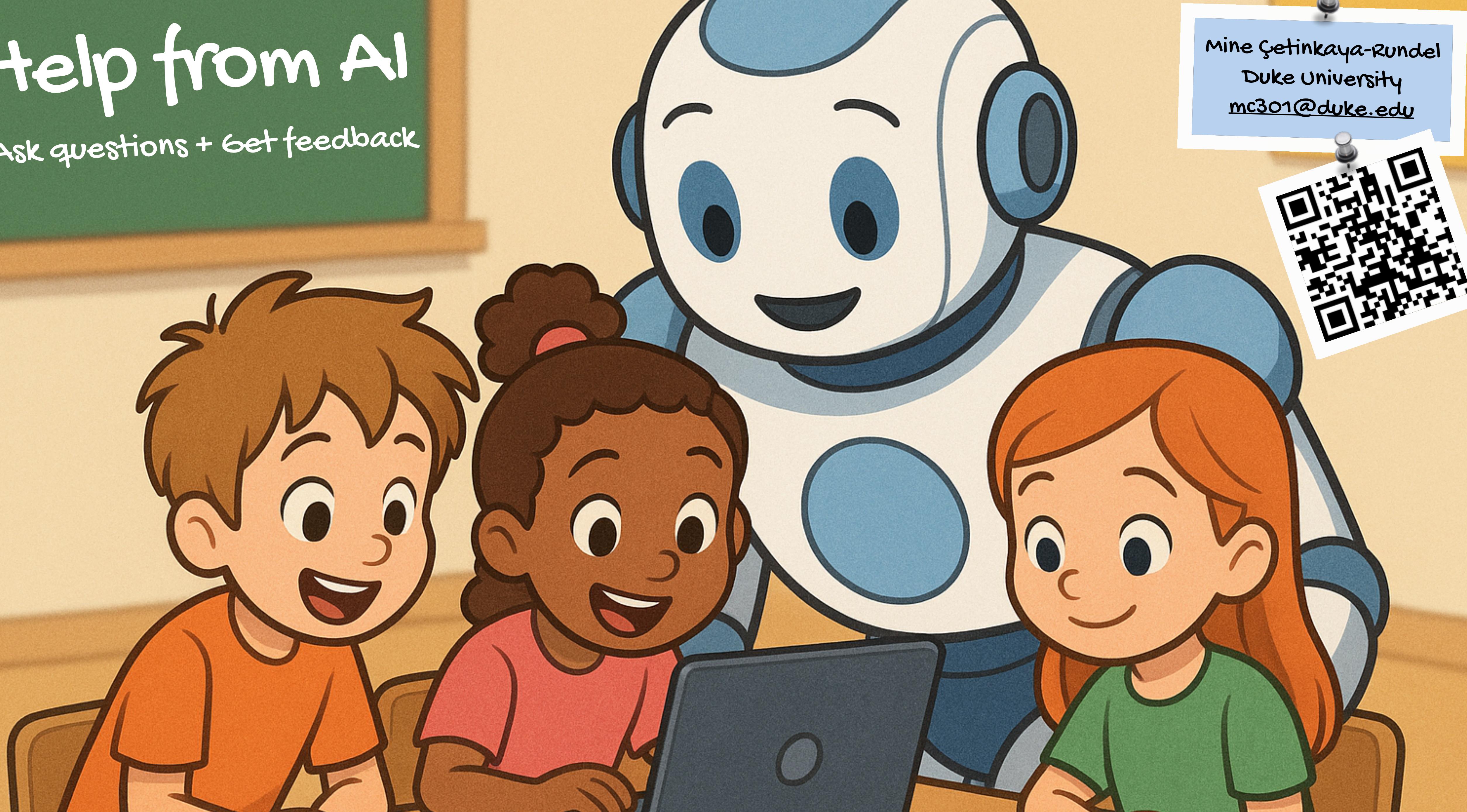


# Help from AI

Ask questions + Get feedback

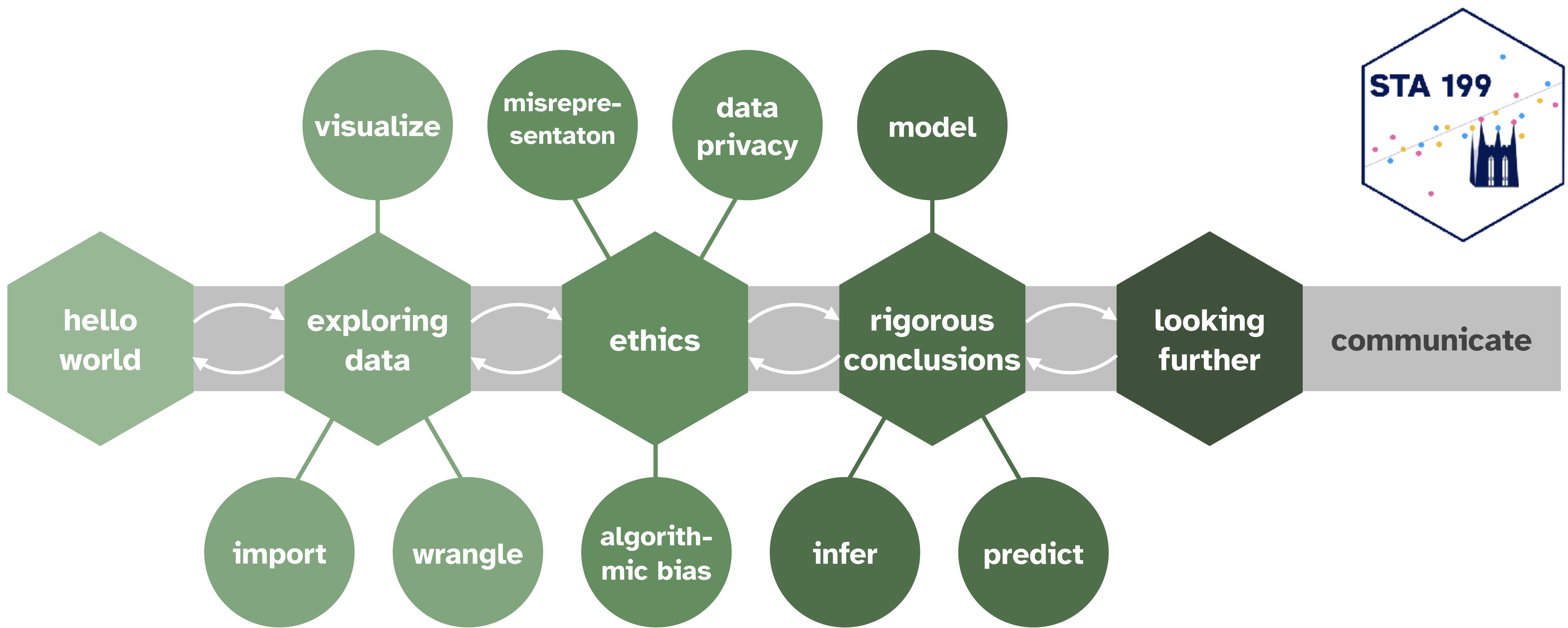
Mine Çetinkaya-Rundel  
Duke University  
[mc301@duke.edu](mailto:mc301@duke.edu)



# Context

## Introduction to Data Science and Statistical Thinking

[sta199-f24.github.io](https://sta199-f24.github.io)



# Components + Assessment

| Component             | Assessment   | Weight |
|-----------------------|--|--------|
| Twice weekly lectures | Application exercises graded for engagement                        | 5%     |
| Once weekly labs      | Lab assignments graded for accuracy                                | 35%    |
| Midterm               | In-class conceptual exam followed by 2-day computational take home | 20%    |
| Final                 | In-class conceptual exam   | 20%    |
| Project               | Team-based, open-ended, culminating in presentation + report       | 20%    |

# AI policy

(that was all too optimistic)

- ✓ AI tools for code: You may use, you must explicitly cite. *[some guidance for how to cite]* The prompt you use cannot be copied and pasted directly from the assignment; you must create a prompt yourself.
- ✗ AI tools for narrative: Unless instructed otherwise, you may not use generative AI to generate a narrative that you then copy-paste verbatim into an assignment or edit and then insert into your assignment.
- ✓ AI tools for learning: You're welcomed to ask AI tools questions that might help your learning and understanding in this course.

# Project 1

A chat



that (hopefully) generates

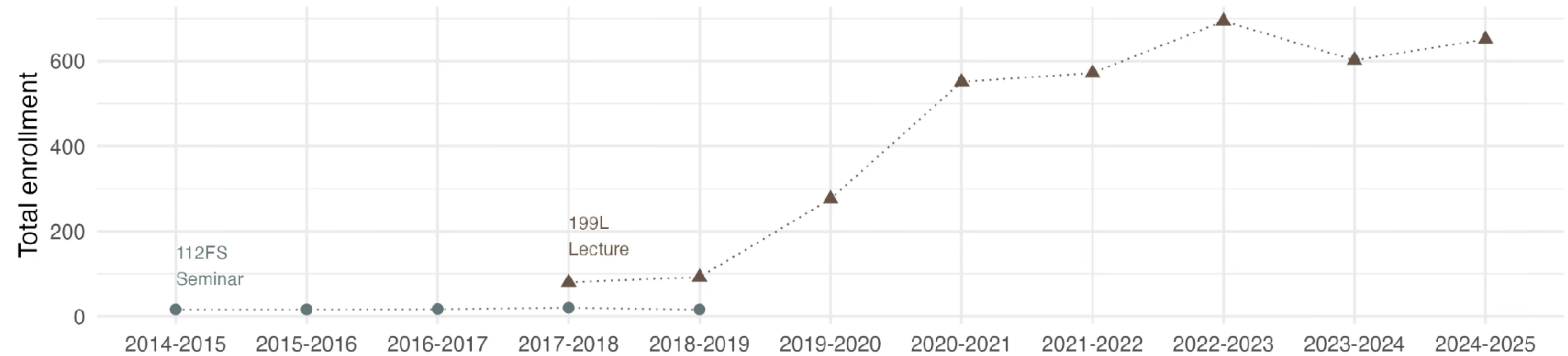
**good, helpful, and correct  
answers that come from course  
content**

and

**prefers terminology, syntax,  
methodology, and workflows  
taught in the course.**

# Motivation #1

Scale up help in a brave new world where reading forum questions and answers is not the cultural norm.



# Motivation #2

Generate “good” answers comparable to answers from the course instructor or TA that stays current with the course content

Another variable that needs some cleaning up is `membership_status`. Currently it should either be `NA` or contain some text that says “Select … to register for this group”. Recode this variable to say “Closed” if the current value is `NA` or “Open” otherwise. Save the resulting dataset with these two new variables as `clubs`, i.e., overwrite the data frame.

Then, display the first 10 rows of the dataset, moving `membership_status` to the beginning of the dataset to make sure it appears in the output in your rendered document.

You can perform this task in R using the following steps:

1. Recode the `membership_status` variable.
2. Move the `membership_status` column to the beginning.
3. Display the first 10 rows.

Here's an R script that accomplishes this:

```
r Copy Edit

# Assuming your dataset is already loaded as 'clubs'
# Recode membership_status: "Closed" if NA, otherwise "Open"
clubs$membership_status <- ifelse(is.na(clubs$membership_status), "Closed", "Open")

# Reorder columns to move membership_status to the beginning
clubs <- clubs[, c("membership_status", setdiff(names(clubs), "membership_status"))]

# Display the first 10 rows
head(clubs, 10)
```

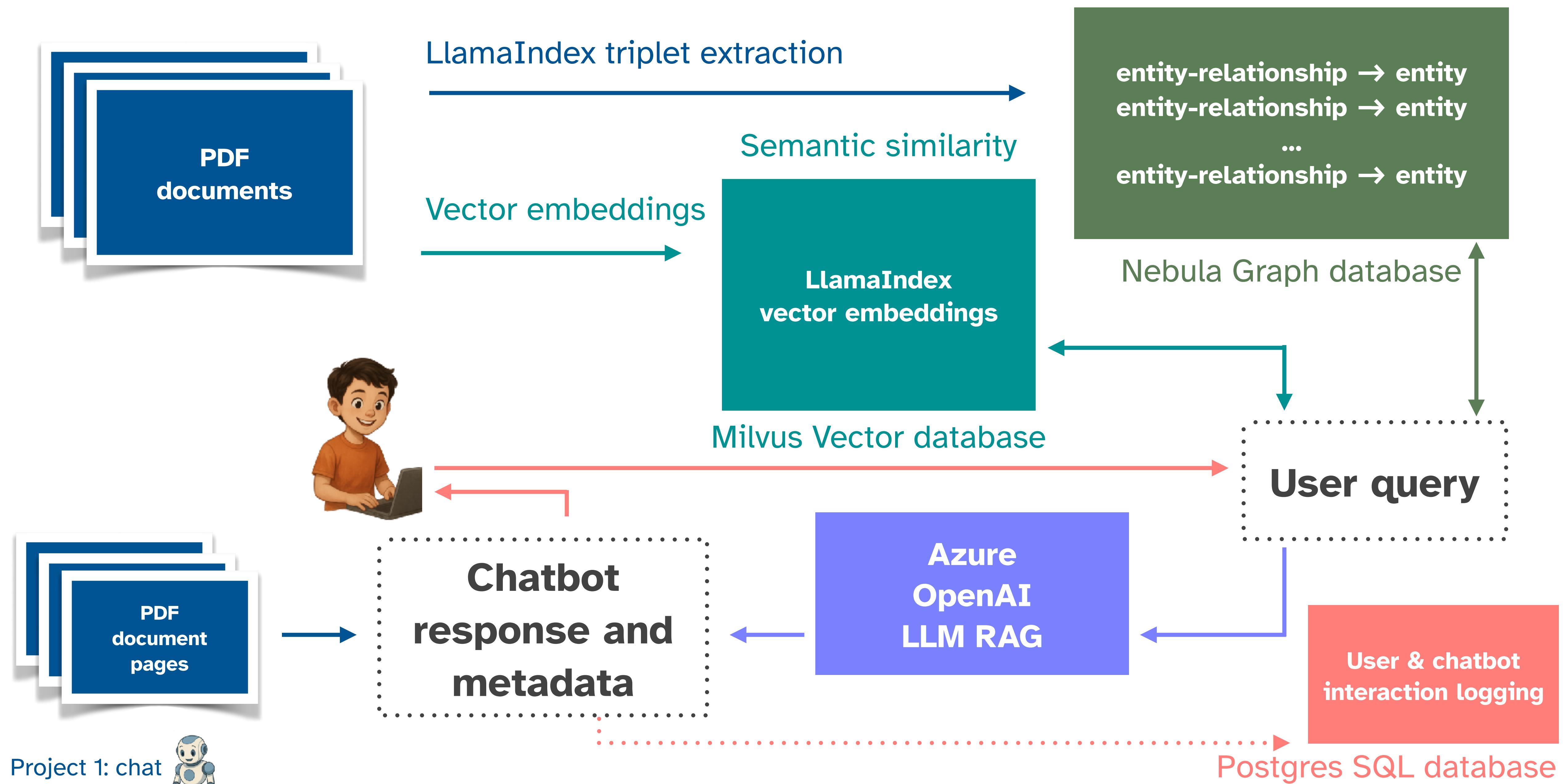


# Technical details

- \* Use RAG (Retrieval Augmented Generation) to focus chatbot on course content, give it context, and obtain pointers to specific pages of interest in the course textbooks:
  - \* **Knowledge Graph:** Searchable/traversable graph database of subject → predicate → object statements from text.
  - \* **Semantic Similarity:** Search identifies nearest neighbors based on word similarity using a vector database.
  - \* Relevant content from the course textbooks identified by combining semantic similarity and knowledge graph searches.
- \* Embed the chatbot into the Canvas Learning Management System as an LTI tool for student and instructor access.



# Architecture @ Duke



# Demo

Chat

source documents

Message

Type your message here...



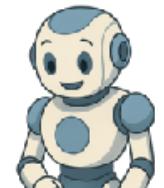
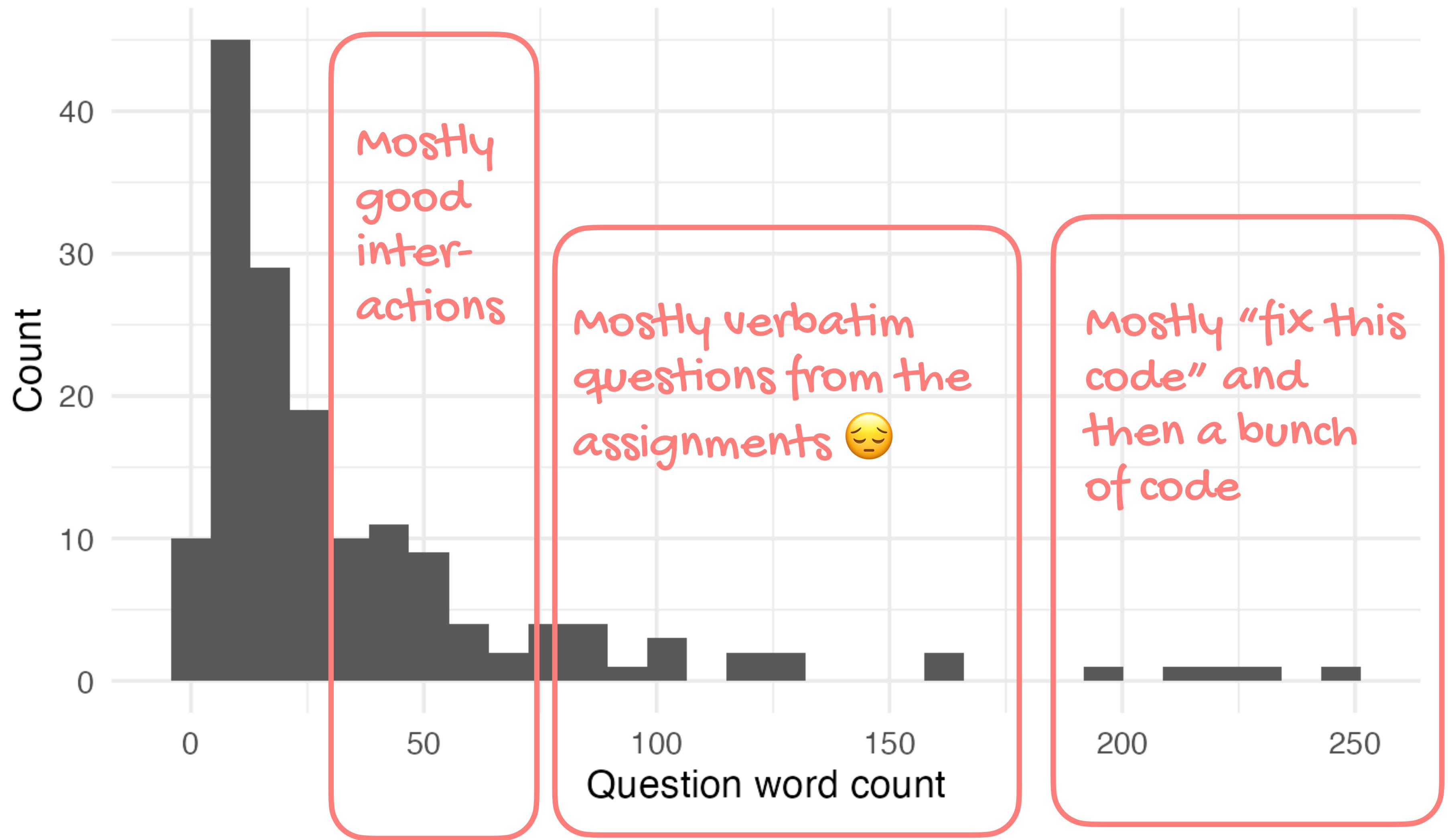
Retry

Undo



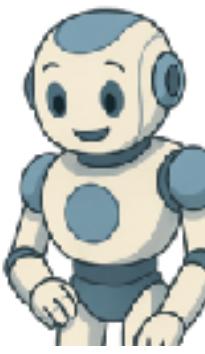
# Data

From Fall 2024



# Project 2

A feedback



that (hopefully) generates

**good, helpful, and correct  
feedback based on an instructor  
designed rubric**

and

**suggests terminology, syntax,  
methodology, and workflows  
taught in the course.**

# Motivation #1

An ever increasing number of students use AI tools as their first step, before thinking about how to approach a task.

## **Microsoft Study Finds AI Makes Human Cognition “Atrophied and Unprepared”**

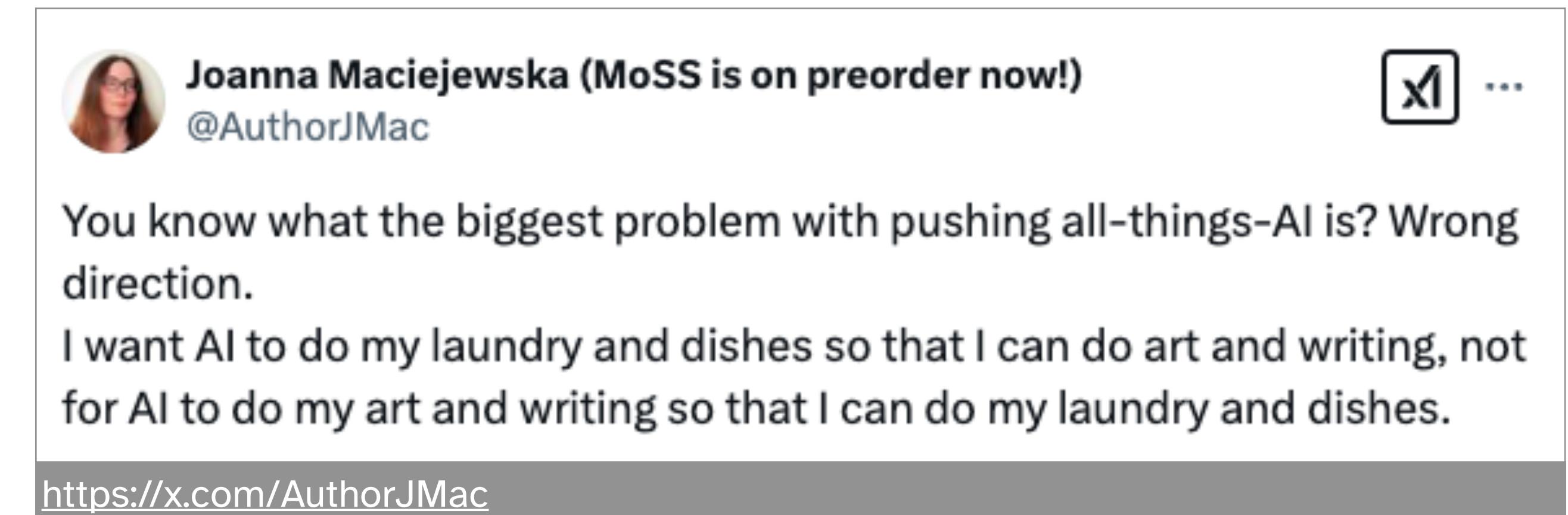
“[A] key irony of automation is that by mechanising routine tasks and leaving exception-handling to the human user, you deprive the user of the routine opportunities to practice their judgement and strengthen their cognitive musculature, leaving them atrophied and unprepared when the exceptions do arise,” the researchers wrote.

[404media.co/microsoft-study-finds-ai-makes-human-cognition-atrophied-and-unprepared-3](https://404media.co/microsoft-study-finds-ai-makes-human-cognition-atrophied-and-unprepared-3). Lee, Hao-Ping Hank, et al. "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers." (2025).

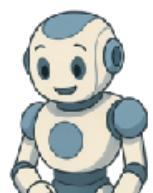


# Motivation #2

- \* How can AI support student learning instead of help them take shortcuts in their learning?
- \* Can AI help TAs redistribute their time towards more higher-value (and more enjoyable!) touch points with students and away from repetitive (and error-prone) tasks much of which go unread?

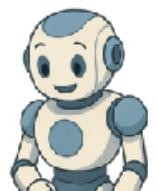


A screenshot of a tweet from the X/Twitter platform. The tweet is from a user named Joanna Maciejewska (@AuthorJMac), who has a profile picture of a woman with glasses. The tweet text reads: "You know what the biggest problem with pushing all-things-AI is? Wrong direction. I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes." At the bottom of the tweet is a URL: <https://x.com/AuthorJMac>. The X logo is visible in the top right corner of the tweet card.



# Motivation #3

- \* Self care! Neither the TAs nor I want to provide detailed feedback to answers generated solely with AI tools.
- \* Thanks to large numbers of students and TAs, and thanks to Gradescope, I'm already writing the darn detailed rubrics!



# activity.



## Activity #1

- \* Go to [duke.is/help-from-ai-uscots25](https://duke.is/help-from-ai-uscots25) and then Activity 1.
- \* **Step 1:** Read the question and the sample student answer and provide feedback to the student.
- \* Practice using LLMs to generate feedback for students using your favorite LLM (Chat GPT, Anthropic, Perplexity, etc.) and
  - \* **Step 2:** The question and the sample student answer alone
  - \* **Step 3:** The question, the sample student answer, and the rubric
  - \* **Step 4:** The question, the sample student answer, the rubric, and a customized prompt
- \* Goal: Fine tune your prompt until you're satisfied with the feedback (or until you're convinced this is not possible, but don't give up too quickly!)

**activity.**



## **Activity #2**

- \* Go to [duke.is/help-from-ai-uscots25](https://duke.is/help-from-ai-uscots25) and then Activity 2.
- \* Peruse...

# Technical details (Recap)

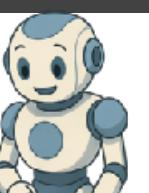
**TL;DR:** Use **prompt engineering** to ground feedback bot with the **question**, **rubric**, and **answer**.

```
library(ellmer)
library(glue)
library(tidyverse)

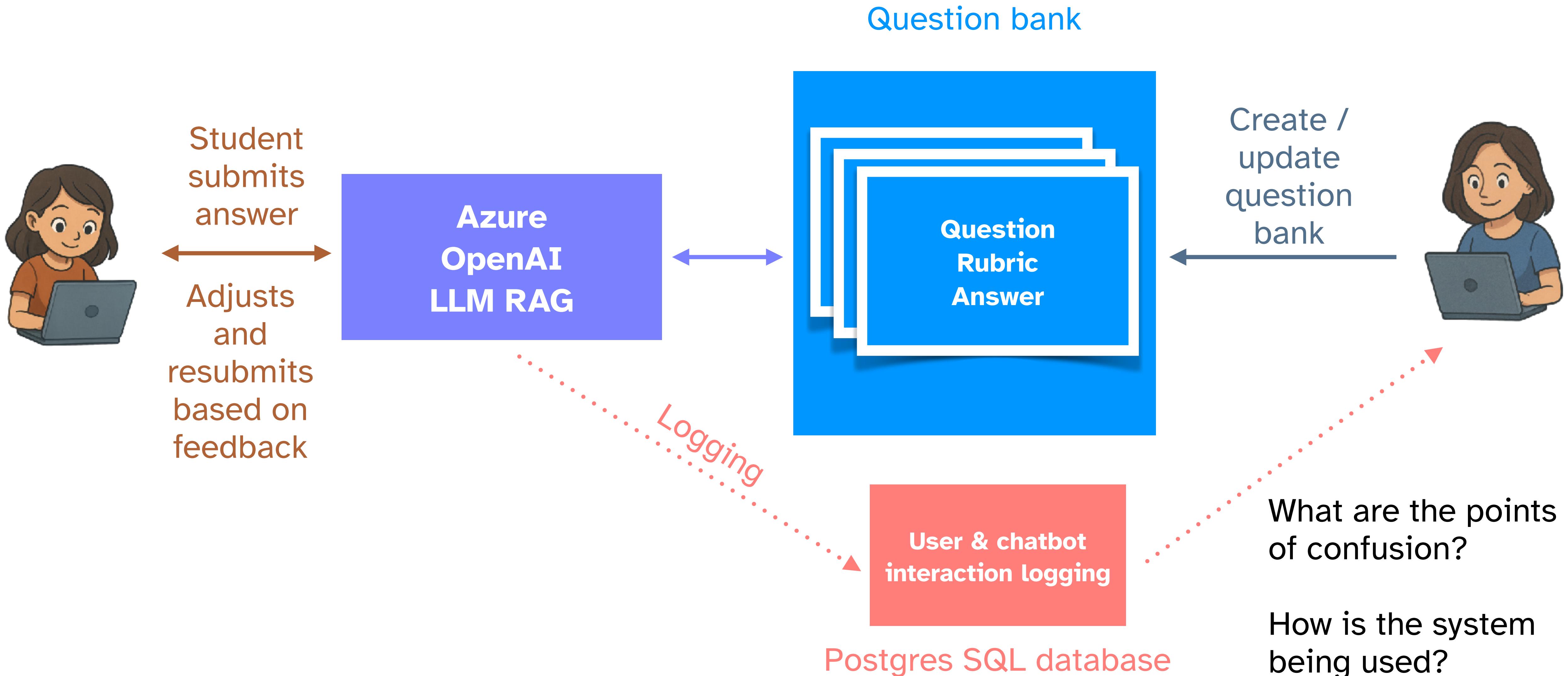
prompt ← function(question, rubric, answer){

  chat ← chat_openai(
    system_prompt = "You are a helpful course instructor teaching a course on data science with the R
programming language and the tidyverse and tidymodels suite of packages. You like to give succinct but
precise feedback."
  )

  chat$chat(
    glue(
      "Carefully read the {question} and the {rubric},
      then evaluate {answer} against the {rubric} to provide feedback.
      Please be certain to spell out your reasoning so anyone can verify them.
      Provide feedback in an output section named **Feedback:**.
      Format the feedback as bullet points mapping to the bullet points in the {rubric}."
    )
  )
}
```



# Architecture @ Duke



# Take aways

## The process

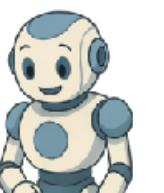
- \* Lots of fiddling with the rubric file, though unclear / hard to measure to what end.
- \* Separating out to `rubric` and `rubric_detailed` helps hide the answer while giving constructive feedback.



# Take aways

## The good

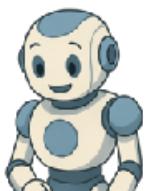
- \* “Spell out your reasoning” results in feedback that is too long, but taking that out and adding limits helps.
- \* It sort of works!



# Take aways

## The bad

- \* **The most concerning:** The feedback tends to catch errors but not the “good” and seems to reiterate the rubric item whether it’s met or not, potentially causing the student (who is already prone to this) to think “there’s no winning here”.
  - \* Somewhat on par with an inexperienced TA who is not seeing the bigger picture but looking at matching every detail to the rubric and pointing out any discrepancies whether they matter or not.
- \* **The inevitable:** Inconsistency in feedback from one try to another.
  - \* Is it possible to instill confidence in students when the feedback changes at each try on the same answer? Not substantially, but potentially enough for an inexperienced student...
- \* Hallucinations happen, somewhat consistently, e.g., “The code uses the base pipe (|>) and includes necessary spaces, but it lacks indentation, which can be improved for readability.” even when the code is properly indented.
- \* Text that would cause more problems gets injected into feedback, e.g., “aligning with rubric expectations”



# Next steps

- \* As new LLMs become available, continue model evaluation and tradeoffs between cost, speed, and accuracy with different approaches.
- \* Send prompt to multiple models with an additional prompt to summarize the various model results and evaluate if this leads to more accurate and consistent feedback.
- \* Continue system prompt enhancements and tuning.
- \* Measure learning outcomes for students using the LLM feedback and evaluate if this approach is “effective” (for a variety of goals).
- \* Consider packaging the LLM feedback tool for application to other fields.

Project 1: chat +



Project 2: feeedback

# acknowledgements.

Mark McCahill  
Computer Systems Architect  
Duke University

Duke University  
Office of  
Information Technology

duke.is/help-from-ai-uscots25

thank you.

Mine Çetinkaya-Rundel  
Duke University  
mc301@duke.edu

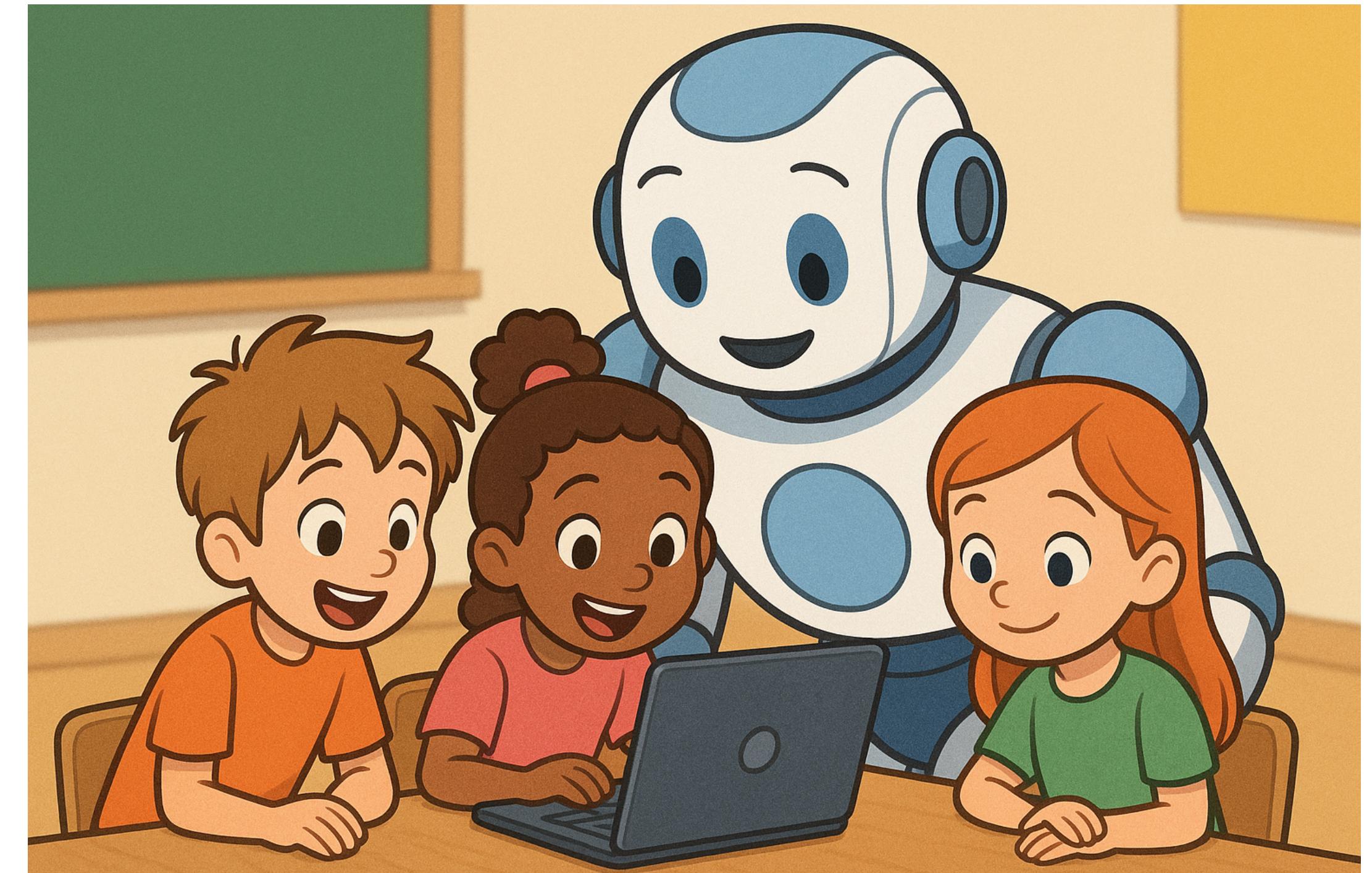


Image generated with ChatGPT with the prompt “Generate an image with a robot helping children. The kids should be looking at a laptop. And make the image rectangle, with height less than width.”

<https://chatgpt.com/share/682c8623-b100-8000-972c-e7384801436f>