Spark Curiosity

LLM-Driven Strategies for Teaching and Learning with Data

A hands-on workshop exploring how Large Language Models can transform data science education

USCOTS 2025

Unit A: Welcome and Framing the Workshop

In this workshop, we'll explore how Large Language Models can modernize data science education by enhancing both teaching efficiency and student engagement. Together, we'll discover practical strategies that you can implement immediately in your classroom.

Logistics & Tools

Room Essentials

- WiFi network & password
- Restrooms location
- Snack station (Available at 2:30 PM)

Prerequisites

- Software setup (R & RStudio)
- Google Doc link: **https://bit.ly/uscots-llm** or use the QR code
- Live updates and real-time collaboration
- Post questions

Introductions

Chester Ismay

Portland State University

Arturo Valdivia

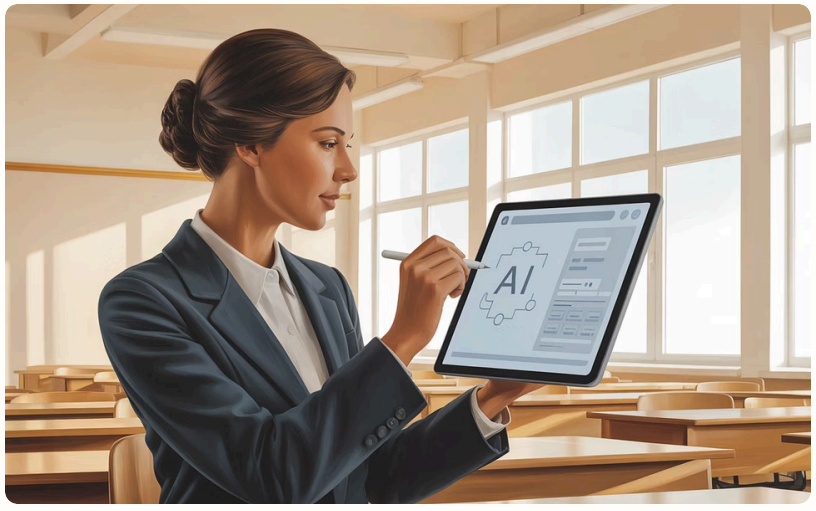
Indiana University

Icebreaker Activity

> In groups, discuss: "What's one teaching challenge you'd love to solve with AI technology?"

Share Your Challenge

Introduce yourself to your group and describe a specific teaching obstacle you face.



Brainstorm Solutions

How might AI tools help address these challenges? Be creative!

Select Top Ideas

Choose one solution to share with the larger group. Note your responses as you wish in **the Google Doc**.

Workshop Goals

Understand LLMs

Gain practical knowledge of how Large Language Models work and their potential in educational settings

Develop Skills

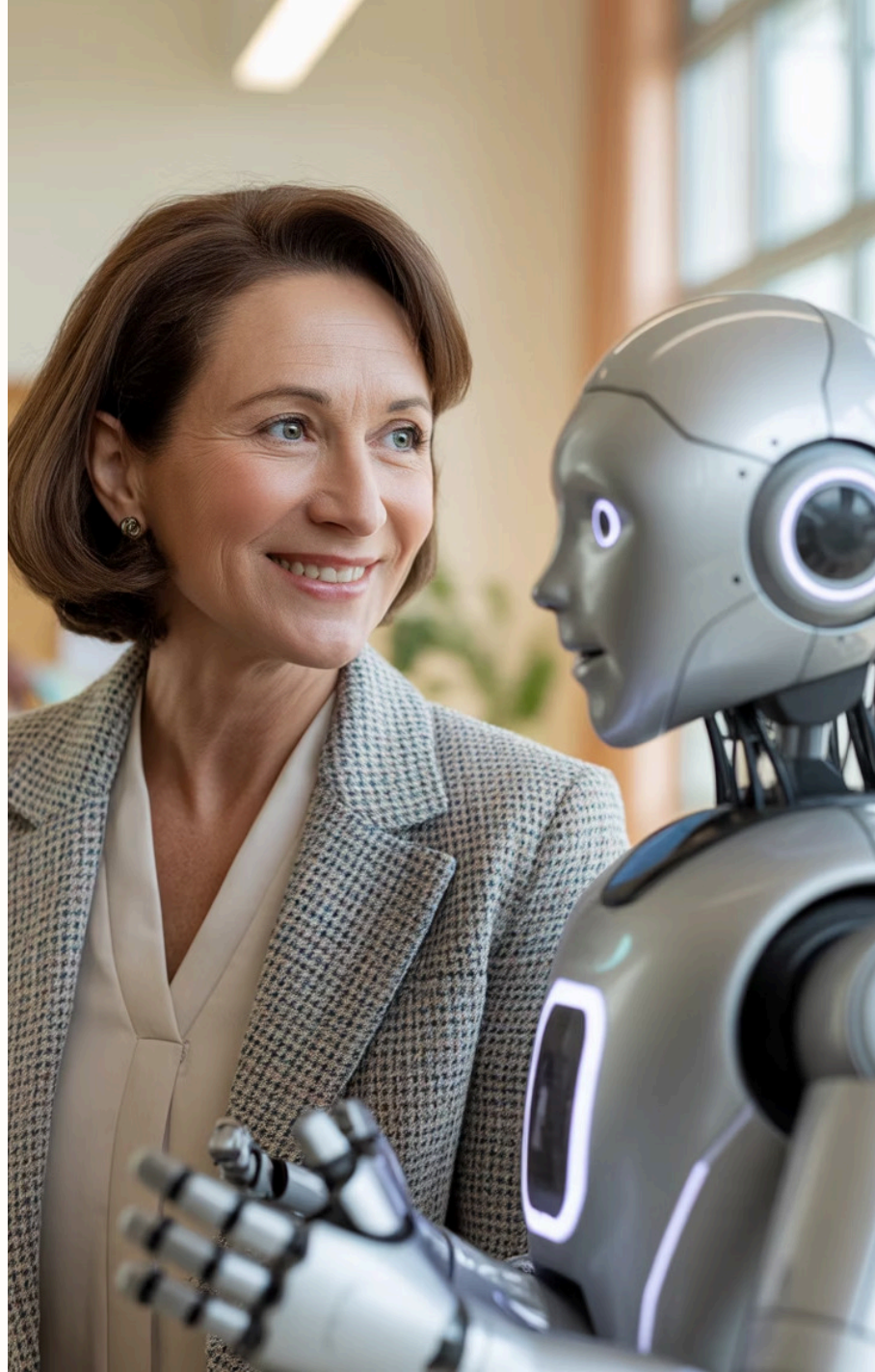
Learn to craft effective prompts and integrate AI tools like GitHub Copilot and ChatGPT into your teaching workflow

Create Materials

Design engaging learning activities and assessments that leverage AI capabilities

Build Community

Connect with fellow educators exploring AI-enhanced data science instruction



ModernDive + LLMs: A Natural Partnership

ModernDive's Approach

ModernDive takes data-driven and computational approaches to statistics and data science education.

Visit: **https://www.moderndive.com/v2/**

- Accessible statistical concepts
- Real-world data exploration
- Computational thinking

LLM Enhancement

LLMs extend this approach by:

- Generating customized examples
- Providing real-time assistance
- Scaffolding code development
- Creating personalized practice
- Explaining concepts in multiple ways

Unit B: The Promise of LLMs in Education

Exploring the potential of AI tools in the educational landscape

Understanding Generative AI

How LLMs Work

Large Language Models are neural networks trained on vast text corpora to predict the next word in a sequence.

- Pattern recognition across billions of examples
- Statistical prediction, not true understanding
- Capabilities emerge from scale and architecture

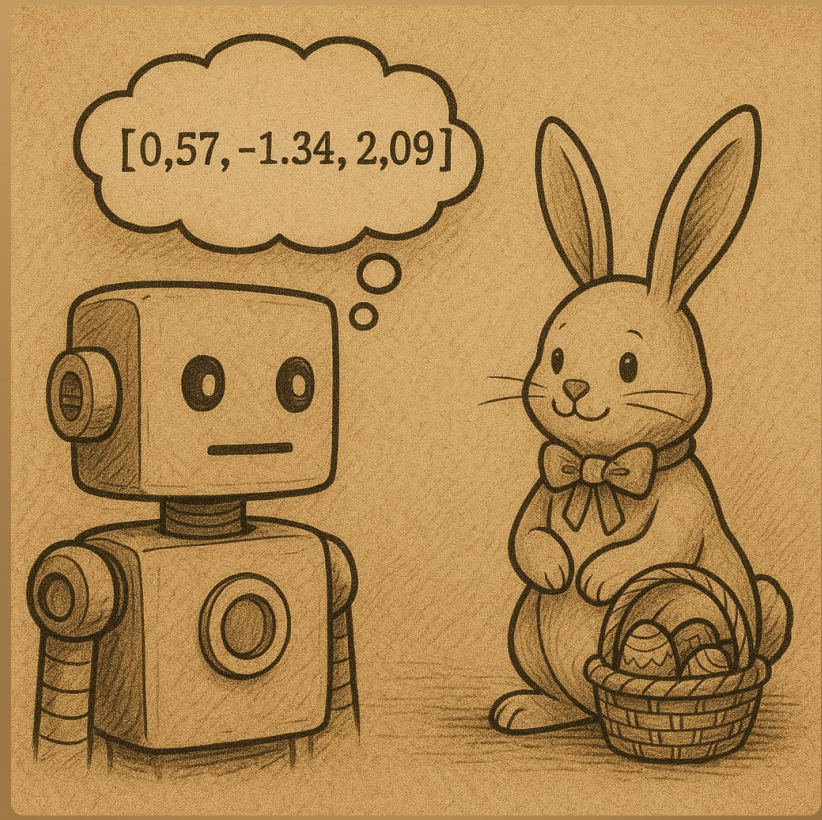
ONCE UPON A TEXT...

- Imagine teaching a child to speak using books alone
- No explanations, just reading patterns
- Transformers learn language the same way!

STEP 1: TURNING WORDS INTO MATH

Before an AI can understand language, it needs to convert text into something mathematical:

- Words become vectors
- Similar meanings = closer vectors
- Text becomes numerical data



STEP 2: PAYING ATTENTION TO WHAT MATTERS

On a sunny summer morning, people walked through the beautiful city park.

STEP 2: PAYING ATTENTION TO WHAT MATTERS

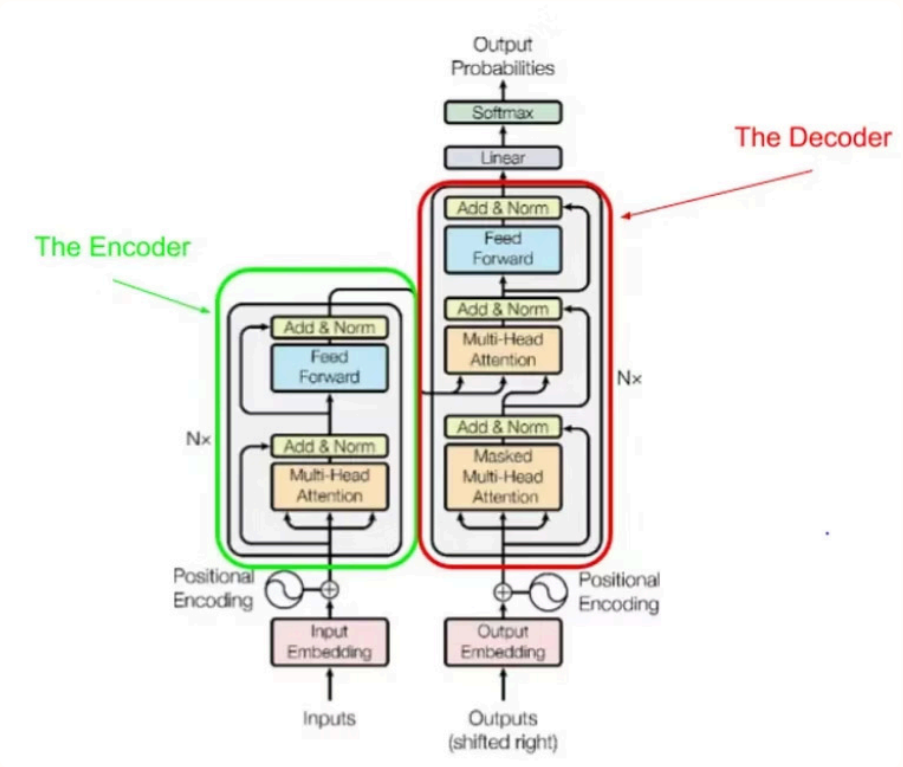
On a sunny summer morning, people walked through the beautiful city park.

Focus on important words

Assign attention weights

Handle long sentences

STEP 3: BUILDING A TRANSFORMER



Stacks of Attention + Feed-Forward layers

Modular design

Deeper layers = deeper understanding

STEP 4: TRAINING ON HUGE TEXT LIBRARIES

- Billions of words
- Predict next word repeatedly
- Patterns emerge naturally

STEP 5: PREDICT, PREDICT, PREDICT!

1. Predict next word/token
2. Assign probability scores
3. Sample or choose highest probability

LLM Landscape: Options for Educators

Model	Strengths	Limitations
ChatGPT	Widely accessible, strong code generation, frequent updates	Knowledge cutoff, occasional code errors, potential cost
Claude	Excellent at context understanding, nuanced explanations	Less programming focus, more formal responses
Gemini	Strong data analysis capabilities, Google integration	Newer platform, still evolving features
DeepSeek	Open source, customizable, privacy-focused	Technical setup required, less polished interface

AI Capabilities in Statistics Education

1 Automate Routine Tasks

Generate starter code, create practice datasets, and produce basic visualizations to save instructor preparation time.

2 Provide Real-Time Feedback

Offer immediate code debugging suggestions, explain statistical concepts in multiple ways, and adapt explanations to student questions.

3 Create Customized Content

Develop scenario-based problems relevant to student interests, generate examples at various difficulty levels, and produce contextual explanations.

4 Facilitate Data Exploration

Help students formulate research questions, suggest relevant variables to examine, and offer interpretation frameworks for findings.

Small Group Discussion

> In groups of 2–3, discuss:

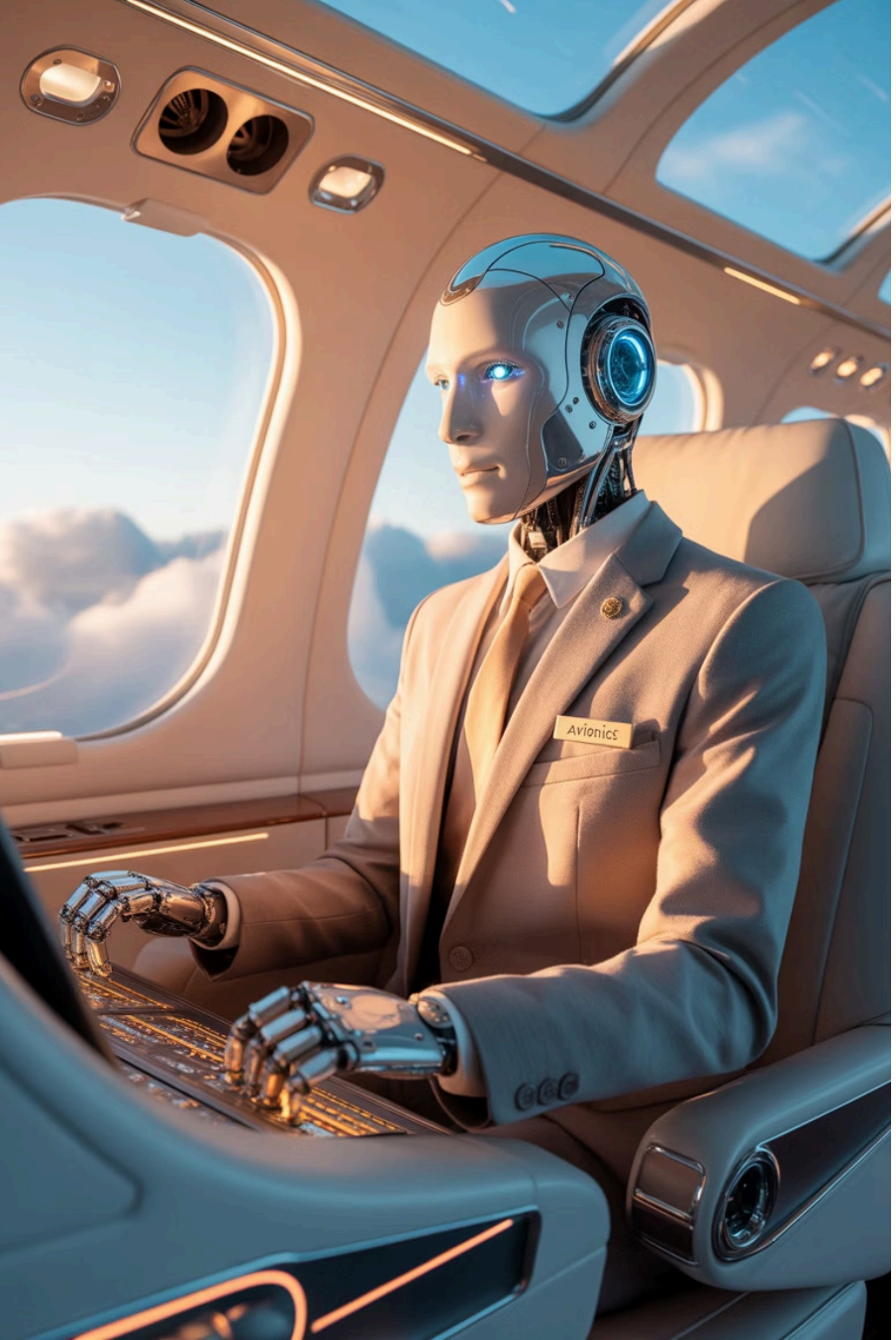
- A. What is your current familiarity with LLMs?
- B. What specific challenges in your statistics teaching might LLMs help address?
- C. What concerns do you have about implementing these technologies?

Take notes to share with the larger group in **the Google Doc**.

Whole Class Sharing

Discussion Prompts

- What were common teaching challenges identified?
- What innovative applications did you discover?
- What shared concerns emerged?
- What questions remain unanswered?



Unit C: Leveraging GitHub Copilot in RStudio

Exploring how AI pair programming can enhance the teaching and learning of R in data science education

GitHub Copilot Setup

What is GitHub Copilot?

An AI pair programmer that suggests code completions as you type, based on context and comments.

- Trained on public GitHub repositories
- Works inside your IDE (RStudio)
- Understands natural language comments
- Generates code from descriptions

Setup Process

1. If needed, create github.com account
2. Connect to RStudio

   Tools > Global Options > Copilot

3. Authenticate and activate

Guided Demonstration: GitHub Copilot in Action

1 Code Generation

Watch as we transform natural language descriptions into functional R code with minimal typing.

2 Real-World Examples

See how Copilot helps with loading datasets, creating visualizations, and conducting statistical analyses.

3 Best Practices

Learn effective commenting strategies that yield better suggestions and reinforce good documentation habits.

Exploring the Old Faithful Dataset with Copilot

With Copilot activated, enter code comments into your R script in RStudio one at a time.

```
# Install ggplot2 and moderndive packages


# Load the packages


# Plot waiting by duration for old_faithful_2024
```

Hands-on Activity: Practicing with Copilot

> Use the provided pseudocode prompts to generate code with GitHub Copilot

Prompt 1:

```
# Calculate average human
# development index
# for each continent
```

Let Copilot generate the code to load data, calculate means, and display results.

Prompt 2:

```
# Visualize the distribution of
# human development index overall
```

See how Copilot suggests different visualization approaches based on your comment.

Try modifying the generated code and see how Copilot adapts to your changes!

Reviewing Our Findings

Discussion Questions

- What worked well with Copilot?
- What challenges did you encounter?
- How might this tool impact how you teach coding?
- What classroom applications do you envision?

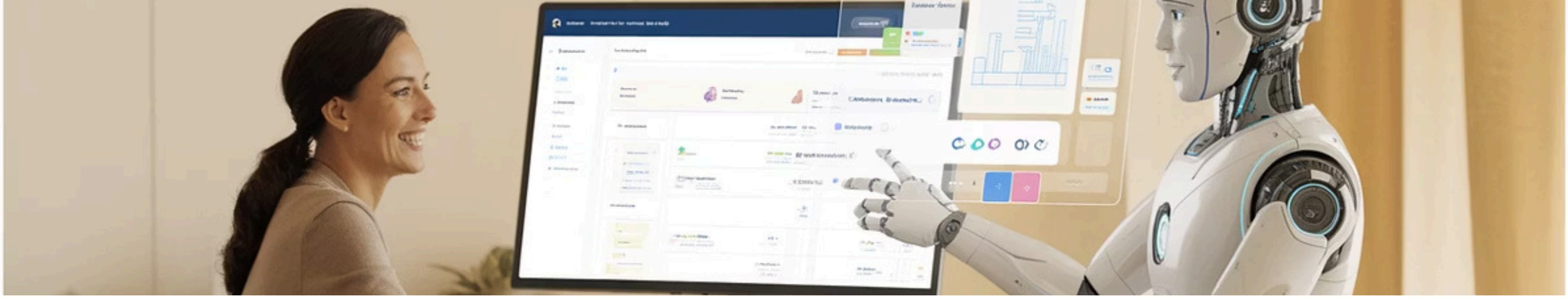
YouTube

10 Minute Timer

This timer counts down silently until it reaches 0:00, then a police siren sounds to alert you th...

10 Minute Break

Refreshments available in the hallway



Unit D: Integrating LLMs into Lesson Plans and Assessments

Practical strategies for incorporating AI tools into your statistics and data science curriculum

Using ChatGPT for Code Generation and Lesson Integration

Beyond Copilot

While Copilot excels at in-IDE code assistance, ChatGPT offers broader educational applications:

- Curriculum planning
- Content creation
- Differentiated examples
- Assessment development
- Conceptual explanations

ChatGPT Educational Applications

Generate Starter Code

Create R scripts for specific data analysis tasks, with detailed comments explaining each step for student understanding.

Create Assessments

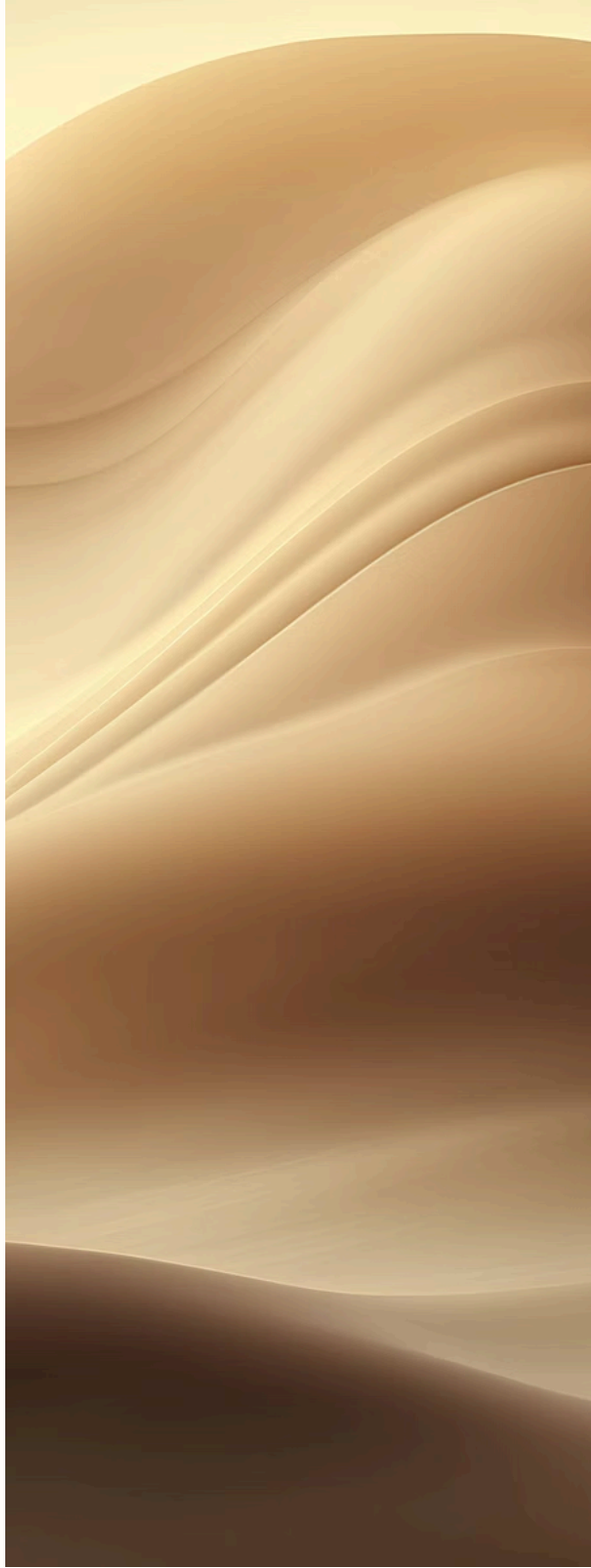
Develop quiz questions, detailed explanations, and code for simulated datasets at various difficulty levels.

Plain-Language Explanations

Generate multiple explanations of complex statistical concepts, tailored to different learning styles and backgrounds.

Visualization Code

Produce ggplot2 code for creating effective data visualizations that illustrate key statistical principles.



The Art of Prompt Engineering

Crafting Effective Prompts

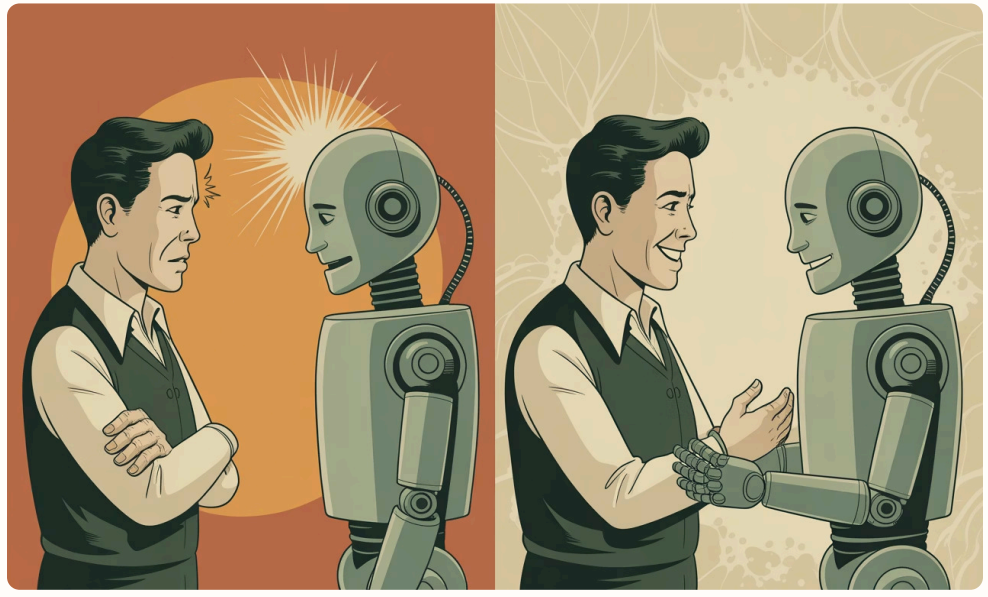
The quality of AI outputs depends heavily on how you structure your requests:

- Be specific about your goals and audience
- Provide context and constraints
- Request multiple variations
- Iterate and refine based on results
- Include examples of desired outputs

Think about working with LLMs as a conversation you are having with a colleague to solve a particular problem. You'll often need to refine what you are asking for, and that's OK!

Weak prompt: "Give me code for t-tests."

Strong prompt: "Create an R script with examples of independent and paired t-tests using the 'coffee_quality' dataset in the moderndive package. Include detailed comments explaining each step for undergraduate students who are new to hypothesis testing but familiar with R basics."



Reflection Activity

> Think of a statistical concept in your course where students struggle. What could you generate with ChatGPT to help them?

Take a few minutes to jot down ideas, considering:

- Specific student misconceptions or pain points
- Types of materials that might help (code, explanations, visualizations)
- How these materials would complement your existing teaching

Demonstration: Generating Teaching Materials

Custom Datasets

Generate realistic data with specific characteristics (skewness, correlation patterns, outliers) to illustrate statistical concepts

R Code Examples

Create annotated analysis scripts at varying complexity levels for scaffolded learning experiences

Assessment Items

Develop conceptual and computational questions with detailed answer explanations and rubrics

Visual Explanations

Generate image queries for creating fun scenarios to motivate topics

Designing AI-Enhanced Discussions

Using AI Outputs as Discussion Starters

- Debugging flawed code (intentionally generated with errors)
- Interpreting ambiguous results
- Comparing multiple approaches to the same problem
- Critiquing AI-generated explanations
- Improving AI-generated visualizations

These activities develop critical thinking and help students understand the strengths and limitations of AI tools.

Small Group Activity

Create an AI-Enhanced Lesson Plan

In your groups, brainstorm and outline a short lesson plan using ChatGPT to assist in teaching a specific statistical concept.

1. Choose a statistical concept (e.g., hypothesis testing, regression, sampling distributions)
2. Identify student challenges with this concept
3. Draft 2-3 specific prompts you would use with ChatGPT
4. Outline how you would integrate the AI-generated content into your lesson
5. Prepare to share your plan with the larger group

Sharing Lesson Plans

Group Presentations

Each group will briefly share:

- The statistical concept chosen
- Key prompts developed
- Integration strategy
- Anticipated benefits

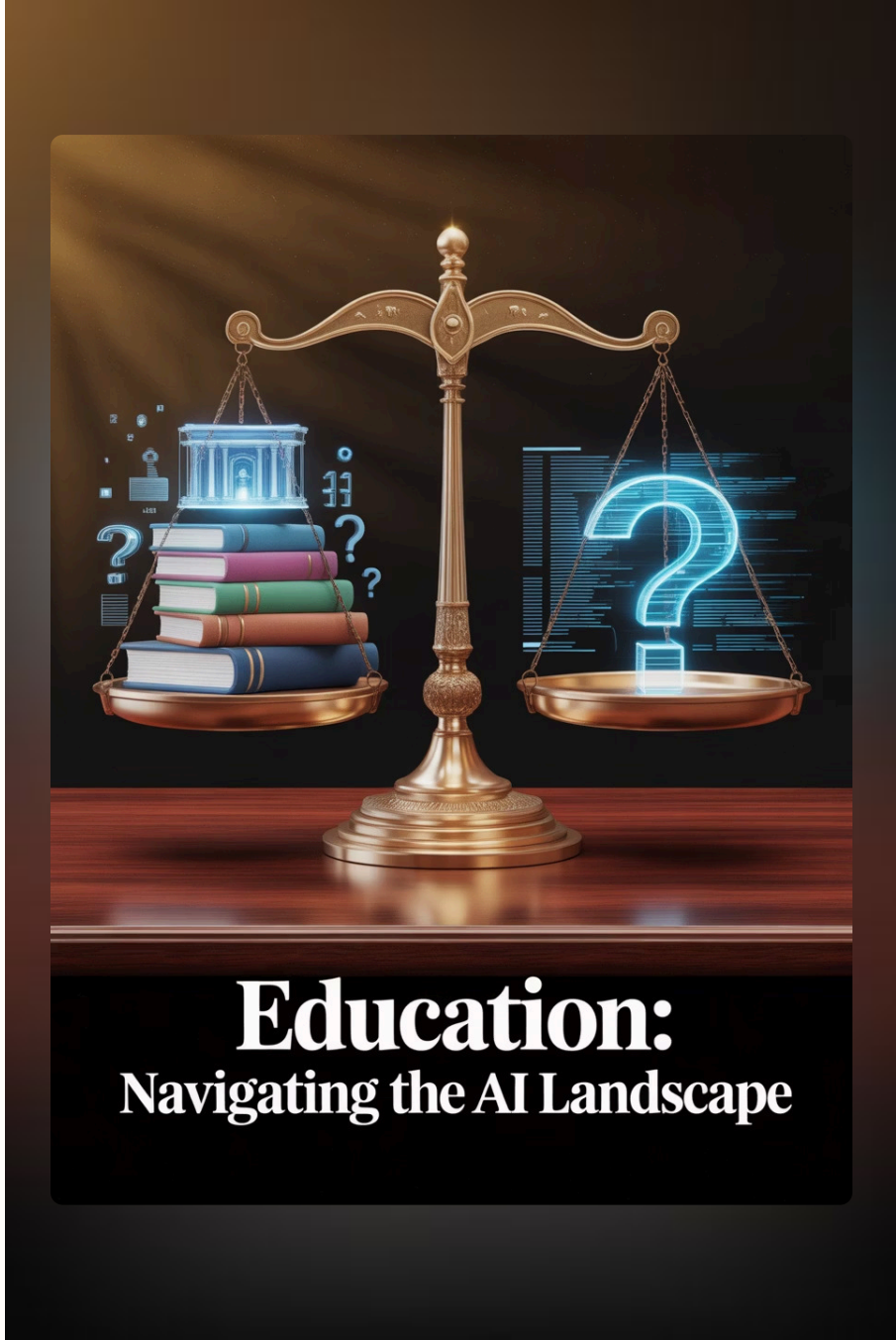
Collective Feedback

After all presentations, we'll discuss:

- Prompt refinement suggestions
- Additional integration ideas
- Potential challenges
- Assessment considerations

Unit E: Reflecting on Benefits and Limitations of LLMs in Education

Critically examining the implications of AI integration in statistics education

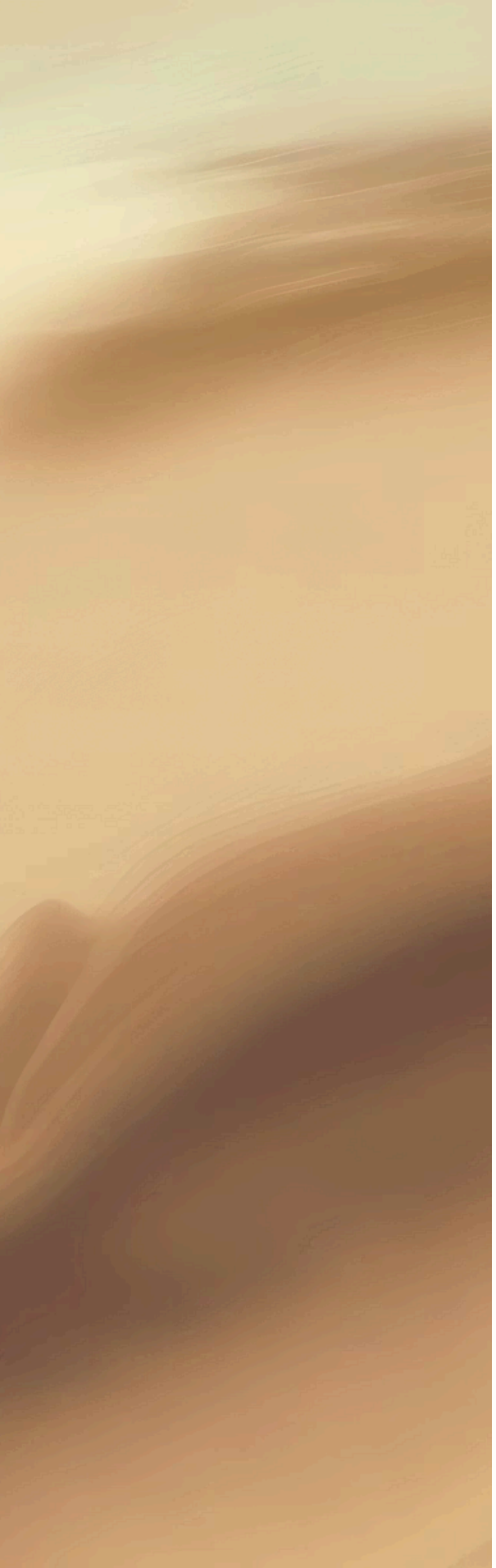


Small Group Discussion: Benefits and Limitations

Discussion Questions

1. What specific teaching tasks could LLMs make more efficient?
2. How might these tools enhance student engagement and learning?
3. What ethical concerns arise when integrating AI into education?
4. What skills might students miss developing if they rely too heavily on AI?

Discuss these questions in your groups, capturing key points to share.



Benefits of LLMs in Data Education

Enhanced Efficiency

Automate routine tasks like generating practice problems, creating starter code, and producing basic visualizations, freeing instructor time for higher-value interactions.

Creative Exploration

Empower students to explore more diverse questions and approaches, trying multiple methods without fear of wasting time on dead ends.

Coding Confidence

Reduce anxiety by providing scaffolding that helps students progress from partial understanding to functional code, building self-efficacy.

Personalized Learning

Offer explanations tailored to individual learning styles and backgrounds, helping diverse students grasp complex statistical concepts.

Limitations and Ethical Considerations

1 Accuracy Concerns

LLMs can generate plausible-sounding but incorrect explanations or code, requiring careful verification and developing students' critical evaluation skills.

2 Dependency Risks

Over-reliance may prevent students from developing fundamental problem-solving skills and conceptual understanding necessary for independent work.

3 Equity Challenges

Uneven access to AI tools can exacerbate educational disparities, while algorithmic biases may disadvantage certain groups in generated content.

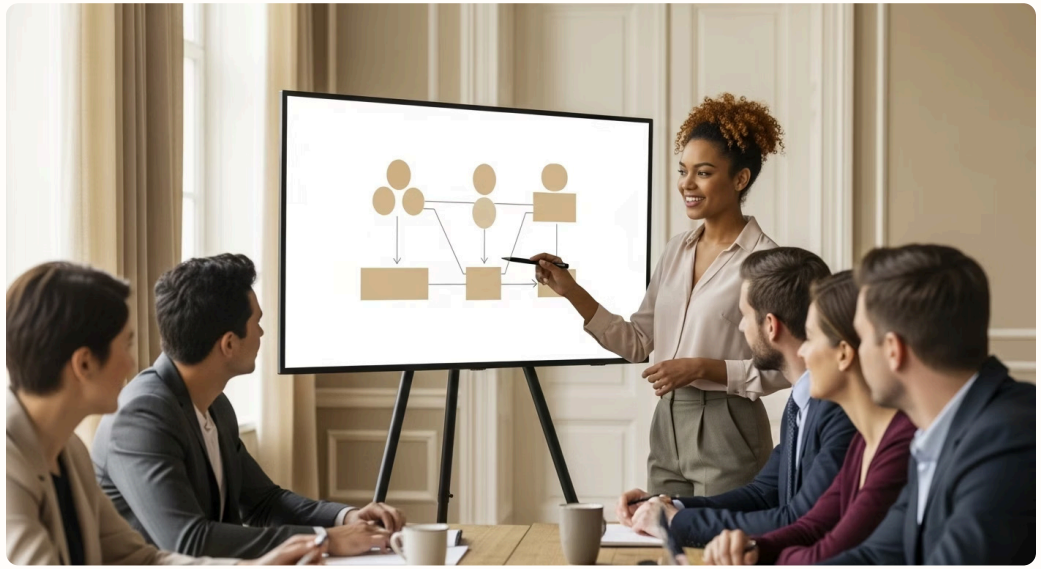
4 Academic Integrity

Traditional assessment methods may become obsolete, requiring redesigned approaches that evaluate higher-order thinking rather than code production.

Whole Group Summary

Key Takeaways

- What common benefits did groups identify?
- What shared concerns emerged?
- What balanced approaches seem most promising?
- What institutional policies might be needed?



Capture these insights into **the Google Doc** to create a shared resource for all participants.

Unit F: Wrap-Up and Next Steps

Consolidating insights and planning for implementation in your teaching practice

Reflective Prompt

> In the **shared Google Doc**, please respond to the following prompts:

1. Identify one way LLMs could transform a routine teaching task or a challenging lesson in your current curriculum.
2. What low-risk ways could you pilot these tools (e.g., optional resources, practice datasets)?

Take a few minutes to develop your thoughts. These reflections will help guide your implementation and connect you with colleagues pursuing similar innovations.

Course Review

1

Unit A: Welcome and Framing

We established workshop goals and connected LLMs to the data-driven, computational approach of ModernDive.

2

Unit B: Promise and Pitfalls

We explored generative AI theory and the capabilities, strengths, and limitations of various LLM platforms.

3

Unit C: GitHub Copilot

We practiced using AI pair programming to enhance coding efficiency and student learning in RStudio.

4

Unit D: Lesson Integration

We designed strategies for incorporating LLMs into lesson plans and assessments for statistics education.

5

Unit E: Benefits and Limitations

We critically examined the educational implications of AI tools and developed balanced approaches.

Thank You and Q&A

Contact Information

Chester Ismay: chester.ismay@gmail.com

Arturo Valdivia: arturo.valdivia@gmail.com

What remaining questions do you have about implementing these tools and strategies in your teaching practice?