ChatGPT, Chain of Thought, and Critical Thinking in the Classroom

Vern Richardson, University of Arkansas

Plan

- Al having dramatic impact on business education.
- How can we think about generative AI?
 - Can we use data analytics models to think about them?
 - Is answering questions we have or doing data analysis actually generative?
- Do we need to teach it to our students?
 - What parts? How do we do it?
- Examples Performing Analytics Using ChatGPT
 - I'll share prompts and these powerpoint slides with you.

How can we Harness Generative Alin Analytics?



I'll use the term ChatGPT, but recognize many other AI Chatbots, etc.



While some universities are trying to shut down all possible access to ChatGPT, we can harness it!



Can we think of it as a decision aid? A really smart assistant that gives us insights?

Ţ

Why do our students need this course?

Will it become a core course?

- All business school disciplines:
 - Have Questions
 - Have Relevant and Unique Data
 - Perform Analysis
 - Share Results with Management
- Students Love Hands-On Learning
 - Much Rather "Do" than "Hear"
 - We Remember (Source: Edgar Dale):
 - 10% of what we read
 - 20% of what we hear
 - 30% of what we see
 - 50% of what we see and hear
 - 70% of what we discuss with others
 - 80% of what we personally experience
 - 95% or what we teach others
- Perfect Class to Let them Personally Experience Asking Questions

Chat GPT – Isn't it just Data Analytics?

The AMPS Model (Also called the SOAR Model):

- <u>A</u>sk the Question
- <u>Master the Data</u>
- <u>P</u>erform the Analysis
- <u>Share the Story</u>



2025 RELEASE

1 0 May 400 350 290 320 45.00 DATA **ANALYTICS** for ACCOUNTING Mc Graw Hill

COST ACCOUNTING

A Data Analytics Approach

Margaret H. Christ | D. Kip Holderness Jr. | Vernon J. Richardson



INTRODUCTION TO DATA ANALYTICS FOR ACCOUNTING





ACCOUNTING INFORMATION SYSTEMS

Richardson | Chang | Smith FOURTH EDITION



FINANCIAL STATEMENT **ANALYSIS**

A DATA ANALYTICS APPROACH





VERNON J. RICHARDSON | MARCIA WEIDENMIER WATSO



Coming January 2026:

- **Data Visualizations** \bullet
- **Financial and** \bullet Managerial Accounting: A Data Analytics Approach

Coming January 2027:

Python for Business • Analytics

Student Comment #1: Tasks that required manual formula input were streamlined with ChatGPT, especially for repetitive data manipulations and summarizations. However, more advanced labs posed challenges, as generating accurate outputs occasionally required refining prompts or combining multiple instructions.

Chat GPT – Isn't it just Data Analytics?

- <u>Ask the Question</u>: Are we asking the right question?
 - Are we really specific about what we're asking?
 - What is the role of prompt engineering?
- <u>Master the Data</u>: What data is it sampling/evaluating? Did we specify it? Do we know if it follows ethical standards? Do we know if it is biased? Where does it come from? Could the data we input be shared with others?

Ask the Ouestion

Perform the

Analvsis

Master

the Data

Share the

Story

- <u>Perform the Analysis</u>: What technique is being used? Is it the correct technique? Is it backward looking or forward looking? What did it just do? Is it a black box? Is it explainable? Do I know what it did? How do we know if it is not a hallucination? Can I understand the underlying python code?
- <u>Share the Story</u>: Is the deliverable something the decision maker can use?
 - Does it communicate findings in an optimal way?
 - Did we think about the consumer?

Student Comment #4: The other conclusion that I came to after this project has been that AI, like a calculator, is only as good as the information that you provide it with.

Prompt Type	Explanation	Example
Zero-Shot	Submitting a single prompt without examples (Kojima et al. 2023)	When sales volume is 2000 units, unit sales price is \$15, unit variable cost is \$9 and total fixed costs are \$20,000, the break-even point is:
Few-shot	Submitting a series of example prompts and responses. Brown et al. (2020)	 When unit sales price is \$25, unit variable cost is \$12 and total fixed costs are \$73000, then the break-even point is 5,616. When unit sales price is \$16, unit variable cost is \$13 and total fixed costs are \$75689, then the break-even point is 25,230. When unit sales price is \$91, unit variable cost is \$70 and total fixed costs are \$8723845, then the break-even point is 415,422. When unit sales price is \$33, unit variable cost is \$24 and total fixed costs are \$33456, then the break-even point is 3,718. When unit sales price is \$15, unit variable cost is \$9 and total fixed costs are \$20000, then the break-even
		point is:
Col (Chain of Thought)	Having the GenAl provide an explanation of its results for verification. Wei et al (2022)	When unit sales price is \$15, unit variable cost is \$9 and total fixed costs are \$20,000, What is the break- even point? Let's think step-by-stepOR When unit sales price is \$25, unit variable cost is \$12 and total fixed costs are \$73000, what is the break- even point in units?
		To get break-even point, divide the total fixed cost by the contribution margin per unit. The contribution margin per unit is the price per unit minus variable cost per unit. So, break-even is \$73,000/(\$15 - \$13) = 5,616. When unit sales price is \$15, unit variable cost is \$9 and total fixed costs are \$20000, what is the break-even point?
Role-Based	Explicitly stating the role the AI is to impersonate. Kong et al.(2024)	You are an expert management accountant. When unit sales price is \$15, unit variable cost is \$9 and total fixed costs are \$20,000, What is the break-even point?

Prompt Engineering – Ask the Question

- *Prompt engineering* is the process of guiding AI to generate the most relevant and high-quality output.
 - We need to be specific
 - If you want to generate specific output, you need to provide a specific prompt or question.
 - For example, the prompt "What are some business risks?" will be vaguer than "What are five business risks from an auditor's perspective?" or from a banker's perspective?
- We need to provide context.
 - <u>Specify the role</u>. For example, if you want to view a problem from an investor's point of view, start your prompt with "Act as an investor..."
 - If you want the AI to ask you for <u>additional context</u> for a complex analysis, include "Ask me any questions you need for context."

Prompt Engineering – Ask the Question

Student Comment #2: One of the challenges I faced was figuring out how to write prompts that ChatGPT could understand. If I wasn't specific enough, it sometimes gave answers that weren't exactly right. This taught me the importance of being clear and detailed when asking questions or giving instructions, just like in school or any project.

Student Comment #3: Unlike Excel, where errors are often a direct result of user formulas, missteps in ChatGPT were likely from broad prompts. This experience emphasized the critical role of human input in guiding AI tools effectively.

Large Language Models are Powerful!

Financial Statement Analysis with Large Language Models

Alex G. Kim¹ Maximilian Muhn² Valeri V. Nikolaev³ This draft: May 20, 2024

Abstract

We investigate whether an LLM can successfully perform financial statement analysis in a way similar to a professional human analyst. We provide standardized and anonymous financial statements to GPT4 and instruct the model to analyze them to determine the direction of future earnings. Even without any narrative or industryspecific information, the LLM outperforms financial analysts in its ability to predict earnings changes. The LLM exhibits a relative advantage over human analysts in situations when the analysts tend to struggle. Furthermore, we find that the prediction accuracy of the LLM is on par with the performance of a narrowly trained state-ofthe-art ML model. LLM prediction does not stem from its training memory. Instead, we find that the LLM generates useful narrative insights about a company's future performance. Lastly, our trading strategies based on GPT's predictions yield a higher Sharpe ratio and alphas than strategies based on other models. Taken together, our results suggest that LLMs may take a central role in decision-making.

A Skill that Needs to Be Taught: Prompt Engineering

Human Processing and Chain-of-Thought Modern large language models can retrieve numbers from structured tables and perform simple calculations. However, they lack the ability to reason like a human and perform judgment. Recent research suggests that chainof-thought prompting can significantly enhance the reasoning and problem-solving abilities of large language models (Wei et al., 2022).

We implement the CoT prompt as follows. We instruct the model to take on the role of a financial analyst whose task is to perform financial statement analysis. The model is then instructed to (i) identify notable changes in certain financial statement items, and (ii) compute key financial ratios without explicitly limiting the set of ratios that need to be computed. When calculating the ratios, we prompt the model to state the formulae first, and then perform simple computations. The model is also instructed to (iii) provide economic interpretations of the computed ratios. Then, using the basic quantitative information and the insights that follow from it, the model is instructed to predict whether earnings are likely to increase or decrease in the subsequent period. Along with the direction, we instruct the model to produce a paragraph that elaborates its rationale. Overall, this set of instructions aims to replicate how human analysts analyze financial statements to determine whether a firm's performance is sustainable (Bouwman et al., 1987).

Source: Citation on prior slide. Kim et al. Working paper.

How do we teach this students?

- Could your students do a "Chain of Thought"? Could I do a "Chain of Thought"?
 - Could they walk through the steps of a problem? Can you and I?
- That ChatGPT is not always right. And how will they know?
 - Did they get the same conclusion twice asking the exact same question twice?
 - Excel/Python first, and then ChatGPT?
 - Does it use RAG (Retrieval Augmented Generation)?
 - the process of optimizing the output of a large language model, so it references an authoritative knowledge base outside of its training data sources before generating a response. (Source: https://aws.amazon.com/what-is/retrieval-augmented-generation

Student Comment #5: ChatGPT always responds confidently, whether the answer is correct or not. This makes it crucial to double-check the results, especially for technical or high-stakes tasks.

Student Comment #6: This project has been very eye opening for me, because I think that previously I had been under the impression that AI, especially the paid version, would automatically get it right when it comes to doing the kind of work needed. However, I saw first-hand how even a premium generative language model could misinterpret the information provided.

Towards Critical Thinking Skills



Final Project (in my class this semester)

- Perform six of the labs we did this semester in ChatGPT.
 - Good starting point.
- Provide output of the visualizations created and the prompts used.
- Summarize your learnings, addressing these questions:
- Was it harder or easier than doing them in Excel/Power BI/Tableau?
- Is this just another tool? Or is it more than that?

Student Comment #8: I was thoroughly surprised by how simple it was to use ChatGPT to perform a lot of the different analyses we used throughout the semester.

Student Comment #9: Working on these labs with ChatGPT was a very cool experience! It was fascinating to see how it handled different tasks, from quick solutions to harder challenges.

Student Comment #10: It was cool seeing how powerful an AI with the ability to analyze and compute when given a spreadsheet to look at.

Decisions We Must Make (ChatGPT 4.0)

- Ask the Question
 - From a specific point of view?
 - Point of view of an investor, or a regulator? Or a potential lender?
- Master the Data
 - Provide own as attachment (generally requires paid version)?
 - Ask ChatGPT to Gather Data?
- Perform the Analysis
 - Chain of Thought do students know how to get to the end?
- Share the Story
 - Output to the screen, or to the file?
 - Output input prompts?

Master (Read in) the Data

- Was full dataset read in?
 - Run summary statistics to make sure?
 - Can we have it list the last five lines to make sure it read it correctly?
- Generally will not go retrieve data. (Anscombe's Quartet)
- Do we need to provide a data dictionary?

Attach Lab 2-2 Data.

<u>Initial</u> Prompt: "Can you show the first five lines of the dataset?"

<u>Follow</u>-Up Prompt: "Can you show the last five lines of the dataset?"

Perform the Analysis

- What type of analytics am I trying to do?
 - What type of analytics question am I trying to answer?
- Am I doing a hypothesis test?
- Am I doing what-if analysis?
- Can ask ChatGPT what inputs it needs?
- Analyst still needs to lead it.

Perform the Analysis Chain of Thought (with ChatGPT)

- <u>*IBA Lab 5-6</u>
- 2 Uploaded File: "Would you show me the last five lines of the dataset?"
- I Create Holiday Variable: "I'd like to create a dummy variable titled Holiday, where Holiday equals 1 in the month of November and December and equals 0 otherwise."
- 2 Create Trend Variable: "Please also create a trend variable counting days for each row starting at 2021-01-01."
- 2 Perform Regression: "Can you now do a regression of Monthly Product Demand on GDP, Weather, Holiday and Trend?"
- 2 Create Scatterplots: "Would you provide a scatterplot of Monthly Product Demand and Weather as well as a scatterplot of Monthly Product Demand and Holiday?"
- 2 Output Files: "Could you output the regression results as a file? And the scatterplots as another file?"

Share the Story

- Output on screen or in file?
- Output table or visualization?
- Who ordered up the visualization the analyst or ChatGPT? Do your axes to start at zero, etc.?
- Can students then tell the story?
- Output all prompts?

Student Comment #7: It is also not the best if you want visualizations. It can perform some visualizations great like bar, histogram, and pie charts but when it comes to more advanced visualization like a filled map chart it does not perform well.

Examples (as time permits)

- Mortgage Amortization Schedule
- Sensitivity Analysis Sensitivity of Profits to Differing Levels of Variable and Fixed Costs and Sales Volume
- Evaluating the Relationship between R&D Expenditures and Sales

• Benford's Law - Chi-square Statistic

GenAl Labs Being Built into Texts

	Number of Labs (Excel, Tableau and Power BI)	GenAl (ChatGPT and Gemini) Labs by Summer 2025
Accounting Information Systems	15	+5
Cost Accounting	78	+10
Data Analytics for Accounting	55	+10
Financial Statement Analysis	41	Summer 2026
Introduction to Business Analytics	92	+10
Introduction to Data Analytics for Accounting	85	+10

Your thoughts?

- Is this just another tool?
- How/when do we teach this?
- Start now? Next year?
 - That is a really good question about Excel vs. ChatGPT. I don't think the capability of ChatGPT is quite complete and will continue to grow in the future with easy access to competitor and industry data, and near effortless analysis. I guess we'll learn into it.

Have input? Need ideas?

Contact me at <u>vjricha@uark.edu</u>

Prompts for Labs we Did

Mortgage Amortization Schedule

- I would like to create a mortgage amortization schedule? What input do you need?
- The loan amount is \$200,000. The annual interest rate is 6%. The loan term is 360 months. Start date is 1/1/2025.
- What would be the total interest paid if the payments were over 360 months, 180 months or 120 months?
- Can you show a graphical comparison?
- Please output the mortgage amortization schedule in an Excel file. And output the prompts in a word document.

Prompts for Labs we Did

Cost-Volume-Profit Analysis

Prompt 1: Start a New Chat. This resets the chat to start with a fresh set of data and instructions.

Prompt 2: I'd like to perform a sensitivity analysis analyzing changes in profit as sales, costs and prices change. How do I do that?

Prompt 3: The base case is units sold is sales volume of 2000 units, unit sales price is \$15, unit variable cost is \$9 and total fixed costs are \$2000. Calculate the profit for this base scenario.

Prompt 4: For the base case scenario, how many units must be sold to break even?

Prompt 5: Allow sales price to vary between \$10 and \$20. And allow unit variable price to vary between \$6 and \$12. Calculate the profits for each scenario.

Prompt 6: Please show this sensitivity analysis as a heat map visualization.

Prompt 7: Can you change the heat map to go from lowest profits in red, middle profits in yellow and highest profits in green?

Prompt 8: Show the level of breakeven in sales units for each level of variable prices.

Prompt 9: Please output the heat map as a .png file titled "Cost Volume Profit Heat Map".

Prompt 10: Please output the prompts I gave you in a Microsoft Word file titled "Cost Volume Profit Prompts".

Prompts for Labs we Did

What is the nature of the relationship between R&D expenditures and sales revenue?

Prompt 1: Start a New Chat.

Prompt 2: (Use the virtual paper click to attach IDAA Lab 7-9 GenAI Data.xlsx.)

Prompt 3: Please display the last five lines of data.

- Prompt 4: Please create a scatterplot for each data point with revenues on the y-axis, and R&D Expenditures on the x axis.
- **Prompt 5: Please insert a trend line into the scatterplot and provide the regression equation and related** statistics for that trend line.

Prompt 6: If the R&D expenditures are 5000000, what could we expect sales to be?

Prompt 7: Please output that scatterplot with the trendline as a .png file titled "IDAA Lab 7-9 GenAI Scatterplot and Regression Results".

Prompt 8: Please display the regression results file titled "IDAA Lab 7-9 GenAI Regression Results".

Prompt 9: Please output the prompts I gave you in a Microsoft Word file titled "IDAA GenAI Lab 7-9 Prompts".