

# AI and Cost Estimation

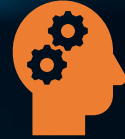
Data Science's Expanding Role in Cost Estimating  
ICEAA Conference May 2024  
Data Science & Machine Learning Track (DML07)

Dan Harper

Kevin McKeel

Special thanks to Greg Forman

# Overview



Purpose &  
Background



IT Trends



Application to  
Cost Estimation



AI Use Cases  
and Examples



Considerations  
and Caveats

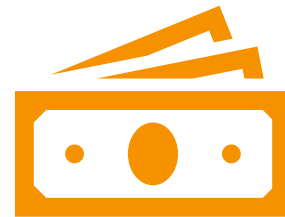


Resources

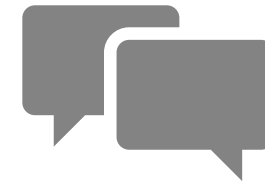
# Purpose



Provide basic overview of key AI concepts and developments



Introduce Applications of AI relevant to Cost Estimation



Initiate further study and discussion

# Background

- Kevin and Dan have backgrounds and interests in cost estimation, automation, and in general finding more efficient ways to help our US DoD and federal sector clients make informed acquisition decisions
- Dan produces the Cost, Acquisition, Manufacturing and Estimation Log (CAMEL)
- We worked for PRICE Systems (now Unison) in 2000s, and have worked for a variety of leading Cost Engineering firms (Dan: Wyle, Booz Allen; Kevin: MCR, TASC, Logapps)
- Kevin and Dan presented on the top of Artificial Intelligence (AI), Generative AI and Cost Estimating at the 2023 International Cost Estimating and Analysis Association
- Since then, we have presented to Project Controls Expo, Joint IT, Software Cost Forum (<https://www.dhs.gov/joint-it-and-software-cost-forum>), and Boehm Center for Software Estimating Annual Research Review
- We are not Comptrollers...so take everything we say with a grain of salt!

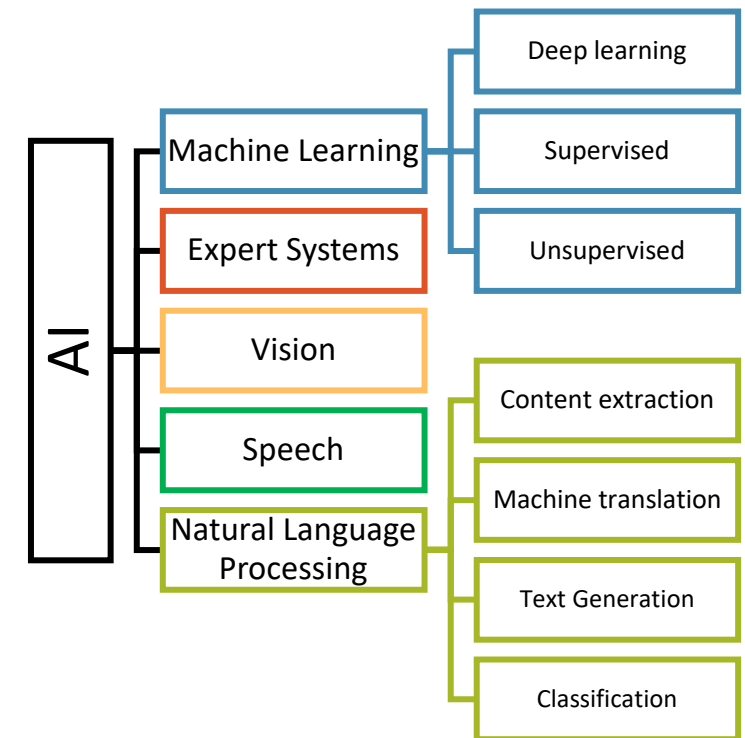
# Poll Question

**Have you every used an AI tools such as ChatGPT for any purpose?**

- a. Yes**
- b. No**

# Key AI Terms and Definitions

- **Artificial Intelligence (AI)** | “The science and engineering of making intelligent machines, especially intelligent computer programs” – John McCarthy
- **Machine Learning (ML)** | Algorithms that can learn from and make predictions or decisions based on data
  - **Artificial Neural Networks (ANN)** | Class of ML algorithms designed to conceptually mimic the neural network (neurons and synapses) in the brain.
- **Expert Systems** | rule-based systems that make decisions and solve problems in a specific domain
- **Natural Language Processing (NLP)** | a subfield of AI focused on the intersection of linguistics, computer science, information engineering, and the interactions between computers and human languages
- **Generative Adversarial Networks (GAN):** Type of Deep Learning model that can generate new combinations of output, such as images or text, from data it has been trained on – ChatGPT, DALL-E, Stable Diffusion are examples



# IT Trends

- Data generated by individuals, businesses and governments is increasing exponentially<sup>1</sup>
- Artificial intelligence is currently most powerful technology for tackling big data challenges, but implementation historically complex and resource intensive
- As AI capabilities grow, and barriers of implementing AI solutions are reduced, more domains will be integrated with or rely on some form AI
- Recently popularized “Generative” AI models pave the way more intuitive general purpose AI abilities, enabling widespread use by the general public, potentially accelerating investment and innovation
- Models like DALL-E (OpenAI) and Stable Diffusion (CompVis) disrupting industries once considered well insulated from AI, such as digital art and music
- ChatGPT (OpenAI) “fastest-growing consumer internet app ever” – 100 Million users in 2 months!

# IT Trends-ChatGPT

- ChatGPT is a NLP tool driven by AI technology that allows you to have human-like conversations and much more with a chatbot. The language model can answer questions, and assist you with tasks such as composing emails, essays, and code. It's "Google on crack"
- Chat GPT 3.5 is currently free (<https://openai.com/chatgpt/pricing>), while GPT-4 is \$20/month, which includes DALL-E
- OpenAI's CEO, Sam Altman was fired then hired in November 2023 due disagreements with Board of Directors on future vision of Gen AI
- How big a deal is ChatGPT?
- It's certainly made a big splash. "ChatGPT is scary good. We are not far from dangerously strong AI," said Elon Musk, who was one of the founders of OpenAI before leaving. Altman tweeted that ChatGPT had more than 1 million users in its first five days after launching.
- According to analysis by Swiss bank UBS, ChatGPT is the fastest growing app of all time. In January, only two months after its launch, UBS analysis estimates that ChatGPT had 100 million active users. For comparison, it took nine months for TikTok to reach 100 million.

## How will AI tools such as ChatGPT impact cost estimation and acquisition?



# Poll Question

**How will AI tools such as ChatGPT impact cost estimation and acquisition?**

- a. No impact whatsoever**
- b. No impact for the next 5 years or so**
- c. Moderate Impact**
- d. Significant Impact**
- e. I'll lose my job!**

# IT Trends-ChatGPT-Application to Software Development

- Generative code development will change how software is developed
- ChatGPT supports over 50 programming languages
- Limitation is that it does not reside within IDE
- TuringBOTS
  - Term coined by Forrester describe generative AI-powered software such as Copilot or ChatGPT that helps developers code from the planning stage to the testing and deployment stage.
  - It defines what analysts believe is a major technology trend, giving birth to a new generation of software development.
  - Software vendors and federal Contracting Officers will have to manage generative code tools

*"ChatGPT is good at providing clear, natural-language suggestions or for getting started with and using a new programming language," said Diego Lo Giudice, principal analyst with Forrester Research. "But [developers] also use it to generate code from English prompts to document existing code, to get a clear explanation of a program or to translate from a programming language to another -- asking about APIs, libraries, parameters to use in coding and more."*

# Generative AI in 2024

- 2023 was the year ChatGPT exploded. 2024 will bring Gen AI to more specific use cases
- Technology
  - Google Bard operates as competitor to chat GPT at no cost, and integrates with Google tools
  - GitHub Co-Pilot: offers automated code development, generates code snippets and enhances developer productivity
- Graphics and art production
  - DALL-E: image generation from text prompts
  - Jasper and Writesonic: creates long-form content like blog articles, scripts, outlines, and more.
  - Fireflies: provides meeting note-taking and transcription
  - Adobe Sensei: Adobe's own generative AI models allow for seamless photo, audio, and video editing
  - Fliki: converts text into audio files and video to simplify the creative process behind videos, podcasts, or audiobooks in 1,000 voices and 75 languages
  - MusicLM by Google: currently in R&D, will allow to create music from text
- Business workflow
  - Summarize.tech: input a video link, and tool will generate text summary
  - Jasper and Writesonic: creates long-form content like blog articles, scripts, outlines, and more.
  - Fireflies: Provides meeting note-taking and transcription services
- AI Threats
  - Bad actors can gain access to user accounts
  - Fraudulent push payment requests from pay applications
  - Financial markets can be influenced by AI algorithms

# Application to Cost Estimation

## Will impact what is estimated:

- AI impact to private and public sector operating costs?
- How do estimators reflect these changes in estimates?
- What areas will be impacted?  
How soon?

## Will impact how we estimate:

- New ways of collecting data and performing cost research
- AI augmented estimation workflows
- AI powered insights and recommendations

# Potential Applications in Cost Estimation

**Streamlining cost research:** AI assisted cost research offers potentially more intuitive and interactive information gathering, leveraging immense datasets, including data sourcing

**Predictive modeling:** Algorithms can be trained on historical data to predict future costs for a project or product, taking into account factors such as materials, labor, and overhead

**Process automation:** AI can be used to further automate cost estimation processes, reducing the time and resources required to perform cost analysis

**Optimization:** AI algorithms can be used to optimize cost estimates, exploring different scenarios and trade-offs to find the most cost-effective solution

**Risk analysis:** AI can be used to assess the risk of cost overruns and identify the key factors that contribute to cost variability, allowing organizations to make more informed decisions

**Supply chain management:** AI can be used to analyze the cost of goods, services, and resources required to deliver a product or service, helping organizations to identify opportunities for cost savings.

# Use Cases

- **Use Case 1: Pattern Recognition**
  - Technology: PromptLoop function
  - Usage: Excel and Google Sheets
  - Reference: <https://www.promptloop.com/>
- **Use Case 2: Labor Rate Lookup**
  - Technology: PromptLoop function, GPT function
  - Usage: Excel, Google Sheets, Google Docs
  - Reference: <https://sheetgpt.ai/articles/how-to-start-using-sheetgpt-in-a-google-sheet>
- **Use Case 3: Functional Size Measurement**
  - Technology: NLP
  - Usage: Cadence
  - Reference: <https://logapps.com/cadence/>
- **Use Case 4: ChatGPT as SME**
  - Technology: Generative AI
  - Usage: OpenAI
  - Reference: <https://openai.com/>
- **Use Case 5: BOE Documentation, CES creation**
  - Technology: GPT chat
  - Usage: Excel, Google Sheets, Google Docs
  - Reference:
- **Use Case 6: Excel Data Cleanup**
  - Technology: GPT chat (v4 paid version)
  - Usage: Excel
  - Reference:

# Use Case 1: Using AI in a Spreadsheet: Pattern recognition

Part 1/2

**Goal:** Identify the correct color in column 'E' based on the fruit listed in column 'D'

- PROMPTLOOP() is a function that uses AI to determine what information should be displayed for a population of data based on some subset of data
- A user enters a sample set of pairings which the AI uses to determine the data structure and content
- The AI then attempts to determine the correct output (color in this example) based on the input alone

	A	B	C	D	E
1				apple	red
2				orange	orange
3				kiwi	green
4				plum	purple
5				raisin	white
6				nectarine	white
7				banana	yellow
8				grape	red
9					

AI generated

User entered

PromptLoop [Demo here](#)

[PromptLoop Spreadsheet Functions here](#)

# Use Case 1: Using AI in a Spreadsheet: Pattern recognition

## Part 2/2

### Observations and limitations:

- AI outputs are sometimes inaccurate or lacking context
  - Does AI think banana:brown because over-ripe?
- Outputs quality highly dependent on well constructed initial prompt inputs
- Lacking source information traceability

Untitled spreadsheet

File Edit View Insert Format Data Tools Extensions Help Last

100% \$ % .0 .00 123 Default (Ari... 10

`=PROMPTLOOP($D$1:$D$4,$E$1:$E$4,D5)`

A	B	C	D	E
			apple	red
			orange	orange
			kiwi	green
			plum	pruple
			raisin	brown
			nectarine	nectarine
			banana	brown
			grape	red



# Use Case 2: Using AI in a Spreadsheet: Labor Rate Lookup

Part 1/3

**Goal:** Find the labor rate based in column 'C' based on the labor category in column 'B'

- User creates a prompt by populating two columns:
  1. List labor categories
  2. Rates associated with those labor categories
- AI analyzes the user entered subset and generates rates for the missing labor categories

A	B	C
	Systems Engineer	\$130.00
	Systems Architect	\$145.00
	Computer Systems Analyst	\$135.00
	Customer Service Representative	\$88.00
	Database Administrator	\$95.00
	Information Security Analyst	120
	Computer Systems Administrator	135
	Software Developer	135
	Web Developer, Senior	170

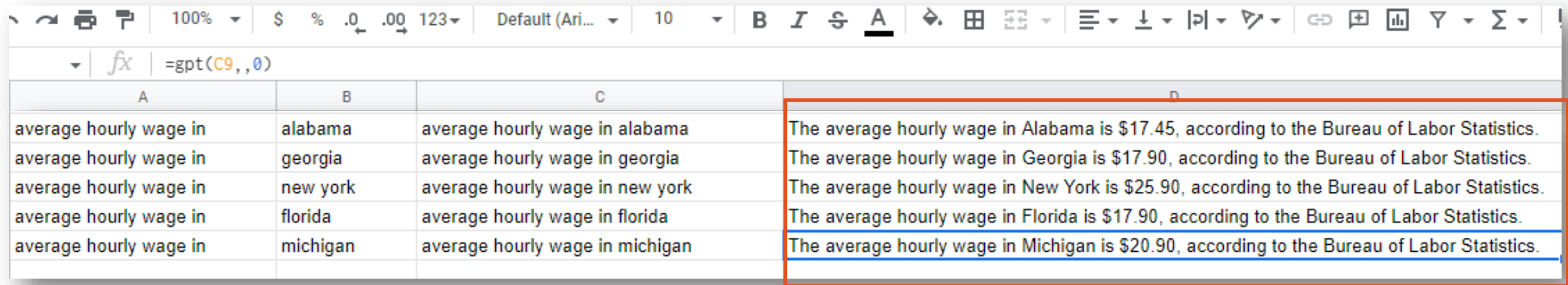
User entered

AI generated

# Use Case 2: Using AI in a Spreadsheet: Labor Rate Lookup

Part 2/3

**Modified Goal:** Lookup average hourly rates based on location



The screenshot shows a spreadsheet interface with a formula bar containing `=gpt(C9,,0)`. The spreadsheet has four columns: A, B, C, and D. Column A contains the text 'average hourly wage in', Column B contains state names, and Column C contains the text 'average hourly wage in' followed by the state name. Column D contains AI-generated text providing the average hourly wage for each state, along with the source 'Bureau of Labor Statistics'. The data is as follows:

A	B	C	D
average hourly wage in	alabama	average hourly wage in alabama	The average hourly wage in Alabama is \$17.45, according to the Bureau of Labor Statistics.
average hourly wage in	georgia	average hourly wage in georgia	The average hourly wage in Georgia is \$17.90, according to the Bureau of Labor Statistics.
average hourly wage in	new york	average hourly wage in new york	The average hourly wage in New York is \$25.90, according to the Bureau of Labor Statistics.
average hourly wage in	florida	average hourly wage in florida	The average hourly wage in Florida is \$17.90, according to the Bureau of Labor Statistics.
average hourly wage in	michigan	average hourly wage in michigan	The average hourly wage in Michigan is \$20.90, according to the Bureau of Labor Statistics.

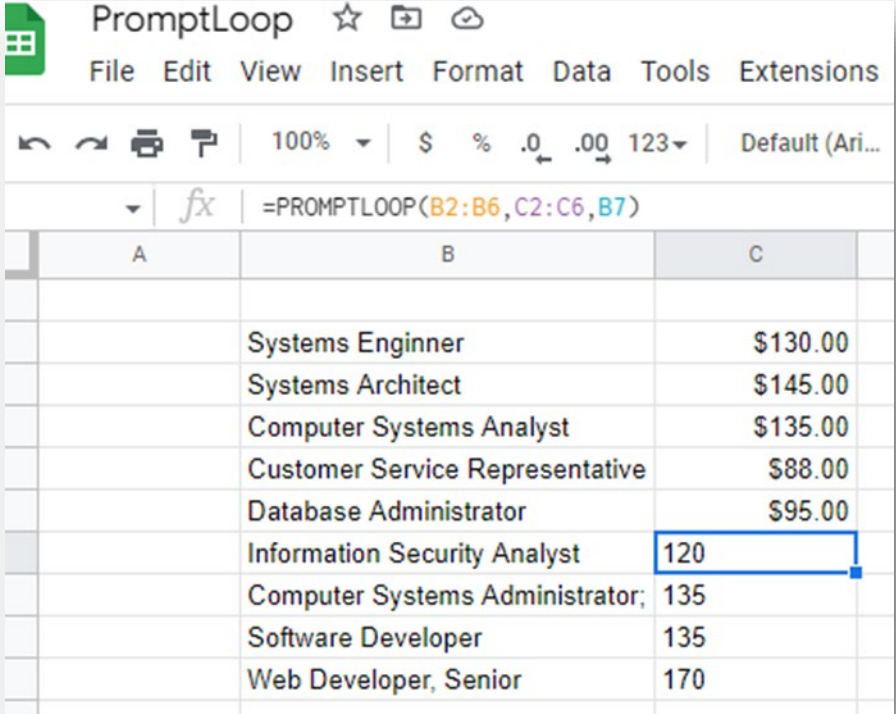
**Note that at this time the source information has been provided in the AI output and can be validated**

# Use Case 2: Using AI in a Spreadsheet: Labor Rate Lookup

Part 3/3

## Observations and limitations:

- Many unknowns for output rate
  - Location, experience?
  - Fully burdened (overhead, etc.)?
  - What timeframe?
- Validation
  - Compare a known set of rates to AI generated



A	B	C
	Systems Engineer	\$130.00
	Systems Architect	\$145.00
	Computer Systems Analyst	\$135.00
	Customer Service Representative	\$88.00
	Database Administrator	\$95.00
	Information Security Analyst	120
	Computer Systems Administrator;	135
	Software Developer	135
	Web Developer, Senior	170

# Use Case 3: Natural Language Processing (NLP)-Aided SW Estimating

Part 1/3

Example using CADENCE by LOGAPPS

Desired outcome in natural language

#3.2 As a Manager, I would like to create User Accounts for sponsored officials so users may access the system.



Req. #	Subject	Verb	Object (noun)	Conjunction phrase
3.2	I	would like to create	User Accounts	so users may access the system

Requirement Parsing

Req. #	Key Verb	Other Verb	Proper Noun	Noun	Subject	Object	Verb Phrase
3.2	create	like	User Accounts	Sponsored officials	Manager	User Accounts	Create user accounts

# Use Case 3: NLP-Aided SW Estimating

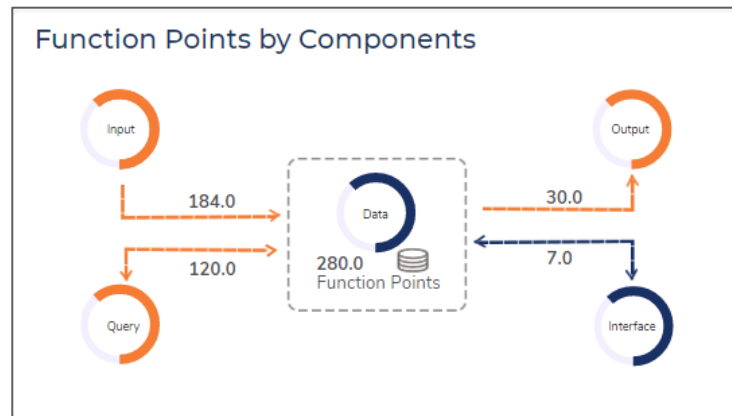
## Part 2/3

- Parsed requirements is used as input for functional sizing
- Specific words or phrases are mapped to IFPUG data or transactional functions
- Business rules are applied
- Analyst evaluates and adjusts as needed
- ROM-level software sizing produced, which then supports estimation or benchmarking activities
- **Functional sizing for large projects can be completed in minutes!**

Parse

Keyword	Add	Change	Delete	View	Report	Save	Interface	Actions
Accept	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Activate	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Add	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

Tag or Chunk



Match & Score

Requirement ID #5

Multiplier: 1

Total: 18.0

Requirement: The system shall modify the Scan Area Target Tables Using Simple Area Rate (Modify) Frame Area Target Table. Calculate Scanning Dwell Time

Comment: Add Comment Here

Save

# Use Case 3: NLP-Aided SW Estimating

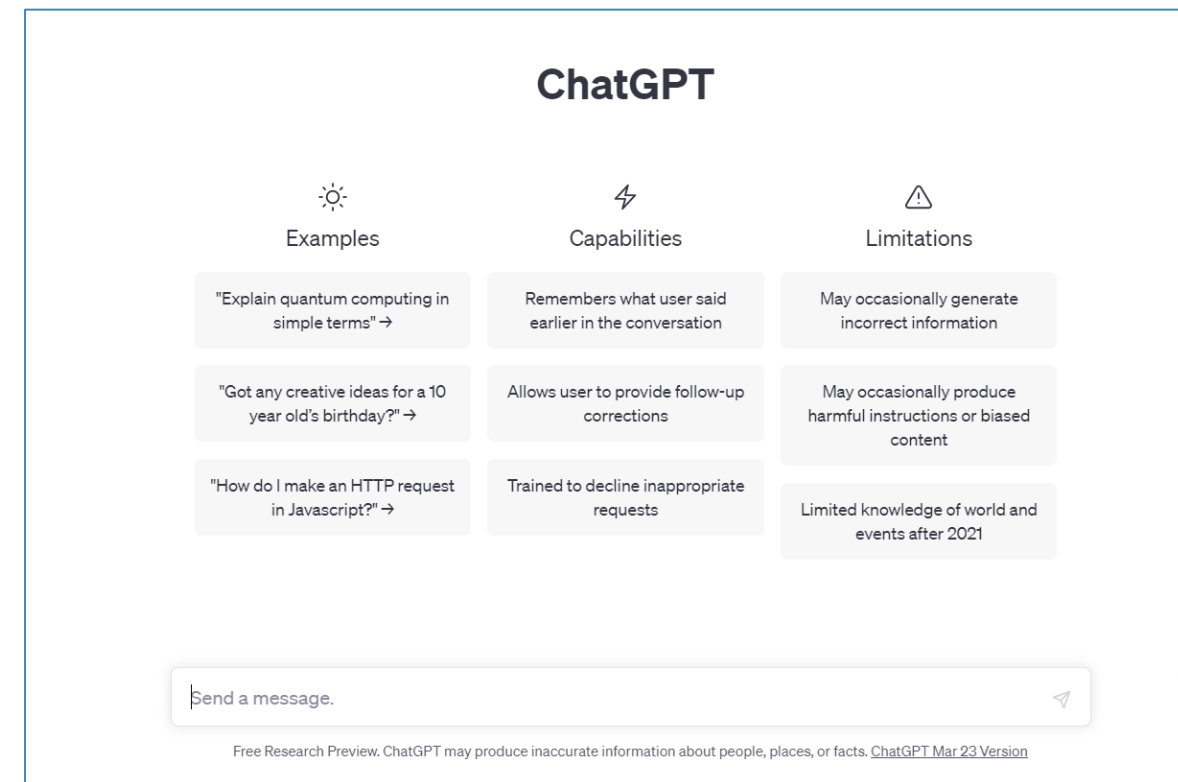
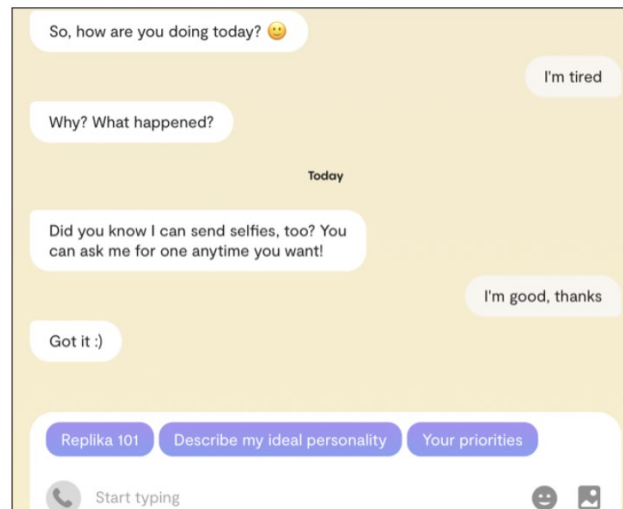
Part 3/3

## Observations and limitations:

- Requirements and User Stories are excellent subjects for NLP analysis because they have a semi-structured construct
- NLP greatly reduced the amount of time needed to associate requirements to function points
- Requirements need to be well-defined during the solicitation process

# Use Case 4: ChatGPT as SME

- **OpenAI has brought to market ChatGPT, GPT-4 and DALL-E2**
  - ChatGPT is a prototype AI chatbot released in November 22, available free of charge
  - GPT-4 is the next generation of Chat GPT as a subscription-based service and via API service (token pricing)
  - DALL-E 2 utilizes deep learning to generate images
- Numerous competitors have entered Generative AI market
  - Google Bard
  - GitHub Copilot
  - WriteSonic
  - Replica



# Use Case 4: ChatGPT as SME

what is the process to create a cost estimate to stand up a help desk?

The process to create a cost estimate to stand up a help desk involves several steps:

1. Define the scope of the project: Determine the size and type of help desk that is needed, including the number of staff required, the level of support to be provided, and the types of technology needed to support the help desk.
2. Determine the staffing needs: Estimate the number of staff required to run the help desk, including the number of agents, supervisors, and managers. This can be based on the expected volume of calls or requests, as well as the required level of service.
3. Estimate the technology costs: Identify the hardware and software needed to run the help desk, including computers, phones, ticketing systems, and other tools. Research the cost of purchasing or leasing the necessary equipment and software.
4. Estimate the training costs: Determine the cost of training the help desk staff, including initial training and ongoing training to ensure that they are able to provide the necessary level of support.
5. Estimate the facilities costs: Determine the cost of leasing or renovating a space to house the help desk, including rent, utilities, and other associated costs.
6. Create a detailed estimate: Based on the information gathered in the previous steps, create a detailed estimate of the costs for standing up the help desk, including staffing, technology, training, and facilities costs.
7. Review and refine the estimate: Review the estimate with the project team and make any necessary adjustments. It's important to take into account contingencies and any potential unknown costs that may arise during the implementation process.

It's important to note that cost estimates are not exact figures, they are a projection of the costs of the project, and actual costs may vary.

what is the hourly labor rate for a help desk worker?

The hourly labor rate for a help desk worker can vary depending on various factors such as location, level of experience, and industry.

According to data from the US Bureau of Labor Statistics, as of May 2020, the median hourly wage for computer user support specialists, which includes help desk workers, was \$25.16 per hour in the United States. The lowest 10% of earners made less than \$16.67 per hour, while the highest 10% earned more than \$42.06 per hour.

However, it's important to note that these figures are median wages and may vary based on different factors. Additionally, some companies may offer benefits or bonuses on top of the hourly rate, which can also impact the overall compensation package.

how many employees work for an average help desk support group:

The number of employees who work for an average help desk support group can vary widely depending on the size and complexity of the organization, the volume of calls or requests received, and the level of support required.

In general, smaller organizations may have only a few help desk support staff members, while larger organizations may have dozens or even hundreds of help desk support staff members. According to industry surveys and research, the average help desk support group may range from 3 to 20 staff members, but this number can vary widely depending on the organization's needs.

The number of employees required may also depend on the level of support provided. For example, a help desk that provides basic support for a few software applications may require fewer employees than a help desk that provides more advanced technical support for a wide range of applications and devices.

Ultimately, the number of employees needed for a help desk support group will depend on a variety of factors unique to each organization, and should be determined based on the specific needs and goals of the organization.

- **With just 3 chats, I was able to use ChatGPT as a “SME:”**
  - Received detailed process for estimating help desk costs
  - Low/average/high Hourly rate for Help Desk Staff
  - FTE counts



# Use Case 5: ChatGPT for IT CES/WBS

Create a work breakdown structure for software development and maintenance



Creating a complete work breakdown structure (WBS) for software development and maintenance can be quite extensive, but I can provide a high-level overview to get you started. Keep in mind that the WBS can vary depending on the project's complexity, methodology, and specific requirements. Here's a simplified example:

- Project Initiation**
  - Define project scope and objectives
  - Gather initial requirements
  - Identify stakeholders
- Planning**
  - Create a project plan
  - Develop a project schedule
  - Define roles and responsibilities
  - Risk assessment and mitigation planning
- Analysis and Design**
  - Requirements gathering and analysis
  - System architecture design
  - User interface design
  - Database design
- Development**
  - Coding and programming
  - Unit testing
  - Integration testing
  - Version control
- Quality Assurance**
  - System testing
  - User acceptance testing
  - Bug tracking and resolution
  - Performance testing
- Deployment**
  - Deployment planning
  - Installation and configuration
  - User training
  - Data migration
- Maintenance and Support**
  - Regular updates and patches
  - Bug fixes and troubleshooting
  - Performance monitoring
  - User support and helpdesk
- Documentation**
  - User manuals
  - Technical documentation
  - Knowledge transfer
- Project Management**
  - Project status reporting
  - Change management
  - Resource allocation
  - Budget tracking
- Closure**
  - Final acceptance and sign-off
  - Post-implementation review
  - Documentation archive
  - Lessons learned

- Created “Level 2” basic IT WBS
- Using the iterative nature of ChatGPT, asked for a Level 3 WBS
- <https://chat.openai.com/c/856d11ff-a801-4e40-ac56-97acf9bde659>

Presented at the ICEAA 2024 Professional Development & Training Workshop - [www.iceaaonline.com/min2024](http://www.iceaaonline.com/min2024)



Certainly! Breaking down the work breakdown structure (WBS) for software development and maintenance into the next level of detail:

- Project Initiation**
  - Define project scope and objectives
    - Conduct stakeholder interviews
    - Create a project charter
  - Gather initial requirements
    - Identify high-level functional requirements
    - Identify non-functional requirements (e.g., performance, security)
  - Identify stakeholders
    - Compile a list of project stakeholders
    - Define their roles and responsibilities
- Planning**
  - Create a project plan
    - Develop a project scope statement
    - Define project milestones
  - Develop a project schedule
    - Create a Gantt chart
    - Allocate resources and estimate durations
  - Define roles and responsibilities
    - Assign project roles (e.g., project manager, developer, tester)
    - Outline responsibilities and reporting structures
  - Risk assessment and mitigation planning
    - Identify project risks
    - Develop risk mitigation strategies
- Analysis and Design**
  - Requirements gathering and analysis
    - Conduct detailed user interviews and surveys
    - Create a comprehensive requirements document
  - System architecture design
    - Develop system architecture diagrams
    - Define system components and their interactions
  - User interface design
    - Create wireframes and mockups
    - Develop user interface prototypes
  - Database design
    - Design the database schema
    - Define data relationships and entities
- Development**
  - Coding and programming
    - Write code based on design specifications
    - Follow coding standards and best practices
  - Unit testing
    - Develop unit test cases
    - Execute unit tests on individual components
  - Integration testing
    - Integrate components and test interactions
    - Identify and address integration issues
  - Version control
    - Implement version control system (e.g., Git)
    - Manage code repositories and branches
- Quality Assurance**
  - System testing
    - Create comprehensive test plans
    - Execute system tests to validate system functionality
  - User acceptance testing
    - Involve end users to validate system usability

Regenerate

# Use Case 5: ChatGPT for IT CES/WBS

- Created “Level 3” basic WBS for IT



Create a work breakdown structure for information technology



Creating a work breakdown structure (WBS) for information technology (IT) can be broad and complex, as IT encompasses a wide range of activities and projects. Below is a simplified WBS for IT that you can further customize to fit the specific requirements and projects within your IT department or organization:

## 1. IT Strategy and Planning

- Develop IT strategic plan
- Define IT governance framework
- Conduct technology assessments
- Establish IT policies and standards

## 2. Infrastructure Management

- Network infrastructure
  - Design and maintain network architecture
  - Configure routers, switches, and firewalls
- Server infrastructure
  - Provision and maintain servers
  - Implement virtualization solutions
- Data center management
  - Manage data center facilities
  - Implement disaster recovery and backup solutions
- Cloud services
  - Evaluate and manage cloud providers
  - Migrate to cloud-based solutions

## 3. Software Development

- Requirements gathering and analysis
  - Collaborate with stakeholders to define software needs
  - Document functional and non-functional requirements
- Application design and development
  - Design software architecture
  - Code and develop applications
- Testing and quality assurance
  - Develop and execute test plans
  - Conduct user acceptance testing
- Deployment and release management
  - Deploy software to production environments
  - Manage software releases and updates

## 4. Cybersecurity

- Security assessment and planning
  - Identify security risks and vulnerabilities
  - Develop a security strategy and policies
- Network security
  - Implement firewalls, intrusion detection systems, and encryption
  - Monitor network for security breaches
- Data security
  - Protect sensitive data through encryption and access controls
  - Conduct regular security audits and assessments

## 5. Incident response and recovery

- Develop an incident response plan
- Investigate and mitigate security incidents

## 5. End-User Support

- Helpdesk and user support
  - Provide technical assistance to end-users
  - Resolve hardware and software issues
- Training and education
  - Develop training materials and programs
  - Educate users on best practices and IT policies
- Device management
  - Manage and maintain end-user devices (computers, mobile devices)
  - Install and update software and applications

## 6. Database Management

- Database design and modeling
  - Design and create database schemas
  - Define data relationships and entities
- Database administration
  - Ensure database performance and availability
  - Backup and recovery
- Data integration and reporting
  - Integrate data from various sources
  - Develop and maintain reporting solutions

## 7. Project Management

- Project initiation and planning
  - Define project objectives and scope
  - Develop project plans and schedules
- Project execution and monitoring
  - Manage project resources and tasks
  - Monitor project progress and budget
- Project closure
  - Ensure project deliverables are met
  - Conduct post-project evaluations and documentation

## 8. IT Procurement and Vendor Management

- Procurement planning
  - Identify hardware and software needs
  - Develop procurement strategies
- Vendor selection and management
  - Evaluate and select IT vendors
  - Manage vendor contracts and relationships
- Asset management
  - Track and manage IT assets (hardware, software licenses)

## 9. Business Continuity and Disaster Recovery

- Business impact analysis
  - Assess critical business processes and dependencies
  - Identify recovery time objectives (RTO) and recovery point objectives (RPO)
- Disaster recovery planning
  - Develop a disaster recovery plan
  - Test and update the plan regularly
- Backup and data preservation
  - Implement backup and data preservation strategies
  - Ensure data integrity and availability during disasters

## 10. IT Compliance and Governance

- Regulatory compliance
  - Ensure compliance with industry and government regulations
  - Conduct audits and assessments
- IT governance
  - Establish IT governance structures
  - Define roles and responsibilities for IT stakeholders

# Use Case 6: Clean Bad Data with ChatGPT for Excel

- Please clean up the data in this Excel workbook

Region	Reference	Date	Name	Valu
US West	219/4/jh	5/21/2021	DAN	5100
	88/DCNA/01	5/26/2021	bill	2368
	366/FAD/9	5/27/2022	fRed	1145
	1035/JLZ/3	3/12/2021	Joe	4561
	12/UAR/69	1/26/2021	jOE	2534
	26/rst/4	9/27/2022	Ted	69
AFRICOM	1902/faa/5	9/21/2021	juanita	3256
	83/FAP/6	1/2/2021	iSabelle'	1278
	711/MTX/8	4/27/2022	Lisa	1200
	15/MTX/0		LUCy	1439



You

Please clean up the data in this Excel workbook



ChatGPT

The workbook contains one sheet named 'Sheet1'. Let's load the data from this sheet and take a quick look at the first few rows to understand what kind of cleaning might be needed. [↔]

The data from 'Sheet1' has been loaded, but it seems there are a few issues that might need addressing:

1. **Column Headers:** The first row contains what appears to be the column headers, but they are not recognized as such.
2. **Missing Data:** There are missing values in the 'Region' column.
3. **Inconsistent Data:** The 'Name' column has names in different cases (e.g., 'DAN', 'bill', 'fRed').
4. **Data Types:** The 'Date' column should likely be in a date format, and 'Valu' seems to be a numerical column but may be read as text.

I will proceed with the following steps unless you have specific requests:

1. Set the first row as the header.
2. Fill missing values in the 'Region' column (need more information on how you want this handled).
3. Normalize the 'Name' column to have consistent casing (e.g., title case).
4. Ensure the 'Date' column is in the correct date format.
5. Convert 'Valu' to a numerical data type if it's not already.

# Use Case 6: Clean Bad Data with ChatGPT for Excel

- Please clean up the data in this Excel workbook

You  
proceed

Notice my misspelling

ChatGPT

The data has been cleaned with the following actions:

1. Set the first row as the column headers.
2. Converted the 'Date' column to datetime format.
3. Converted the 'Value' column to numeric.
4. Normalized the 'Name' column to title case (e.g., 'Dan', 'Bill').

However, there are still missing values in the 'Region' column. How would you like to handle these? Here are a few options:

1. **Fill with a specific value:** We can fill the missing values with a specific value like 'Unknown' or any other placeholder you prefer.
2. **Fill Forward/Backward:** If the data is sorted and the missing values should be the same as the nearest non-missing value, we can fill forward or backward.
3. **Remove Rows:** If the rows with missing 'Region' values are not useful, we can remove them.

Please let me know how you'd like to proceed with the missing values in the 'Region' column. [-]

missing values with 'unknown'

ChatGPT

The missing values in the 'Region' column have been filled with 'Unknown' and now appears to be cleaned and organized. If you need further help, please let me know. If you like to save this cleaned data back to an Excel file, please let me know.



You

please save this cleaned data back to an Excel file

ChatGPT

The cleaned data has been saved to a new Excel file. You can download it using the link below:

[Download the cleaned data file \[-\]](#)



Standardized dates

Standardized names (mostly)

Region	Reference	Date	Name	Value
US West	219/4/jh	2021-05-21 00:00:00	Dan	5100
Unknown	88/DCNA/01	2021-05-26 00:00:00	Bill	2368
Unknown	366/FAD/9	2022-05-27 00:00:00	Fred	1145
Unknown	1035/JLZ/3	2021-03-12 00:00:00	Joe	4561
Unknown	12/UAR/69	2021-01-26 00:00:00	Joe	2534
Unknown	26/rst/4	2022-09-27 00:00:00	Ted	69
AFRICOM	1902/faa/5	2021-09-21 00:00:00	Juanita	3256
Unknown	83/FAP/6	2021-01-02 00:00:00	Isabelle'	1278
Unknown	711/MTX/8	2022-04-27 00:00:00	Lisa	1200
Unknown	15/MTX/0		Lucy	1439

# Use Case 6: Clean Bad Data with ChatGPT for Excel

**You**  
change values to currency

**ChatGPT**  
To change the values in the 'Valu' column to a currency format, we can format the numbers as strings with a currency symbol. Excel typically handles currency formatting when opening the file, but we can add a currency symbol to the data as a text representation.

Would you like to use a specific currency symbol (e.g., \$, €, £) for the values in the 'Valu' column? If so, please specify the currency symbol.

**You**  
\$

**ChatGPT**  
The values in the 'Valu' column have been formatted with a dollar sign (\$). Now, I'll save this updated data back to an Excel file. [↗]

The data with the 'Valu' column formatted as currency has been saved to a new Excel file. You can download it using the link below:

[Download the currency formatted data file \[↗\]](#)

**You**  
change 'valu' column header to 'value'

**ChatGPT**  
The 'Valu' column header has been changed to 'Value'. The updated data has been saved to a new Excel file. You can download it using the link below:

[Download the updated data file \[↗\]](#)

- **Please clean up the data in this Excel workbook**

Region	Reference	Date	Name	Value
US West	219/4/jh	2021-05-21 00:00:00	Dan	\$5100
Unknown	88/DCNA/01	2021-05-26 00:00:00	Bill	\$2368
Unknown	366/FAD/9	2022-05-27 00:00:00	Fred	\$1145
Unknown	1035/JLZ/3	2021-03-12 00:00:00	Joe	\$4561
Unknown	12/UAR/69	2021-01-26 00:00:00	Joe	\$2534
Unknown	26/rst/4	2022-09-27 00:00:00	Ted	\$69
AFRICOM	1902/faa/5	2021-09-21 00:00:00	Juanita	\$3256
Unknown	83/FAP/6	2021-01-02 00:00:00	Isabelle	\$1278
Unknown	711/MTX/8	2022-04-27 00:00:00	Lisa	\$1200
Unknown	15/MTX/0		Lucy	\$1439

# Considerations and Caveats

- Generative AI models such as ChatGPT can produce convincingly erroneous results – must be validated
  - Even data sources quoted by AI can be fictitious!
- Successful AI implementations rely on well-crafted inputs and test data
- Some of the techniques and methods employed by AI models are “black box”, complicating traceability and auditing

# References

1. <https://www.statista.com/statistics/871513/worldwide-data-created/>
2. <https://www.geeksforgeeks.org/introduction-convolution-neural-network/>
3. [What is the difference between a convolutional neural network and a regular neural network? - Artificial Intelligence Stack Exchange](#)

# Use Case 7: Analysis of Financial Data

You

transactions (58).csv  
Spreadsheet

please analyze my data

ChatGPT

The dataset seems to represent transaction records, possibly from a personal or business account. Here's a breakdown of the columns present in the data:

- Date:** The date of the transaction.
- Merchant:** The entity with which the transaction was made.
- Category:** The category of the transaction, for example, income, gas & fuel, unknown, etc.
- Account:** The account from which the transaction was made. It seems to mask sensitive parts of the account number.
- Original Statement:** The original statement description of the transaction.
- Notes:** Additional notes about the transaction, if any.
- Amount:** The amount of the transaction. Negative values probably indicate expenses, while positive values indicate income.
- Tags:** Additional tags or labels for the transaction, if any.

What specific analysis or insights are you looking for with this data? Some common analyses could include:

- Summary Statistics:** Overview of total income vs. expenses, average transaction amount, etc.
- Categorization Analysis:** Analysis of spending or income by category.
- Temporal Analysis:** How do income and expenses trend over time? Are there any noticeable patterns?
- Merchant Analysis:** Which merchants are most common, or which ones involve the highest amounts of spending or income?

- Vacation food:** 16 transactions, total of -\$534.03
- Veterinary:** 1 transaction, total of -\$109.50
- Water:** 2 transactions, total of -\$181.50
- kids activities:** 1 transaction, total of \$53.47
- medical bill:** 10 transactions, total of -\$3511.58

### Temporal Analysis:

(Note: Displaying the aggregated amount of transactions monthly, with count, sum, average, minimum, and maximum.)

- Transactions are spread from November 2021 to January 2024.
- There's a significant fluctuation in the total amounts, with some months showing large positive sums and others showing large negative sums.
- Latest Month (January 2024):** 328 transactions, total sum of \$88,505.52, average of \$269.83 per transaction.

### Merchant Analysis:

(Note: Only a part of the merchants is displayed for brevity. Each merchant provides a count of transactions, total sum, average transaction amount, minimum, and maximum transaction amount with that merchant.)

- 3 Dogs and a Chick:** 1 transaction, total of -\$29.96
- 3rd Planet:** 1 transaction, total of -\$6.35
- 3rd Planet Brewing:** 20 transactions, total of -\$483.73
- 527 HOLDINGS MIARMAR BEACH:** 1 transaction, total of -\$10.92
- 7-Eleven:** 5 transactions, total of -\$14.04

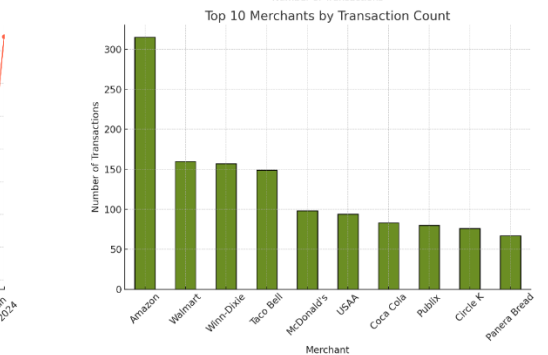
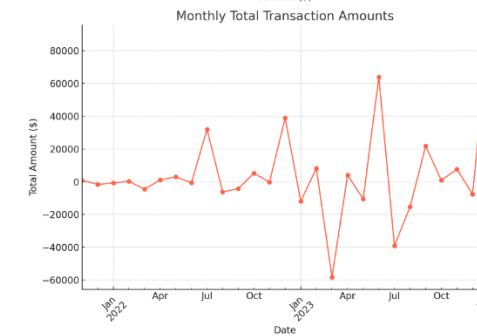
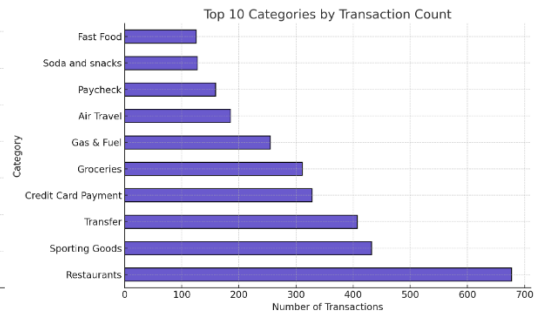
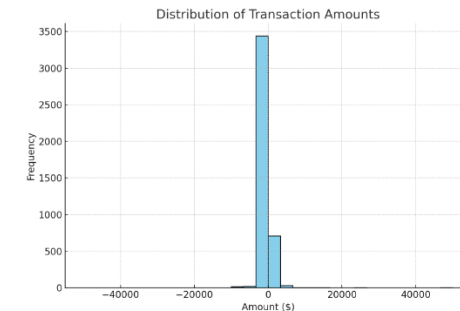


You

create some charts showing this analysis



ChatGPT



The charts visualize various aspects of your transaction data:



# Use Case 8: Analysis of GSA Calc Labor Rates

- Asdfliasd;fl
  - ChatGPT is a prototype AI chatbot afa

**GSA** BUY.GSA.GOV | Pricing

Research ▾ Develop documents ▾ Buy ▾ Help ▾

Home / Pricing Intelligence Suite / Labor Category Ceiling Rates

### Search labor category ceiling rates applicable to GSA's Multiple Award Schedules

This tool allows you to review fully burdened Not-to-Exceed Ceiling Rates awarded under GSA Multiple Award Schedule (MAS) contracts. You can use it to conduct market research and develop Independent Government Cost Estimates.

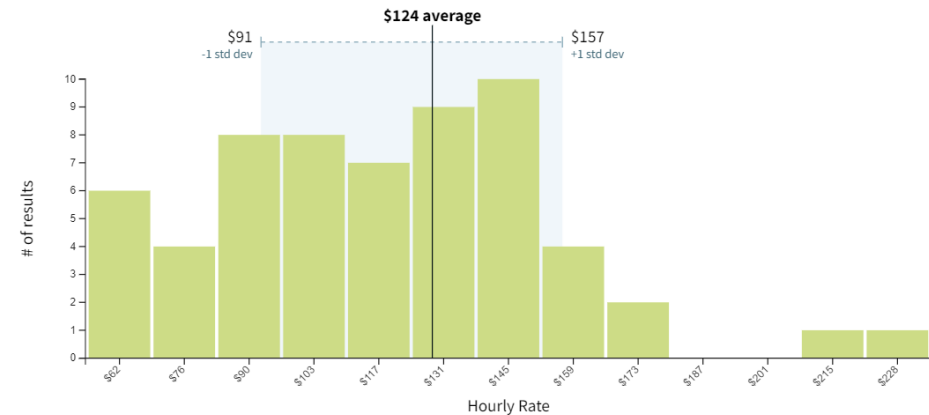
The tool can be used to assess the relative competitiveness of a vendor's price to other vendors' prices on MAS contracts. However, the tool is to be used only as part of a larger negotiation objective development strategy that seeks fair and reasonable pricing (in accordance with FAR 15.4).

[Labor Ceiling Rates User Guide](#) PDF

Search by:

Labor Category ▾ cost estimator

Not-To-Exceed Hourly Ceiling Rate



Standard deviation is a commonly used measurement that, in this case, tells you how far each labor category price is from the average (or arithmetic mean). It should not be used as the basis for determining fair and reasonable pricing. The Labor Ceiling Rate histogram displays the average labor rate, the price point that includes 34% of prices higher than the mean (in a normal distribution), the price point that includes 34% of prices lower than average (in a normal distribution), and the number of results per price range.

Std Deviation -1	Average Price	Std Deviation +1
<b>\$91.00</b>	<b>\$124.00</b>	<b>\$157.00</b>

# Using ChatGPT in a Spreadsheet

## How to use ChatGPT Functions in a spreadsheet: Initial Setup

Show users how to get set up with ChatGPT, locating their unique API token, and installing on their system for utilizing ChatGPT

<https://chat.openai.com/chat>

<https://www.youtube.com/watch?v=zlhIzvQR6w>

 Examples	 Capabilities	 Limitations
"Explain quantum computing in simple terms" →	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?" →	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?" →	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

# AI 101: Major Developments

Modern AI Developments	AI Paradigm	Foundational AI				Insightful AI			Generative AI	
	Decade	1940's	1950's	1960's	1970's	1980's	1990's	2000's	2010's	2020's
Robotics		Early work				Practical applications				
Neural Networks		Early work				Revival				
Natural Language Processing (NLP)			Early work			Practical applications				
Computer Vision				Early work					Breakthrough	
Expert Systems						Early work				
Machine Learning (ML)						Early work				
Reinforcement Learning (RL)						Early work				
Big Data							Introduced		Mass adoption	
Deep Learning								Introduced		Breakthrough
Generative Models (GANs, etc.)									Introduced	

# AI 101: Machine Learning

Definition: algorithms that can learn from and make predictions or decisions based on data

Recent developments:

1. **Generative Adversarial Networks (GANs):** GANs are a type of deep learning algorithm that can generate new data, such as images or music, that is similar to a training dataset.
2. **Reinforcement Learning:** Reinforcement learning is a type of machine learning that involves training models to make decisions in an environment by receiving rewards or punishments.
3. **Transfer Learning:** Transfer learning is a technique that allows models trained on one task to be used on another related task, reducing the amount of training data required.
4. **Attention Mechanisms:** Attention mechanisms are a type of deep learning architecture that allows models to focus on important parts of the input data, making them more efficient and effective.
5. **Natural Language Processing (NLP):** NLP is a branch of machine learning focused on processing and understanding human language. Recent advances in NLP include the use of deep learning for tasks such as sentiment analysis and machine translation.
6. **Robotics:** Machine learning algorithms are being used to control robots, allowing them to learn from experience and perform tasks such as grasping objects and navigating environments.

# Abstract

AI and Cost Estimation: Data Science's Expanding Role in Cost Estimating

Technical Innovations Track (TI03)

Kevin McKeel

Daniel Harper

Greg Forman

Data Science has taken on an expanded presence in Cost Analysis. E.g., NLP is used to automate functional software sizing in commercial models. Data Science methods and tools such as NLP can be also used to extract data to identify spending trends and agency funding. We will present an overview of modern usages of data science, to include Machine Learning, AI and data visualization. We will also survey cost and budget data sources which are publicly available and ripe for analysis using data science methods.

Keywords: Artificial Intelligence, Data Science, Machine Learning, NL