

AI and Cost Estimation

Data Science's Expanding Role in Cost Estimating

Dan Harper

Kevin McKeel

Greg Forman



Overview



Purpose &
Background



IT Trends



Application to
Cost Estimation



AI Use Cases
and Examples



Considerations
and Caveats

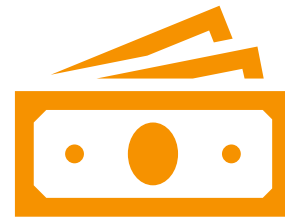


Resources

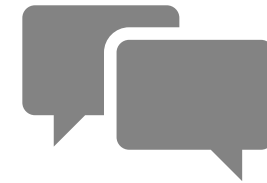
Purpose



Provide basic overview of key AI concepts and developments



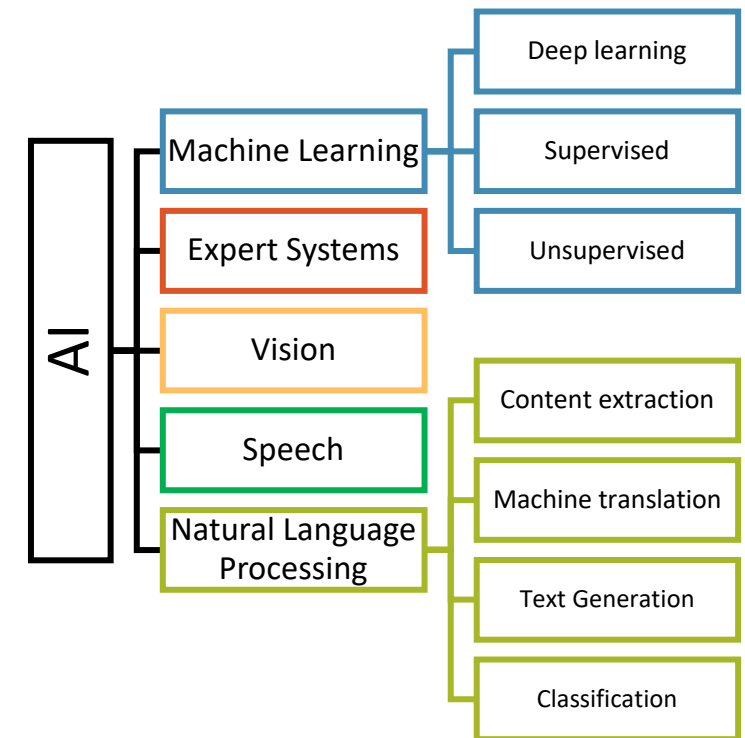
Introduce Applications of AI relevant to Cost Estimation



Initiate further study and discussion

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 exponentially¹
- 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"
- Usage is currently open to public free of charge because ChatGPT is in its research and feedback-collection phase. As of Feb. 1, there is also a paid subscription version called ChatGPT Plus.
- ChatGPT was created by OpenAI, an AI and research company [OpenAI launched ChatGPT on Nov. 30, 2022]
- 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. Sam Altman, OpenAI's chief, said on Twitter 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?

Application to Cost Estimation

Will impact what is estimated:

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 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

<https://www.youtube.com/watch?v=C-WKlchYZKA>

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(D1:D4,E1:E4,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

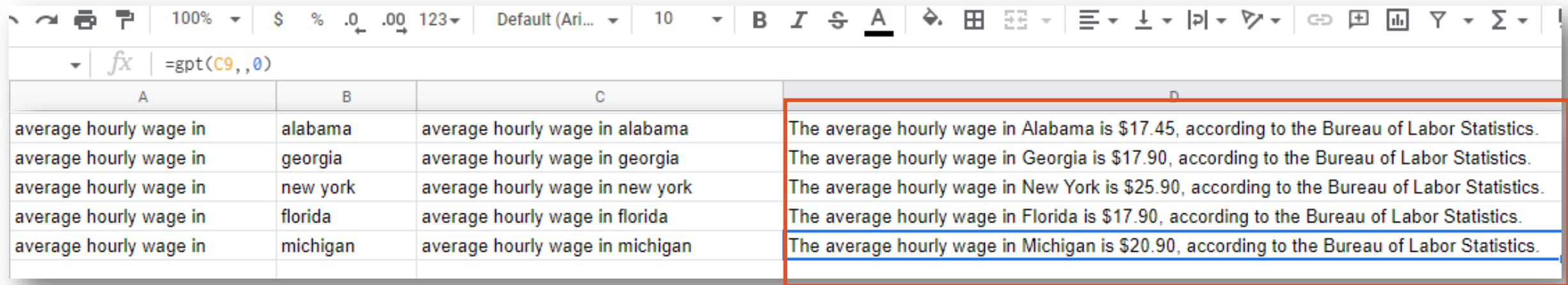
User entered

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: "The average hourly wage in [State] is \$[Rate], according to the 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.

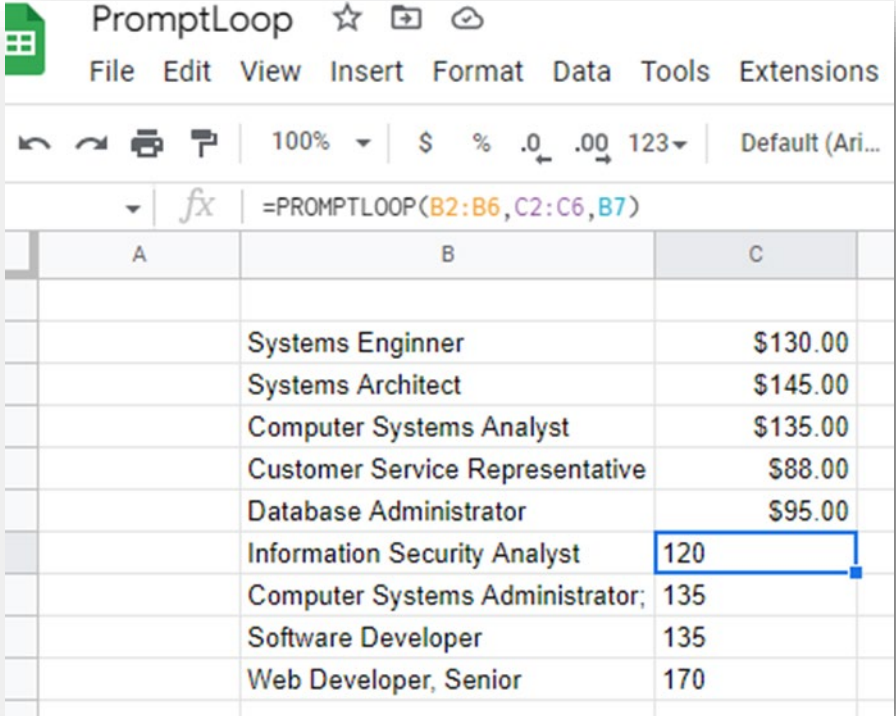
Note that at this time the source information has been provided in the AI output and can be validated

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

Settings

Whitelist Keywords

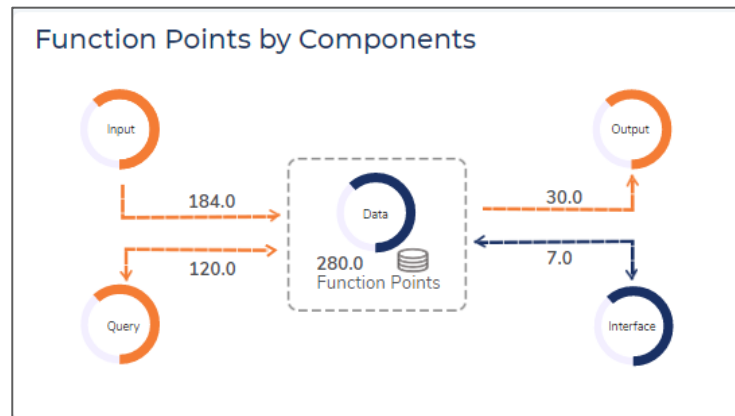
Programming Language

Function Point Weights

Keyword Settings

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

Input: 4.0, Query: 4.0, Data: 10.0, Total: 18.0

Multiplier: 1

Comment: Add Comment Here

Save

Unique Data Groups: Deposit Account

Logical group of data is maintained by the application

Logical group of external data is maintained by another application

Transactional Data: Make Deposit, View Deposit

Input transaction adds data to a maintained data group

Input transaction changes data in a maintained data group

Input transaction deletes data in a maintained data group

Inquiry transaction presenting retrieved data to user

Use Case 3: NLP-Aided SW Estimating

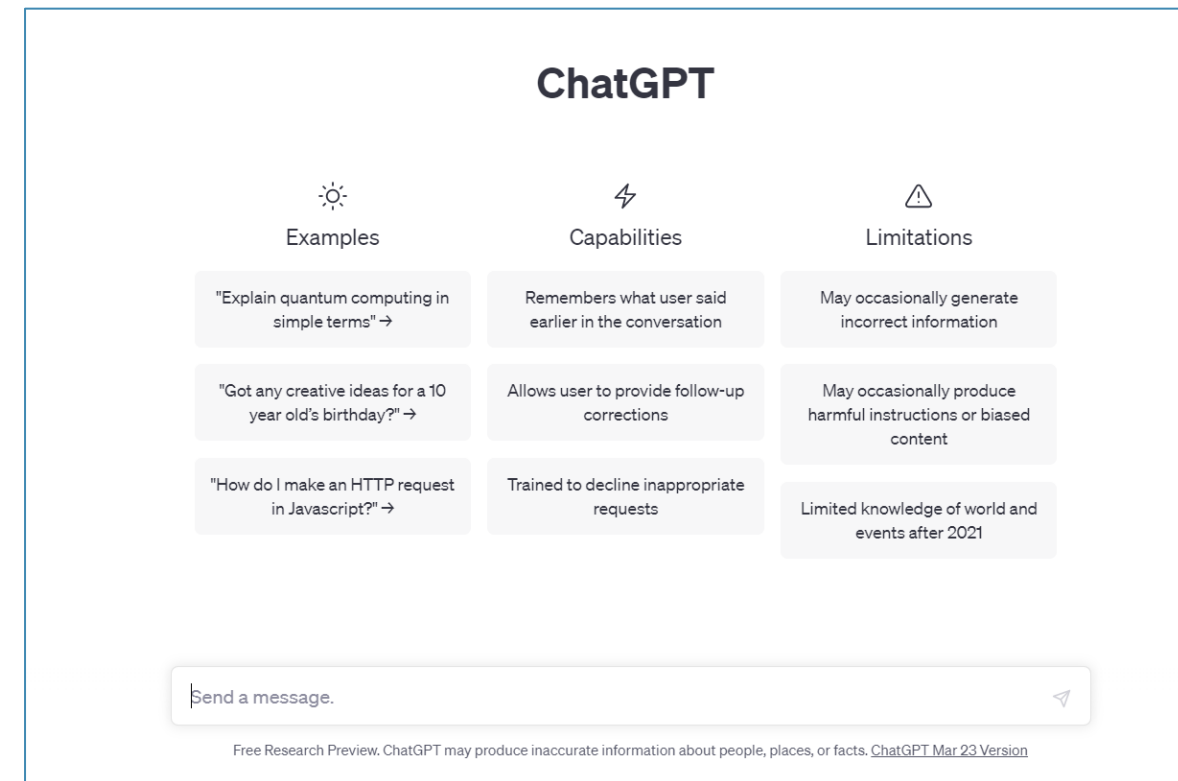
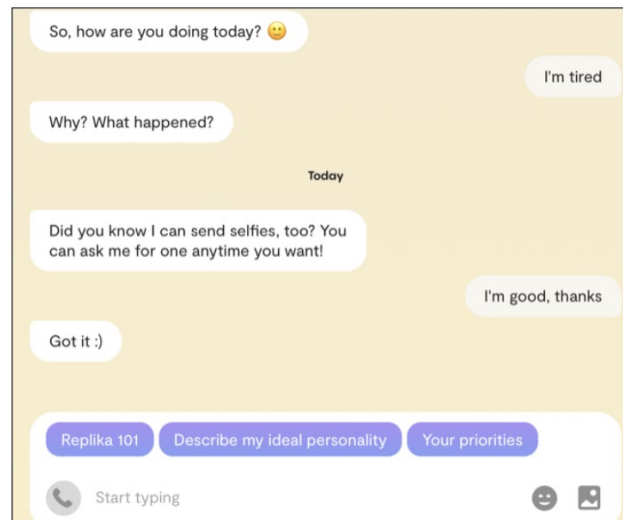
Part 3/3

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

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](#)

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