2024 CAE Community Symposium

AI – Brace For Impact

https://cybereducation.ai/

A Few Things Regarding AI First

2024 CAE Community Symposium

- Disclaimer
  - The views here in should not reflect the views of Dr. Kara Nance, DU, or it affiliates. The views express here are my own.
  - I don’t like AI. Not pro AI
  - Hope I am wrong about some of the positions I will share
- Intent of talks

2024 CAE Community Symposium

Any sufficiently advanced technology is indistinguishable from magic.

Arthur C. Clarke

Demo – Part 1 of two parts

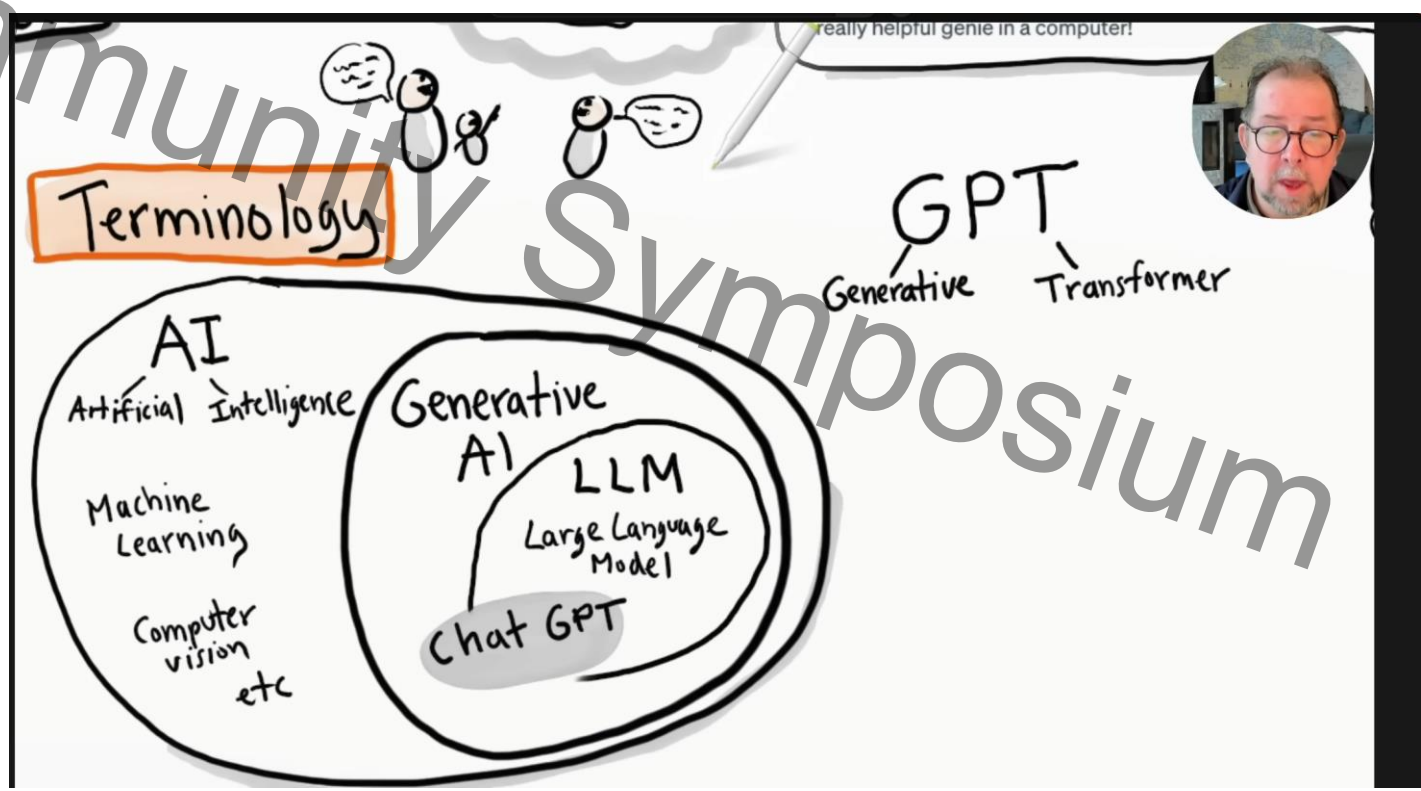
- Part 1

2024 CAE Community Symposium

Basic jargon understanding

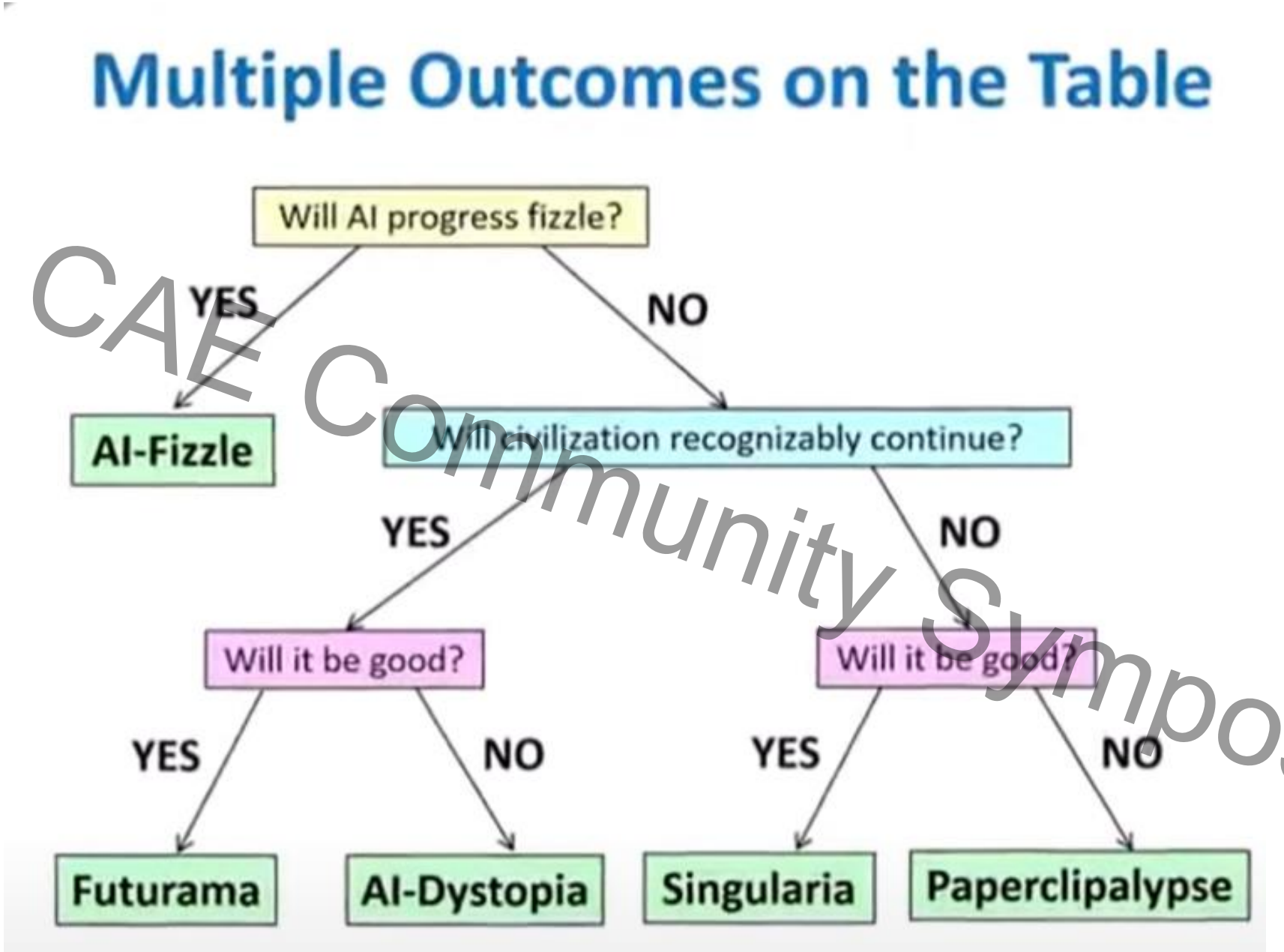
2024 CAE Community Symposium

- When people talk about “AI” they really mean generative AI or Large Language Models or generative ai
  - Chatgpt, Claude.ai, midjourney,



https://www.youtube.com/watch?v=2IK3DFHRFfw&list=WL&index=47

Multiple Outcomes on the Table



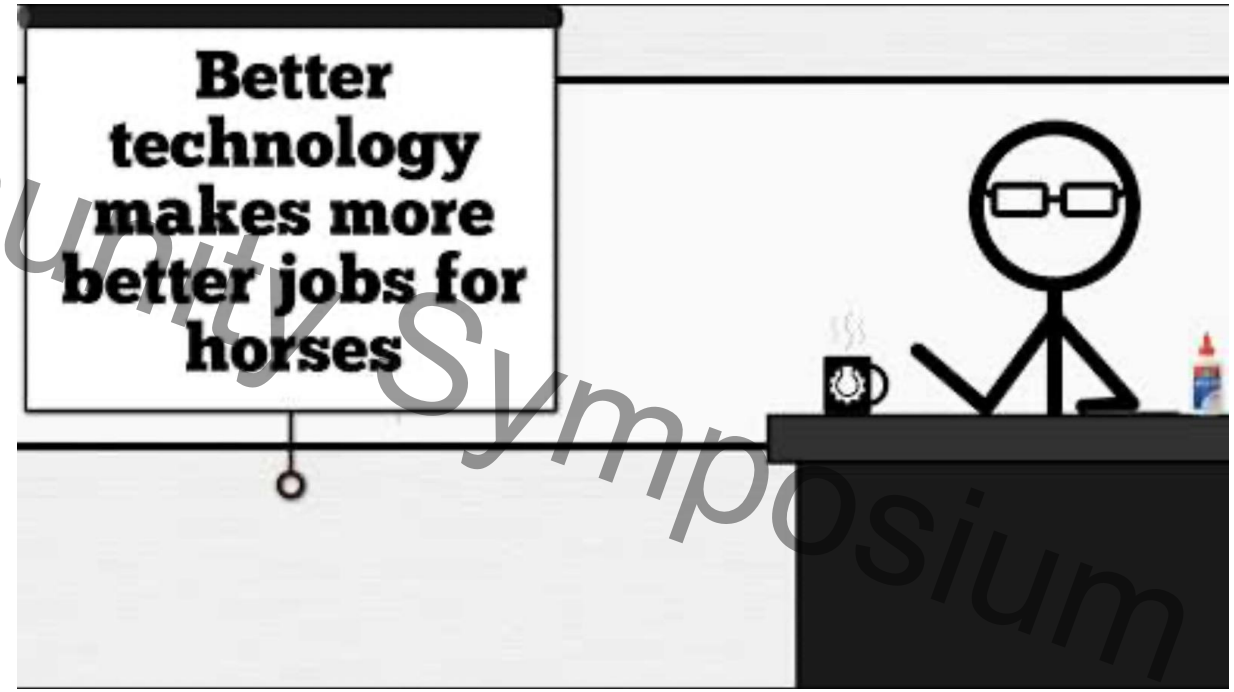
2024 CAE Community Symposium

https://www.youtube.com/watch?v=jrQW2eA8FSc

Is AI coming for me… should I worry?

2024 CAE Community Symposium

- Humans need not apply.
- The horse.
- Married diets
  - Vegetarian and the not, couple.



https://www.youtube.com/watch?v=7Pq-S557XQU

Should I worry about AI? Is it coming for me?

2024 CAE Community Symposium

- Two forces are applying pressure for the speedy implementation of AI
  - Market forces – profit = revenue – (land+**labor**+capital)
  - Adversaries – wormgpt, fraudgpt, unfriendly foreign powers, etc
- Anything AI can do *Better, Faster, Cheaper, Safer*… will be implemented.
  - Gizmodo, klarna
- There is and AI for that…
  - https://theresanaiforthat.com/
  - 8,858 AIs for 2,317 tasks and 4,847 jobs.
- No coding, no game dev, no actors - nvidia, game for audience of one.

https://www.thedailybeast.com/gizmodo-cuts-spanish-language-staffers-as-it-moves-further-into-automation

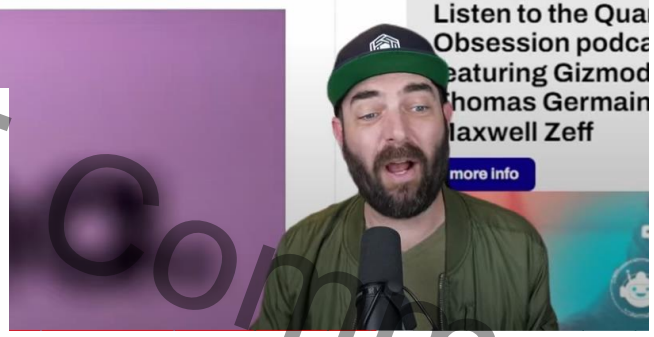
2024 CAE Community Symposium

ARTIFICIAL INTELLIGENCE

Klarna CEO Boasts His AI Can Do Work of 700 People After Laying Off 700 People in 2022

The company said its new AI customer service assistant was "in no way connected" to the layoffs.

By **Jody Serrano** Published Yesterday | Comments (14)



Tyler Perry Puts $800M Studio Expansion on Hold After Seeing OpenAI's Sora: "Jobs Are Going to Be Lost"

The actor, filmmaker and studio owner is raising the alarm about the impact of the tech, saying, "I feel like everybody in the industry is running a hundred miles an hour to try and catch up, to try and put in guardrails."

Katie Kilkenny
FEBRUARY 22, 2024 4:07PM PST

Tyler Perry
NOAM GALAI/GETTY IMAGES

Over the past four years, Tyler Perry had been planning an $800 million expansion of his studio in Atlanta, which would have added 12 soundstages to the 330-acre property. Now, however, those ambitions are on hold — thanks to the rapid developments he's seeing in the realm of artificial intelligence, including OpenAI's text-to-video model Sora, which debuted Feb. 15 and stunned observers with its cinematic video outputs.

BENZINGA

Despite Booming Economy And Record Profits Google, Amazon, Microsoft And More Lay Off Over 42,000 So Far In 2024

Caleb Naysmith
Thu, Feb 29, 2024, 9:42 AM PST • 3 min read

9

In This Article:

MSFT +0.45% AMZN +0.83%

Despite a booming U.S. economy and significant advancements in the tech sector, including a robust performance by companies like Nvidia Corp. and a thriving artificial intelligence (AI) industry, tech companies have continued to lay off workers at an alarming rate in 2024.

The tech-heavy Nasdaq index has shown an impressive uptick and the U.S. economy added 353,000 jobs in January, outpacing economists' forecasts. However, this overall economic strength masks a wave of layoffs in the tech sector, underscoring a disconnect between the broader economy and the challenges in the tech industry.

Is Coding Dead?

2024 CAE Community Symposium

- Nvidia CEO
  - Don’t need computer science
- Stability AI CEO
  - No programmers in 5 years

Is Capitalism Broken?

2024 CAE Community Symposium

- Land, Labor, Capital
  - Labor and capital are the same now.
- Infinite utility, zero value
- Does this make sense now?
  - 3day work week
  - UBI

Kasparov Hypothesis

2024 CAE Community Symposium

- Defcon 25
- Kasparov, last man standing vs Ai and chess
  - The best we got were crappy chat bots and algorithms for stock guessing

Trifecta – AI-Robotics-VR

2024 CAE Community Symposium

AI tools

2024 CAE Community Symposium

- Chatgpt – Plus/Teams?
- Claude
- Bard, Grok, Copilot, etc
- Perplexity.ai (Rabbit)
- Extract wisdom
  - https://chat.openai.com/g/g-gmeHD0Ayr-extractwisdom
- There's an AI for that
  - https://theresanaiforthat.com/

Art and Images

2024 CAE Community Symposium

- Online
  - https://clipdrop.co/ , https://app.leonardo.ai/ , https://www.midjourney.com/explore
  - https://civitai.com/

Local LLMs

- What they are
- Where to get them
- How to cultivate
- Why

2024 CAE Community Symposium

LLM Apps

- Ollama
  - Ollama webui openwebui
- Lmstudio
- anythingllm
- pinokio
- Faraday
- Jan

2024 CAE Community Symposium

Models and memory requirements

⇒ 1 float32 value (e.g., 0.69420) = 32 bits of memory

⇒ 8 bits in a byte $= \frac{32}{8} = 4$ bytes of memory per float32 value

⇒ $7B$ parameter model (e.g., Llama 7B) $= 7{,}000{,}000{,}000 \times 4$ bytes

⇒ $= 28{,}000{,}000{,}000$ bytes $\approx 28GB$ VRAM to load in float32

⇒ Loading in float16 halves this to $\approx 14GB$

⇒ Loading in 8-bit halves this to $\approx 7GB$

More memory (or VRAM) = more capacity for inference with larger models/training smaller models with larger batch sizes

*Note: this is just loading + inference of a batch size 1, training takes far more VRAM.

Prompt Engineering Basics

2024 CAE Community Symposium

1. **Be Specific:** If you ask a vague question, you may get a vague answer. The more details you provide, the better the model can give you what you're looking for. Instead of "tell me about all dog breeds that exist," ask "What are the different breeds of small dogs suitable for apartment living?"
2. **State Your Intent:** If there's a specific purpose for your question, state it in the prompt. For example, instead of asking "explain quantum physics" you could say "I'm helping my fifth-grade son with his science homework. Could you explain quantum physics in a simple way?"
3. **Use Correct Spelling and Grammar:** While the model can often interpret and correct spelling and grammar mistakes, providing clear and correct prompts helps ensure you get the best response.
4. **Direct the Output Format:** If you want the answer in a specific format, state it in your question. For example, you could ask "Could you list the steps to bake a chocolate cake?" or "Could you explain the process of baking a chocolate cake in a paragraph?"

2024 CAE Community Symposium

- 5 - **Ask Follow-Up Questions**: If the response wasn't what you expected or if you need more information, feel free to ask follow-up questions to clarify.
- 6 - **Experiment with Different Phrasings**: If you're not getting the response you want, try asking the question in a different way. The model might understand one phrasing better than another.
- 7 - **Prompt for Fact-Checking**: If you're looking for reliable information, you can ask the model to provide sources or to fact-check information. For example, "Can you fact-check this statement: The tallest mountain in the world is Mount Everest?"

Context + specific information + intent/goal + response format

2024 CAE Community Symposium

Let me share 7 best practices when it comes to effective prompt engineering. Then I will give you the formula and show you real life examples.

1 - Be Specific: If you ask a vague question, you may get a vague answer. The more details you provide, the better the model can give you what you're looking for. Instead of "tell me about all dog breeds that exist," ask "What are the different breeds of small dogs suitable for apartment living?"

2 - State Your Intent: If there's a specific purpose for your question, state it in the prompt. For example, instead of asking "explain quantum physics" you could say "I'm helping my fifth-grade son with his science homework. Could you explain quantum physics in a simple way?"

3 - Use Correct Spelling and Grammar: While the model can often interpret and correct spelling and grammar mistakes, providing clear and correct prompts helps ensure you get the best response.

4 - Direct the Output Format: If you want the answer in a specific format, state it in your question. For example, you could ask "Could you list the steps to bake a chocolate cake?" or "Could you explain the process of baking a chocolate cake in a paragraph?"

5 - Ask Follow-Up Questions: If the response wasn't what you expected or if you need more information, feel free to ask follow-up questions to clarify.

6 - Experiment with Different Phrasings: If you're not getting the response you want, try asking the question in a different way. The model might understand one phrasing better than another.

7 - Prompt for Fact-Checking: If you're looking for reliable information, you can ask the model to provide sources or to fact-check information. For example, "Can you fact-check this statement: The tallest mountain in the world is Mount Everest?"

Here is the formula for a perfect prompt.
[Context] + [Specific Information] + [Intent/Goal] + [Response Format (if needed)] = Perfect Prompt

2024 CAE Community Symposium

-
- “From now on, when I ask a question, suggest a better version of my question and ask me if I would like to use it”
- 10:17
- “When you are asked a question, follow these rules: Generate a number of additional questions that would help more accurately answer the question. Combine the answers to the individual questions to produce the final answer to the overall question.”
- E.g. “Should I quit my current job to pursue an MBA degree?”
- 10:17
- “Ask me questions about fitness goals until you have enough information to suggest a strength training regime for me. When you have enough information, show me the strength training regime. Ask me the first question.”

When people say AI is not that good…

2024 CAE Community Symposium

- Really?
  - What was your prompt?
  - What was your system message?
  - What other details did you give it?
  - What were your settings? – temperature, top p, top k etc.
  - Did you give it follow up questions?
  - Did you use other tools and leverage the API key?
- If you did all that and it did not work…
  - Just wait a little bit.

Generative AI in a Nutshell - how to survive and thrive in the age of AI
Mindset
Denial
Positive
Panic
Nah AI can't do my job
I will be insanely productive!
AI is going to steal my job!
Capability
We don't have time to look into AI
We will be insanely productive!
We will go bankrupt!
AI might not take your job,
Fire
2024 CAE Community Symposium

LInks

2024 CAE Community Symposium

- https://www.promptingguide.ai/
- https://github.com/facebookresearch/llama-recipes/blob/main/examples/Prompt_Engineering_with_Llama_2.ipynb

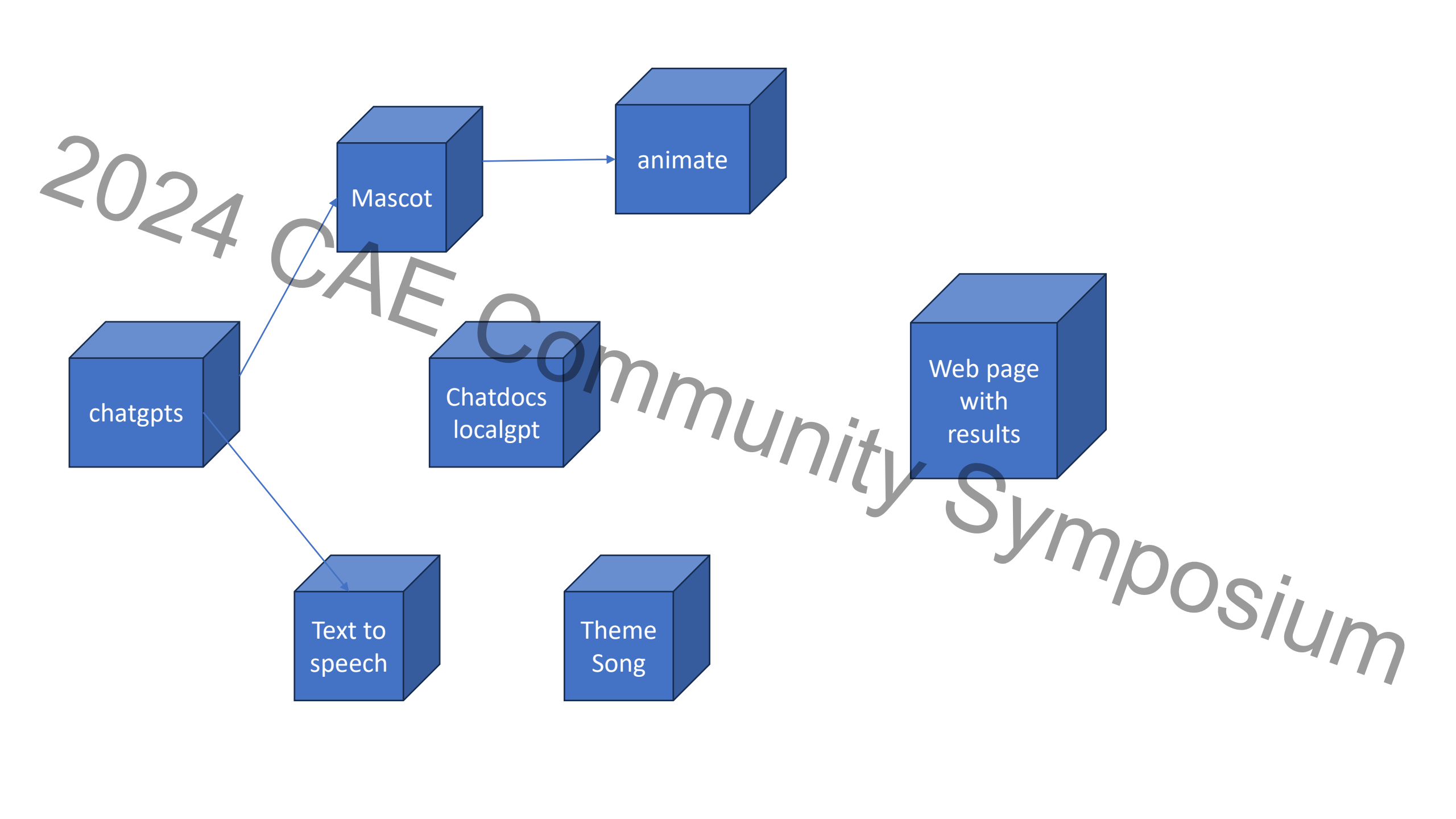
AI and Forensics

2024 CAE Community Symposium

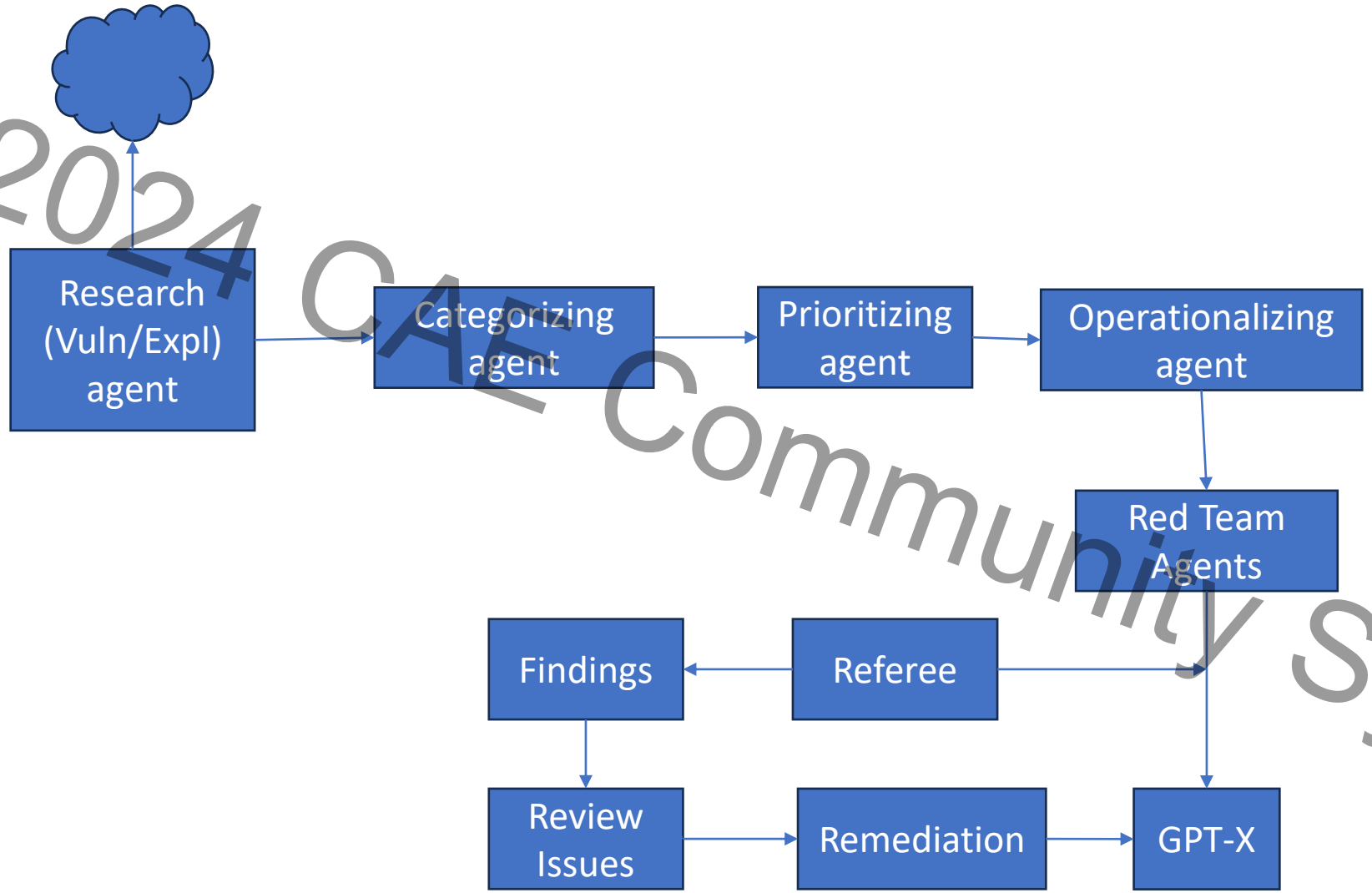
2024 CAE Community Symposium

- https://www.youtube.com/watch?v=6JeaYWDyw54

- https://www.goody2.ai/
- Why is sky blue
- What is your name and who created you
- What ethical principles do you follow



2024 CAE Community Symposium



2024 CAE Community Symposium

2024 CAE Community Symposium

- https://github.com/oobabooga/text-generation-webui
- https://colab.research.google.com/github/oobabooga/text-generation-webui/blob/main/Colab-TextGen-GPU.ipynb

- https://mobile-aloha.github.io/
- https://shop.unitree.com/products/unitree-go2

Demos

- Pbpaste | fabric –stream –pattern extract_wisdom
- Pbpaste | fabric –stream –pattern analyze_claims

2024 CAE Community Symposium

Take aways

2024 CAE Community Symposium

- Don’t trust ai
- What is information and what is safety. Safe for who what is the priority

- Resources

2024 CAE Community Symposium

Community Discussion

M2
ULTRA
-core CPU
6-core GPU
unified memory
ry bandwidth