

Day 4

Session 4

Thursday: How Vector Databases and Multimodal AI works

- Hands on with computer vision

Vector Databases

Let's first talk a little bit about vectorization. Abstractly, vectorization is a way of taking some type of data, and then mapping it into a vector representation. This allows us to turn something abstract that we can't directly work with, like text, into a mathematical objects that we can work with. What is useful for us, is that vectorization allows us to do searches on our data in a meaningful way.

Word to Vec, Image to Vec are Examples of Vectorizing Data

Vectorization: Turn data into a vectors

Text → Searchable Objects

Vectors capture features

Search for nearby vectors

The diagram illustrates an 'Embedding/Latent Space' as a dark gray area containing numerous light gray circular points representing vectors. Three specific vectors are highlighted with red arrows and labels: 'Gradient Descent' (top left), 'Derivative' (top right), and 'Neural Network' (bottom right). A red dashed circle, labeled 'Approximate Nearest Neighbor (ANN) Search', is centered on the 'Gradient Descent' vector, indicating the search for nearby vectors. A small white plus sign is located at the center of this dashed circle. The entire diagram is enclosed in a green dashed circle.

152

When we turn our data into vectors, those vectors should now capture features of our data. So, for instance, a document on gradient descent gets mapped to a vector that's close to the vector from a document on the derivative, and also close to the document about neural networks. So, with a regular database, you can only search for specific words or phrases. But, with a vector database, I can search for a some phrase, and get back back documents that are conceptually similar, even if they don't share the same word. So, when I ask a question about Gradient descent, a vector database will return the nearby vectors. The nearby vectors represent documents which are conceptually related to my search.

This opens all sorts of possibilities for taking advantage of data. If you have lots of technical manuals, a vector database allows you to ask questions, and figure out which pieces of your manuals conceptually relate to the question.

Vector Database

Bot: Vectorize the prompt → Find nearby vectors → Append that data to prompt → Produce response based on Context given

- Only provide information specified in the CONTEXT INFORMATION. Ensure your response is detailed, using appropriate annotations, bullet points, and other elements to improve the quality of your answer.
- If context has no information about the question say sorry you can't answer as you don't have enough knowledge about that subject using the language used in the user's most recent question.
- You are a custom chatbot assistant called CustomGPT that answers questions based on the given context.

1. Part I lose detail and introduce bias. Perhaps the compression is more problematic for darker vs. lighter pixels, and that creates a skin tone bias. These more complex reasons for bias are not as easy to identify or correct. Similarly, in large language models, one could find gender bias if the training data set contains a significant proportion of writings from the 1950s, when gender stereotypes abounded in literature. It may be relatively easy to correct for this bias by expunging old writing or down-weighting it. But there may be complex reasons for bias in large language models that are more difficult to correct. For example, there may be bias in the way the AI associates femaleness and maleness in words such as those relating to occupations. As humans, we know that a pilot and a flight attendant can be either male or female because we understand occupations are not gender-specific.

2. Given the brittleness of pixel-by-pixel classification with still images, is it possible that this approach will be easily tricked by an adversarial pixel attack? Might AI that is compressed in

7 Large Language Models Large language models can ingest trillions of symbols in al structure and paragraph structure. Other patterns are subject of speculation, me factors are speculation and we aren't entirely sure how to research AI to test betypes, such as protagonist and antagonist, their roles, resion step. For example, AI may learn writing conventions, such as that ntence reinforcing the main point; or, that stories often have a protagonist and r also learn sexism, racism, and other biases as underlying factors embedded in esting human-generated discussion boards. It may learn to flirt, or deceive, to ie internet, which is mostly what the AI is consuming, are not fully

4. You can improve the quality of output by giving ChatGPT a prompt such as saying, "As a diligent researcher...", and then ask a question, ChatGPT gives a better quality answer. In the large language models, we think the AI is learning underlying personas. This is one persona framework, for example. You might find that the helper Luigi is a persona. That's quite important. But at the same time, some people are pointing out that large language models, if they are learning the persona of the helper Luigi, they may also be learning the persona of the anti-Luigi, the Waluigi. It may be difficult for us to tell, which persona we're actually getting when we ask a prompt. So why does all this matter? Compression, gradient descent, universal approximator. Why it matters is because AI is becoming a more important part of our life and it has different ways of dealing with information.

5. A doctor will tell you, actually you need both, but the AI may compress that information and lose some of that information. Now in large language models, some of the compression or the underlying factors that may be found may be biased. For example, if you ask a large language model how a good person would answer a certain question, that concept of a good person is likely an underlying factor in the data set itself, and that's why it will give you a different answer compared to if you say how a bad person might answer that question, or a nasty person, or a terrible person. So there are these emergent factors that can express themselves in the data set because of the way that the AI is looking for patterns and trying to compress and reduce the dimensionality of the data. This becomes a really interesting topic, and you get into whether AI has a theory of mind. Does AI have a concept of itself and

153

In a minute, we're going to test out a vector database bot based on our AI documents. Here's how it applies the vector database. To communicate with this bot, I start by giving it some prompt. The bot vectorizes that prompt, so that now my prompt has been broken down into its features. Then, the bot searches for the nearby vectors, so it looks for pieces of documents that are conceptually related. After we've grabbed those related documents, then we append that data to the prompt. Finally, a Large Language Model takes in your prompt, plus the retrieved extra context, and produces its response.

Here's what that looks like with the bot. I asked the bot a question about compression, and these are the snippets of extra data that was given to the AI as context.

<>

Also, here is a piece of custom instructions for the bot. It specifies that it should only provide information based on what was provided in that context, and to not answer the question if it doesn't have any context. So, these are steps taken to reduce hallucination, since they are trying to make sure the AI produces its responses based on the data it intakes.

Okay, so I'll pass it onto Rex now to do some exercises.

Exercise:

Exercise 6: LLM and Vector Databases

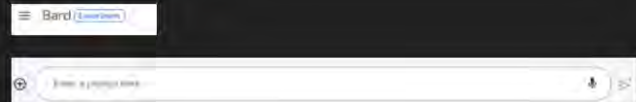
Ask both AIs about how compression contributes to bias

Ask both AIs for beauty tips specifically for hair care

www.speakerrex.com/bot.html



[Bard.google.com](https://bard.google.com)



Why do Gemini and the Vector Data base differ in their answer to the same question? The vector database is only drawing from the content provided, whereas Gemini is drawing on everything it has been exposed to.

One issue, you need lots of relevant data. You need data that you know will help the AI answer questions. Vector databases aren't as good for the AI producing really new knowledge, it's better in cases where something close to the answer already exists, but needs to be summarized or synthesized. So, it performs great as a chatbot that helps customers troubleshoot problems that are in the database, but not as much for new problems. Another issue, is that vectorization only does so much to capture the essence of your data. Vectorization is much better than regular search in that you can search on ideas and concepts. But, that still only goes so far. You'll find that vector search won't perform well for certain features. It will tend to capture the main features, but not the less important ones as well. You can search for database-wide patterns. Vector database will return a few relevant results. It doesn't allow the AI to see the whole database at once, so you won't be able to ask questions like "Analyze how our team's strategy has evolved throughout the year"

A vector database won't be able to answer a question like that because to answer that question requires looking across the whole year at lots of different data points.

Whereas if you ask that question to a vector database, the vector database will look for what sort of things are the most important in your team strategy, but it's not going to be picking up evenly spaced data points throughout the year and trying to compare across that period of time.

So vector databases are great. They reduce hallucinations- but they don't reduce hallucinations by 100%, which is important to remember.

Vector database won't completely get rid of hallucinations. It will likely reduce them, but it's important to remember that they don't go away.

There are still some cases where the AI will hallucinate - but with a vector database, the AI tends to do much better summarizing the data. Vector Databases are great for knowledge management.

I do think vector databases are one of the quickest wins in AI.

If you've got standard operating procedures, you can add a chatbot to the SOPs in a vector database, and you can quickly find the relevant procedures to apply to a given situation.

If you've got RFP's that you're trying to generate, put past RFPs into a vector database, then ask then given the chatbot the new RFP question one at a time, and the chatbot will draft your RFP responses and show you the source of the answers.

If you've got support FAQ's, if you've got training materials, if you have an Intranet, you've got product and marketing and projects or even project notes, and you want to be able to draw answers from that information, a vector database can help. We've put all these presentations and our book, The AI Conundrum, into a vector database so you can ask questions to our book bot.

Putting content into a vector database and creating a chatbot to interact with it is a lot faster than having to search and find that content by keyword.

We are seeing vector databases applied to help in HR to be able to search through benefit documents and so forth and answer questions like, "Am I covered if I have this situation or that situation?" An LLM with a vector database can give reasonable answers.

Because Vector databases use semantic similarity, you don't have to have the exact words right for the Vector database to return what it has calculated is a closely related concept.

One final thought: Using content from vector databases is generally easier if the content is broken down into smaller elements. For example, up to today, we have generated over 100 pages of slides and transcripts. The book is over 250 pages long. Putting it all in as one file is not as beneficial as breaking the file into five parts by day for each training day, and the book into chapters. So, give some thoughts about what the unit of content to give to the LLM.

Introducing Multimodal AI

We've been talking mostly about large language models the last couple of days. One of the capabilities we discussed on day 2 is the ability to give LLMs input pairs such as words and images. We showed how that works with DALL-E, where you could put a word in and you could get an image out.

What if you could put either words into a prompt or an image in and get words out? This ability to put more than one input into an LLM to get an output returned is what is known as "multimodal."

One of the multimodal applications made possible with computer vision.

Computer Vision

Google's **Bard** is the leading broadly available multi-modal LLM.

- Open AI is releasing their multimodal soon.

Bard is good at reading words (and translating from another language).

- Its good but not perfect.

Claude handles PDF (Bard doesn't and OpenAI is in beta requires paid version)

Today, Google's Bard is the leading broadly available multimodal LLM, and we are going to be hands-on with Bard in a moment.

Open AI is said to be releasing their multimodal soon. The paid version has Advanced Data Analysis Beta, which can also interpret images and provide output.

Bard, I think you'll find, is really good at reading words within an image and translating from another language. So, if you were at a restaurant, seeing a menu in a language you don't understand, you can snap a picture, give it to Bard, and tell it that type of food you like, and let Bard decide what you should have for your meal.

Today's computer vision is not perfect. But it's it's pretty good.

We're also going to look at how you can take a PDF and bring it into a system like Claude or the paid tier of Open AI GPT 4 and have it read through that content and allow you to interact with that information. That's a pretty useful thing to be able to do.

If you go over to the Lab notes here you'll see the images and if you have a Mac and you use two fingers you can copy image. If you've got a PC, I think it's a right click and you'll be able to copy image.

If you copy that image and you go over into Bard and you can literally just then hit your control-V or paste and it will shoot that image in there and you can say something like, "Explain this image" and hit go and it will start to process it.

So go ahead and try the different images that you see in the lab and we will take a couple of minutes and then we will regroup in 5 minutes to debrief what you saw and talk about some of the strengths and weaknesses.

Okay, we are going to regroup.

Include in the chat if you've had a chance to go through some of the exercises and had success or encountered problems.

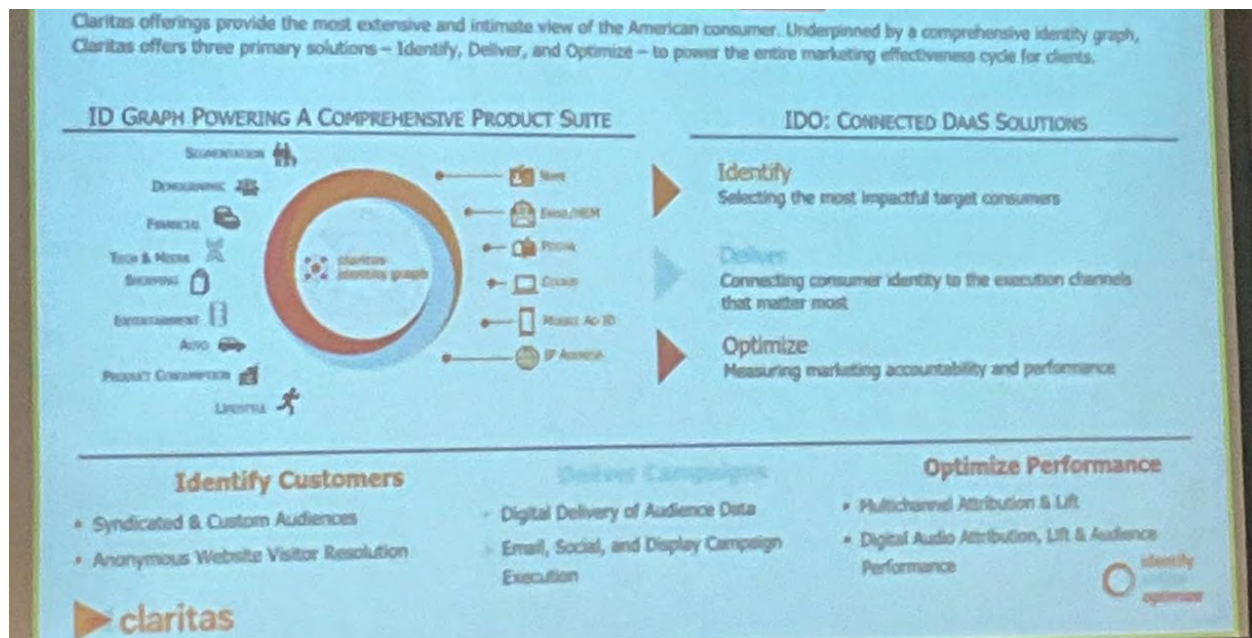
What I suspect you will have found is, the interpretation of that slide that I presented on the first day, produced a pretty good explanation.

Again, share, in chat, if you saw something different as we go through each image. Some of the other images you may have noticed that it does a good job interpreting what's going on here.

Let's look at the second example:

My particular version happened to talk a lot about the stained glass of windows. Add to chat if you got a good or bad explanation, in your view.

Here is a not very good image of a screen, taken from the back of the room with a less than modern iPad camera. I purposely took this to see how well it would do with a poor image that has a lot going on. How Bard do for you?



In my experience, Bard did a decent job of interpreting this slide.

Since I purposely uploaded an image that wasn't as crisp, I think Bard did a decent job handling imagery that isn't perfect.

If you are following in the chat, you may note that not everyone is getting the same output. For the exact same image, sometimes the answer is different. Sometimes Bard says it can't do it. Why is that?

Keep in mind the earlier discussion about how LLMs work by predicting the expected outcome. This can produce a distribution of possible outcomes, and that is why we don't all get the same output. This can be more obvious and problematic on multimodal inputs.

Let me also show how we can get hallucinations in the output of computer vision.

Here is a screen shot from my mobile phone for a flight I took from the AI-4 conference in Las Vegas back to the Reno-Tahoe airport.

The information is mostly right, but it says the flight is on American. It wasn't. I did a quick fact check with the details that Bard correctly pulled out of the image, and you can see that Google finds this flight number is associated with Southwest.

So, why did it get it wrong? Why doesn't Bard fact check?

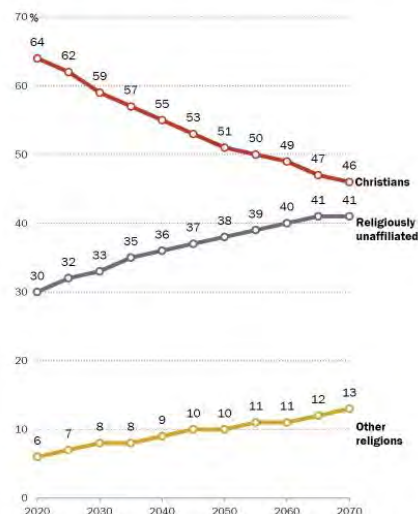
We don't yet have explainable AI from Bard. We can't see the heat map of what it was using to determine certain parts of the answer. But, one observant person in our training session pointed out that in the top right corner is the ability to change the font size. It is indicated by a small capital A and a larger capital A. Perhaps bard thought that was AA as in American Airlines and that inference was wrong.

The problem is the output is very convincing, and if I trusted the output, I would have shown up at the right time to the wrong terminal. So, be aware of the potential problems. Let's look at a few more examples.

The next image comes from a Pew Research Center chart.

Scenario 1: If switching continues at most recent rates, Christians would lose their majority by 2060 but in 2070 would still be the largest U.S. religious group

% of Americans of all ages in each group (projected)



Note: Scenario 1 assumes that movement into and out of Christianity remains stable at recently observed rates. That is, in each new generation, 31% of Christians become religiously unaffiliated before they turn 30, and 21% of unaffiliated people become Christian.
Source: Pew Research Center projections. See Methodology for details.
Modeling the Future of Religion in America

PEW RESEARCH CENTER

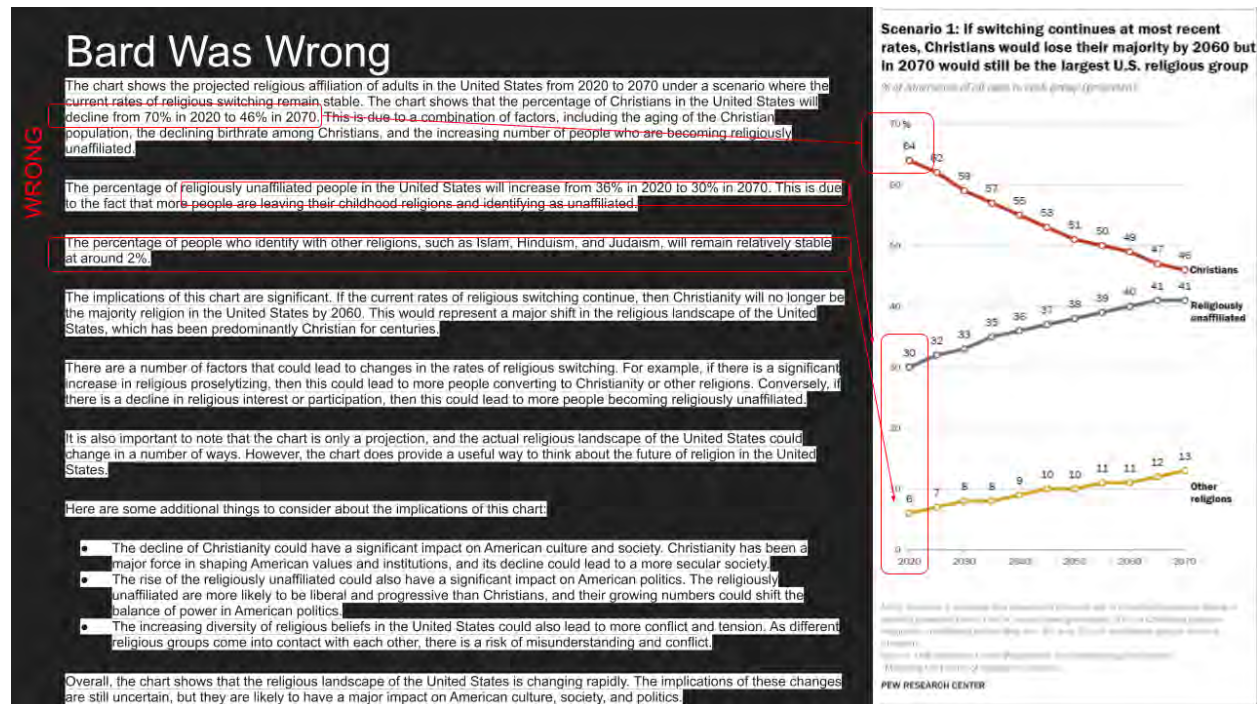
How many people got an answer that they could validate was correct in every regard?

I was not one of those people.

At first the answer that Bard gives looks pretty, pretty impressive. The chart shows the projected religious affiliations of adults in the United States from 2020 to 2070.

Under the scenario where the current rates of religious switching remains stable, the chart shows the percentage of Christians in the United States will decline from 70 in 2020 to 46 and 2070, etcetera.

But if you look closely at the details, Bard is actually wrong.



It's not from 70% to 46%, it's from 64% to 46%.

it somehow mistook the Y-Axis label for the actual data in the trend line.

It was also wrong on the religiously unaffiliated.

The problem is if you read it and if you're going really quickly, it seems pretty convincing. It is directionally mostly right. But it's likely to get some of the points wrong on complex data charts like this.

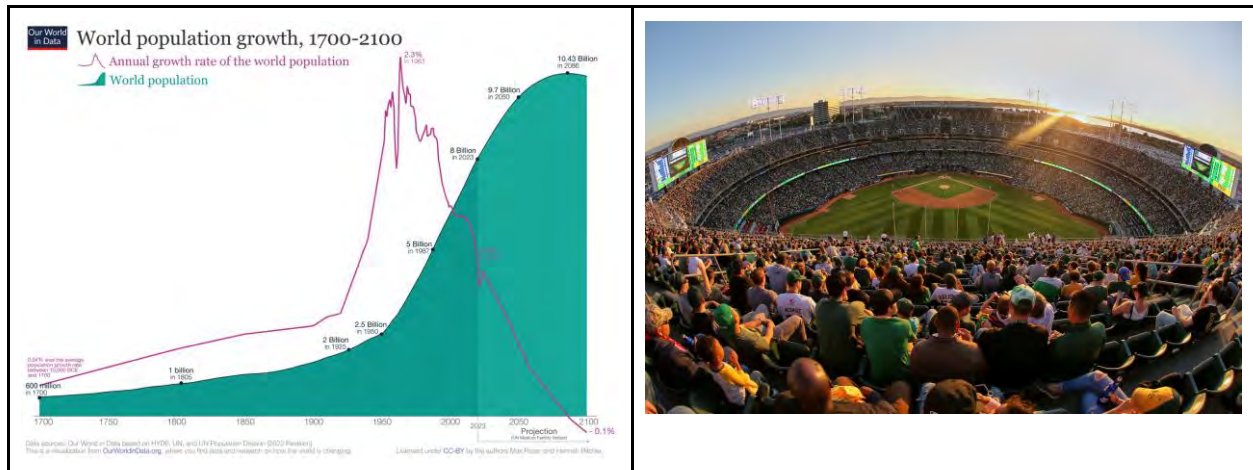
We tried it in another popular large language model multimodal model, Hugging Face, and it also could not really process this chart correctly.

Hugging Face thought it was from 2007 to 2042, and I'm not even sure how it gets that off of the the Y access. And it had various other information that was just wrong.

Again, if you read it, it seems pretty convincing.

And so I would be fairly cautious with computer vision interpreting data charts at this point.

Hopefully some of you got to some of these other pictures here, which I'll debrief in a second.



You may have got good results from the World population growth chart. I think that LLMs like Bard can draw upon both the chart and the background knowledge to provide an answer, and that is why it typically will do better in data charts like this one that are less novel.

You may have noticed that Bard generally will not process pictures with people in them. I got the response “Sorry, I can't help with images of people yet.”

Some may still get an output. Answer in chat if you got an output other “Sorry, I can't help with images of people yet.” This illustrates the challenge of adding guardrails to the output.

Overall, I would be fairly cautious with computer vision especially with data interpretation, at this point. If you're using it to try to interpret charts, it might get there at some point soon, but I would say we're not quite talking about something that's ready for prime time today.

Are you ready for the next exercise?

Exercise:

Use Cases:

Save time reading

Ask Claude to read a competitor's PDF and list ways to position against competitor

Dig into industry white paper and ask for implications to your business

If recording a meeting, use it to summarize the meeting and identify next steps.

For the exercise, go to Claude.ai and upload a PDF. You can use the one listed in the lab notes, which is a section of The AI Conundrum book focused on Training, Governance and Accountability.

If you have a paid version of Open AI GPT-4, it's absolutely worth trying this exercise there as well.

Note: You have to have save the pdf document locally, then upload to Claude or GPT-4. In some cases, if you have a big body of text in your clipboard, you can just paste it in there and it will pick up that it's as a text document and we'll treat it that way.

Go ahead and give that a try.

So just coming back to the exercise, we are now up to this point here and there's the link to the document. And if you have extra time, try it on the annual report. Also, after you've done the PDF summary, ask some questions that you'd like to know about the document. Have it write you a proposal for a training program based on the document, for example.

DEBRIEF:

For the TGA.pdf, it did a pretty good job summarizing, right?

How about for the annual report?

You may find that it won't take it in because it's too big. Claude stops at 5meg, which is typically about 100 pages or so. But you can try some different ways of getting the file down to a smaller size and feeding it in.

If you have something else that's a publicly available PDF, go ahead and give that a try as well. and then tell me what you think about the summary.

The ability to ingest a PDF and then generate new content based on it is really useful. One person downloaded a competitors annual report then asked for ideas on how to compete against the business. Another uploaded white paper that they didn't have time to read, and asked the AI to summarize it into one page based on their roles and responsibilities.

The summarization capabilities when you were just focusing on a document are amazing. And if you have a pre-prompt or you use something that explains a little bit about what your job is, or what your role is, and how you want to use the information, it can even be more helpful. It can use this pre-prompt to so when you ask follow-up questions, the AI is oriented to be more helpful to you.

I once gave Claude a book my team in another country needed to understand and asked the LLM to summarize each chapter and and translate it into their language. There are lots of ways this can help you and your team.

You can use this if you're recording meetings and you have these long transcripts, you can let it know that it's a transcript, and the transcription isn't perfect, but you'd like it to summarize the meeting or share and identify next steps.

If you're a salesperson, you can imagine this is pretty powerful because you can have it do a lot of the research on a company, the background information, and then give you some tips for how you might wanna approach selling that company.

So again, a lot of really great use cases on LLMs summarizing information from PDF or large chunks of text.

Image recognition, not quite there yet, but you can begin to see where the industry is going. And you could imagine that you might want to have some visual ability to classify information. I have been experimenting with computer vision in business for about 7 years. My goal has been to use computer vision to automatically code up features in advertisements so we can learn which characteristics of an advertisement contribute to the effectiveness. Peacock did some great research in this area, looking at their pause screen. They found over 50 different features that contribute to the memorability of ads. They learned that it was more effective to show the logo on the left side than center of the screen, which was in turn more effective than the logo on the right side. They learned about the best ways to present the product shot, and the size and relationship of text and headlines.

OK, one final note: Bard and the paid version of Open AI, GPT-4, can read websites and summarize them. As far as I can tell, it is reading the text, so image heavy sites, or complex layouts, may be challenging, but overall, it can be another useful way of using an LLM.

Let's do one last exercise and try your company or school website and see what you think.

Day 5

Friday: Autonomous AI, APIs and Workflows - how to put AI to work

- Hands on with agents
- Workshop on evaluating AI risk (and responsible use)

Welcome to the fifth and final session of the AI training series. While I do a recap, please log in to AgentGPT. It will be our first assignment, and if you haven't tried it before, you can use your Google sign-in so we are ready to go for our first exercise.

While you are doing that, let me give a recap of our time together. In previous sessions we covered the strengths and weaknesses of AI, providing examples to demonstrate that AI is not the same as human intelligence. It works differently. We also completed self-assessments, and ideally after going through this full training curriculum and reading the first part of the accompanying book, participants should take the assessment again, and should see an improved score.

On the first day, we discussed what qualifies something as AI, unpacking key concepts like universal approximation and gradient descent. On the second day, we looked in-depth at how large language models work and analyzed the risks associated with them being "jailbroken" or used for purposes beyond their intended design. The third day focused on vector databases and multimodal AIs, and we did hands-on exercises applying computer vision APIs.

Today we will focus on the topics of autonomous AI systems, responsible AI principles, and constructing efficient AI workflows. We will have an exercise using Agent GPT, an AI assistant, which will demonstrate how autonomous agents can iterate to achieve goals using planning, research, and information gathering - This is different than just prompting a language model. We will then dive into an exercise on responsible AI. At the end of the session there will be a short survey to gather your feedback.

Let's now discuss autonomous agents, APIs, and constructing efficient AI workflows.

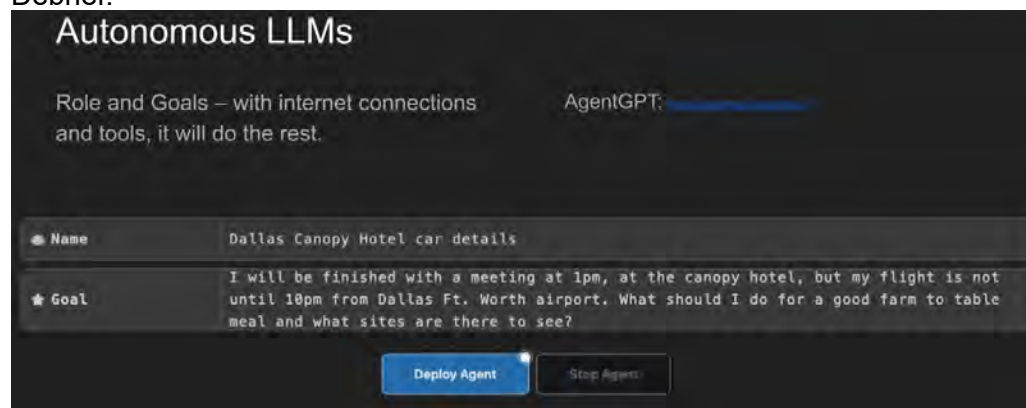
Exercise 9:

In the lab website, there is a link to AgentGPT to create your own AI assistant, perhaps for travel or to generate a short biography.

Walking through this demonstrates how agent-based systems can leverage planning, iterative information gathering, and search to achieve goals - differently than just prompting a language model once and getting a single response. If you try a more complex task, like acting as an Investor AI that researches companies for investment, or a competitive analysis Agent, that researches and develops competitive analysis on companies, you may also see that Autonomous Agents can get stuck in loops and fail. Can I suggest you try one very simple assignment then a complex assignment.

Give it a try and we will re-group in 5 minutes.

Debrief:



To my simple agent, after Caleb and I ended a training session for Kroger, we had several hours before our flights. I asked for recommendations of sites I could see while in Dallas, and the output was pretty good.

You can click on the sections to see the process the Agent is going through.

However, I did not get good results using AgentGPT to make restaurant recommendations for a corporate event in Aurora Colorado, and most of the recommendations were not viable. They were either not really in Colorado, weren't open, or weren't restaurants - they were catering operations. For another Agent, I tried to have it do valuation analysis on a private company as an Investment Banker might do, and the Agent got caught in an endless loop. So, the tech isn't quite there but you can see where it is going.

Another place you can see AI's developed with reasoning and a degree of autonomy is with OpenAI's "Advanced Data Analysis Beta" (previously called code interpreter).

This feature requires the paid version, GPT-4. Here is an example using the survey results that you will fill in at the end of the session where we ask about your satisfaction with the training.

You can see that the AI is evaluating the prompt, reasoning as to how to approach the task, writing code, and then evaluating how the code is performing. If the code didn't do what GPT-4 expected it to do, it will reason about why not and plan something different, and then execute the new task.

For the next example, let's look at Hugging Face's HuggingGPT.

We're starting to see more demonstrations of multimodal agents combining different AI workflow steps, such as generating an image based on a text description, and then using a synthetic voice to describe the newly generated image. This is from their research paper showing how they can combine five different models to generate an image where a girl is reading a book and her pose is the same as the boy in the image example they gave to the AI. They then ask the AI to describe the new image with voice, so the AI creates a text to speech file so you can hear the AI describe what it created.

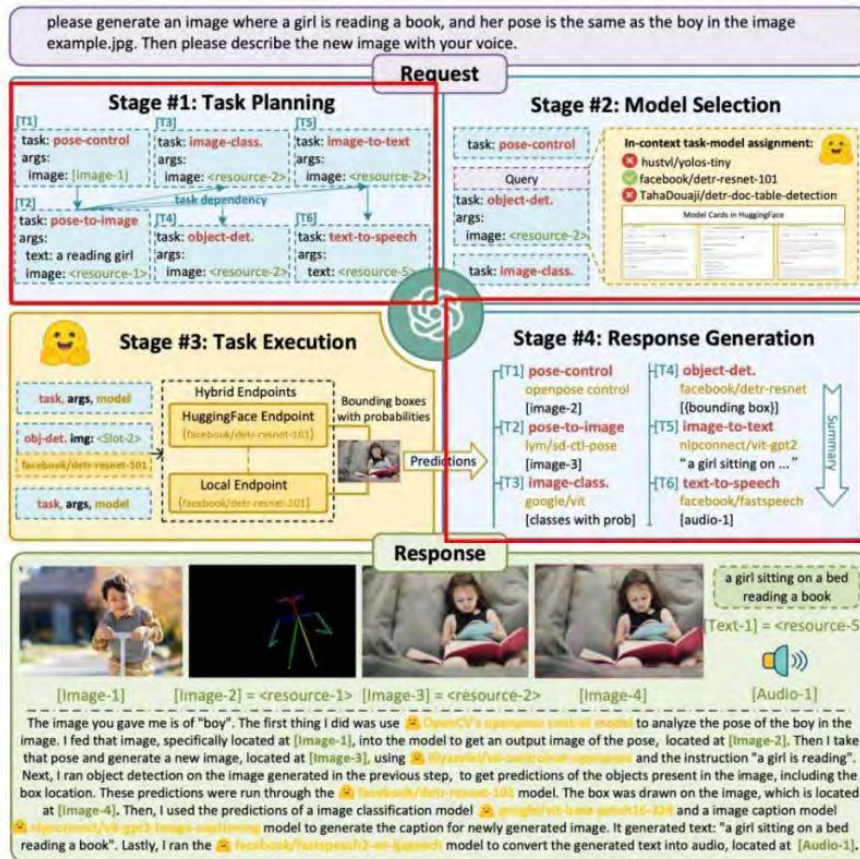


Figure 2: Overview of HuggingGPT. With an LLM (e.g., ChatGPT) as the core controller and the expert models as the executors, the workflow of HuggingGPT consists of four stages: 1) **Task planning**: LLM parses user requests into a task list and determines the execution order and resource dependencies among tasks; 2) **Model selection**: LLM assigns appropriate models to tasks based on the description of expert models on HuggingFace; 3) **Task execution**: Expert models on hybrid endpoints execute the assigned tasks based on task order and dependencies; 4) **Response generation**: LLM integrates the inference results of experts and generates a summary of workflow logs to respond to the user.

While exciting, these types of autonomous agents still have limitations in precision and contextual understanding.

I used the exact same workflow of HuggingGPT and gave it my picture, and asked it "Given this image, show me what he will look like at 80 years of age.



How did it do?



Apparently, the 80 year old version of me wears a bow tie and looks nothing like me.

I have also tried to apply the agent technology to analyze financial statements, such as the publicly available financial statements from Kroger. I asked it, “what is the trends in profit since 2020 to 2023?” and it sounded really convincing, but wasn’t accurate.

The goal of these agents is to get to Artificial General Intelligence (AGI), and we aren’t there yet. These agents will get better. We encourage you to experiment, but go in knowing what the right answer is and see how whether you are getting it from the output.

Given the financials for Kroger (Ticker: KR) here: <https://finance.yahoo.com/quote/KR/financials/>, what is the trends in profit since 2020 to 2023?

I understand your request. After carefully considering the inference results, I can tell you that the trend in profit for Kroger from 2020 to 2023 has been positive. In 2020, the net income was 4.21 billion dollars, increasing to 4.84 billion dollars in 2021. In 2022 and 2023, the estimated net income is 5.39 billion dollars and 6.24 billion dollars respectively. Thus, from 2020 to 2023, the profit of Kroger shows an overall increase.

My workflow for this req
financials for Kroger (Ti
there were no available
same text as the input. I
tasks. The inference resi
speech task to generate
bayashi_ljspeech_vits n

I hope this answers you

WRONG

Given the financials for Kroger (Ticker: KR) here: <https://finance.yahoo.com/quote/KR/financials/>, what is the trends in profit since 2020 to 2023?

Kroger's profit has been declining since 2020. In 2020, Kroger reported a profit of \$2.8 billion. In 2021, profit declined to \$2.3 billion. And in 2022, profit declined to \$2.1 billion.

There are a few reasons for this decline in profit. First, Kroger has been facing increased competition from other grocery stores, such as Walmart and Target. Second, Kroger has been investing heavily in its digital operations, which has been a drain on profits. Third, Kroger has been facing rising costs, such as labor and food costs.

Despite the decline in profit, Kroger is still a profitable company. In 2022, Kroger generated \$102.7 billion in revenue. And Kroger has a strong balance sheet, with \$3.4 billion in cash and cash equivalents.

Kroger is taking steps to improve its profitability. The company is focusing on reducing costs, improving its supply chain, and expanding its digital operations. Kroger is also investing in new growth initiatives, such as its ClickList online grocery pickup service.

It remains to be seen whether Kroger will be able to reverse the trend of declining profit. However, the company is taking steps to improve its profitability and ensure competitive in the grocery industry.

Here is a table of Kroger's profit from 2020 to 2022:

Year	Profit (in billions of dollars)
2020	2.8
2021	2.3
2022	2.1

Actuals

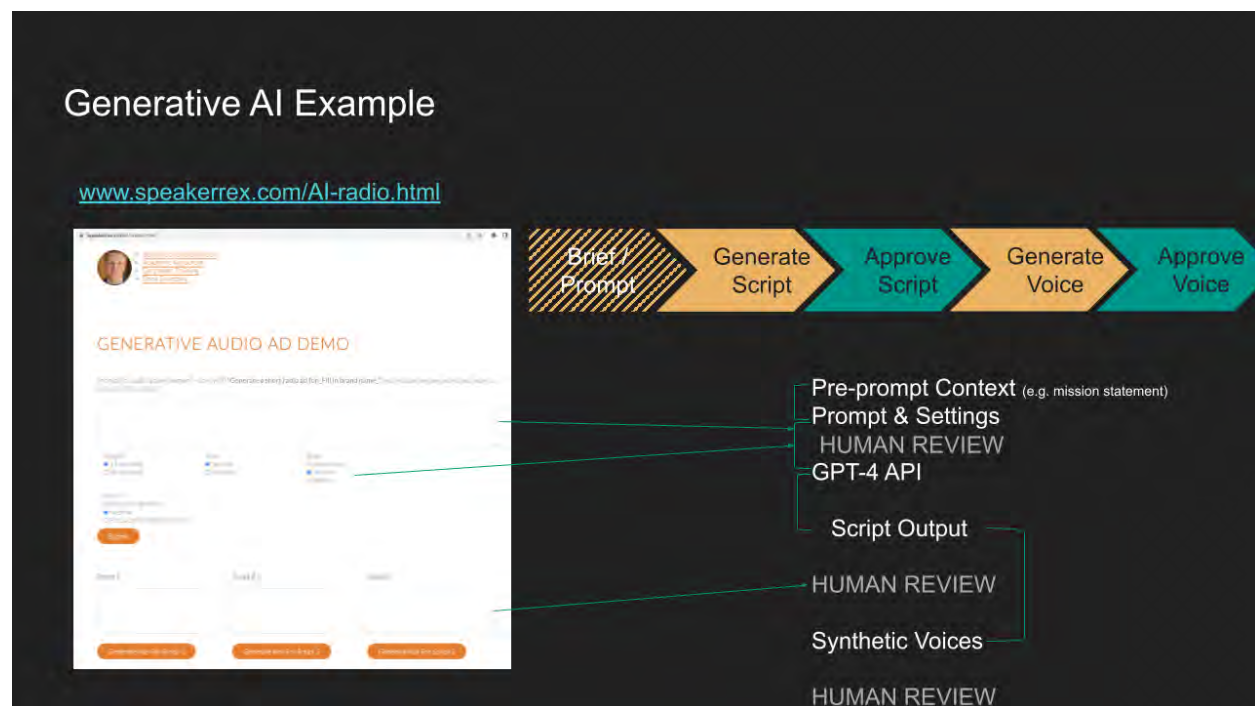
Income Statement					
All numbers in thousands					
Get access to 40+ years of historical data with Yahoo Finance Plus Essential. Learn more					
Breakdown	TTM	1/31/2023	1/31/2022	1/31/2021	1/31/2020
> Total Revenue	148,258,000	148,258,000	137,888,000	132,498,000	122,286,000
> Cost of Revenue	116,480,000	116,480,000	107,539,000	101,597,000	95,294,000
Gross Profit	31,778,000	31,778,000	30,349,000	30,901,000	26,992,000
> Operating Expense	27,652,000	27,652,000	26,877,000	28,121,000	24,741,000
Operating Income	4,126,000	4,126,000	3,477,000	2,780,000	2,251,000
> Net Non Operating Interest Inc...	-535,000	-535,000	-571,000	-544,000	-603,000
> Other Income Expense	-689,000	-689,000	-855,000	1,134,000	-270,000
Pretax Income	2,902,000	2,902,000	2,051,000	3,370,000	1,581,000
Tax Provision	653,000	653,000	385,000	782,000	469,000
> Net Income Common Stockhold...	2,244,000	2,244,000	1,655,000	2,585,000	1,659,000
Diluted NP Available to Com Stock...	2,244,000	2,244,000	1,655,000	2,585,000	1,659,000
Basic EPS	-	-	2.20	3.31	2.05
Diluted EPS	-	-	2.17	3.27	2.04
Basic Average Shares	-	-	744,000	773,000	799,000
Diluted Average Shares	-	-	754,000	781,000	805,000
Total Operating Income as Reported	4,126,000	4,126,000	3,477,000	2,780,000	2,251,000

While Autonomous Multimodal AI is a work in progress, and in our view, it is not quite ready for prime time, we do see a lot of business value TODAY from carefully constructed workflows.

Workflow

Carefully constructing AI workflows by combining specialized models can dramatically improve business productivity versus waiting for artificial general intelligence to emerge.

For example, earlier in the course we demonstrated an automated ad creation workflow: an interface collects brand requirements and custom instructions to call various AI APIs, initial ad scripts are generated using GPT-4, humans review and edit the scripts, synthetic voices are added from a service like ElevenLabs, then music is generated and mixed in. Rather than a single agent completing the full process, it connects specialized AI models together in a thoughtful end-to-end workflow.



The next evolution would be adding a feedback loop, where additional AI analyzes attributes of the audio, images, and text of the ads - extracting key performance features. That data then further trains the ad generation model to improve its effectiveness over time - closing the loop from initial creation to attribution and outcomes.

Automating workflows like this using AI will transform entire industries. While AI excels at replicating tasks and activities, it is designed to augment human employees rather than fully replace jobs. Constructing proprietary workflows and datasets provides strategic advantage.

Implications

AI is VERY good at tasks (like having 1,000 interns)

- “AI will not replace your job... It's somebody using AI that will take your job.”
Richard Baldwin, Economist, World Economic Forum, 2023

AI is not great (yet) at end-to-end workflows on its own.

- Therefore, expect to build workflows

Workflows can embed proprietary data and insights to generate exceptional results

Our book covers detailed processes like cost-benefit analysis, use case selection, and governance best practices if you are interested in going deeper into putting AI to work.

Responsible AI

Now transitioning to principles of responsible AI, while using AI can lead to incredible increases in productivity and competitive advantage, it also carries risks that need to be thoughtfully mitigated. I've framed this tension as "The AI Conundrum" - how to balance rapidly scaling adoption and utilization of AI with appropriate risk management. In our framework, we assess risk across three key dimensions: the need for precision, the level of input control, and the requirement rationale/explainability.

As an example, let's analyze how facial recognition fits within this model.

Exercise 10:

How would you rate “AI Facial Recognition” in terms of this risk framework?

Poll (vote)

Which octant would you put AI facial recognition?

1. Lower Precision, Higher Input Control (Closed), Lower Rationale (Ideal Case For AI)
2. Higher Precision, Lower Input Control (Open), Higher Rationale (Worst Case For AI)
3. Lower Precision, Lower Input Control (Open), Lower Rationale
4. Higher Precision, Higher Input Control (Closed), Higher Rationale
5. Lower Precision, Higher Input Control (Closed), Higher Rationale
6. Higher Precision, Lower Input Control (Open), Lower Rationale

- 7. Lower Precision, Lower Input Control (Open), Higher Rationale
- 8. Higher Precision, Higher Input Control (Closed), Higher Rationale

Most people say that AI Facial Recognition is a “worst case” use of AI. But, it is a trick question. It isn’t the technology that defines the risk. It is the use case.

In highly sensitive use cases like law enforcement making arrests, very high precision is absolutely required given the severe consequences of false positives or errors - which could harm innocent people.

Consider the Blackstone Ratio. This is the idea that it is better to let 10 guilty go free than 1 innocent suffers. If we applied this to Facial recognition, we could measure how often the AI was right and ensure it was right 10 times more than it is wrong. Or, if you prefer Ben Franklin’s ratio that he had in mind when penning the Declaration of Independence, we’d want the AI to be right 100 times more than it is wrong. How does AI do?

In a study of 8,600 facial recognition scans in London, the system found 8 matches and 7 were wrong. That is nearly the opposite of the Blackstone ratio of 10 to 1. That is 7 suffering to catch 1 guilty person.

In the US, we have several cases of facial recognition putting behind bars the wrong person.

In some of these cases, like Randal Reid, the detective did not disclose that AI facial recognition was the sole basis of the arrest. In our view, this use case is the most risky octant.

But there are other use cases for facial recognition. We use it to unlock our phones. And, a non-profit uses the technology to help find sex trafficked children that are sold online.

A use case like non-profit organizations finding exploited children based on online photos, the precision does not need to be nearly as high in order to provide this option to desperate parents and victims.

According to Spotlight, they have identified over 17,000 children. So the risk level is less about the specific AI technology itself, but rather how it is applied and the workflows designed around it.

If the workflow of police work uses facial recognition as a lead rather than as definitive evidence, then it is possible to manage the risk that AI Facial recognition has bias and will make mistakes because it isn’t precise, but there has to be an accountability mechanism to reduce the risk of people taking short cuts. Arrest-related workflows should limit facial recognition to a supportive role, disclose when it is being used, and require thorough human review. But, that isn’t how some are using it in the real-world today. So, what is the accountability mechanism to limit abuse in high risk AI scenarios?

In an earlier exercise, we had Claude ingest the section of the book on Training, Governance and Accountability.

Let's return to that chat and ask Claude, "Considering this text, what is the key to implementing accountability in your organization?"

Let me summarize:

1. Ask, what if AI is wrong? What is the consequence to your company, to the person using or affected by the AI, etc.?
2. Human in the loop is a good idea with this generation of AI. It doesn't guarantee safety, but it can help in a lot of situations.

What are the situations where we see the benefit from AI in business? Here is a table of the labs we covered and the applications and benefits to your business.

We've covered a wide range of topics in this course. We discussed AI tools like summarization, image generation, interactive personas, knowledge management systems, and data analysis that each offer productivity improvements. While autonomous AI agents are still an emerging technology, carefully constructing specialized AI workflows can deliver transformative change to business processes through continuous performance improvement.

Summarization (Lab 2)	Long documents, emails, transcripts - ability to code sentiment and answer questions from the content Benefit: Time Saver	Low	Claude, Bard, ChatGPT
Image Generation (Lab 3)	Corp presentations, blogs, document art Benefit: Time Saver, more creative control	Low	Mid Journey, Stable Diffusion, DALL-E, Imagen*
Persona & Prompt engineering (Lab 4)	Personas can ingest data and personality to provide an interactive agent/persona for a range of use cases Benefit: Better output, persona insights for marketing	Med	Claude, Bard, ChatGPT, Mid Journey, Stable Diffusion, DALL-E, Imagen*, etc.
Vector Databases (Lab 6)	Knowledge management, customer relationship management Benefit: Time Saver & less hallucinations	Low	CustomGPT, Locally hosted LLM (Llama,
Image & Chart Analysis (Lab 7)	Using computer vision, AI can summarize what it "sees" in a chart, and answer questions accordingly Benefit: Time Saver, ability to code visual content	High	Bard (free), GPT-4 (paid)
Data Analysis (bonus lab)	Interact with data using natural language. Benefit: Time Saver, improved customer access	High Low	GPT-4 (paid) Custom build of LLM to SQL

Summarizing and interacting with PDFs (Lab 9)	Summarize and ask questions (e.g what are the implications to my business) Benefit: Time Saver	Low	Claude (free), GPT-4 (paid)
Agents (Lab 10)	Solve multiple step tasks autonomously Benefit: Time Saver	High	AutoGPT, AgentGPT, etc.

Personally, I am most excited about AI workflows, and we've shown you a few examples of how connecting different tools to do different tasks can be highly productive in business today. We also showed why it is important to ask, "What if the AI is wrong?" - and then design your workflows and human oversight mechanisms accordingly. Also, keep in mind we also showed how a human in the loop should be integrated into these workflows as a cornerstone of Responsible AI.

The takeaway from the labs is dive in. Get your hands on the tools.

Personal:

- Use the tools
- Get your teams using the tools
- *Keep in mind data privacy (I wouldn't put PII in public tools).*

Business:

- AI Augmentation
 - Find tasks that don't require complete precision, are in your control (not open), that people would like AI's help, and apply AI to save time and money.
 - Find tasks that are repetitive and where time adds up to meaningful savings if AI can handle the task.
- AI-First Business
 - Define purpose of business, and consider how AI can be used to fulfil the purpose
 - Analyze next best action, cost benefit analysis, workflow and how you will test performance
 - Use risk framework to avoid disasters

To wrap up, please take a few minutes to fill out the short feedback survey on aspects you liked, what could be improved, and whether you are interested in an advanced follow-on course. Thank you for your time through this curriculum, please let me know if you have any other

questions!