Introduction

Prompt Engineering

Welcome!

Mick McQuaid

mcq@utexas.edu

University of Texas at Austin

15 Jan 2025

Welcome!

Setting the stage

Let me set the stage for our adventure

- This is all new, this course, this topic
- There are no models for how it should be run
- We're pioneering new ground
- Congratulations! You can call yourself a pioneer but you could also call yourself a guinea pig!

Pioneering is good and bad

- You can probably guess the good aspects of being a pioneer, but what about the bad?
- We may have to change course often.
- We may go down blind alleys.
- We *will* face a lot of uncertainty.
- If this kind of thing bothers you, you should drop the course and wait until we have done all the pioneering work and you can follow.
- ⟨ Pause to let the room clear ⟩

Constraint

- There are no prereqs. You may not even know how to spell AI
- ⟨ Pause for laughter ⟩
- ⟨ If no laughter, tell mother-in-law joke ⟩
- Since there are no prereqs, we have to start before the beginning.
- I need to tell you a little bit of history, so you get the context.

Today's plan

- We'll do an exercise.
- Then we'll review Dhar (2024) for a bit of history.
- We'll learn how you'll do homework and presentations.
- Finally, we'll do another exercise.
- In fact, the general format of the class will usually be some lecture followed by some exercises.
- Next week, we'll discuss the general framework for the semester, but I want to give you something to think about first—one view of learning is that it is like hanging ornaments on an Xmas tree—you need the tree but you also need to appreciate the ornaments: which should come first?

Before we begin, TANSTAAFL

That stands for “There ain’t no such thing as a free lunch”

Many people think AI will provide a free lunch. I don’t.

You will have to work (think) to make AI work (think) for you

⟨ Pause to let the room clear ⟩

Exercise: What I already know

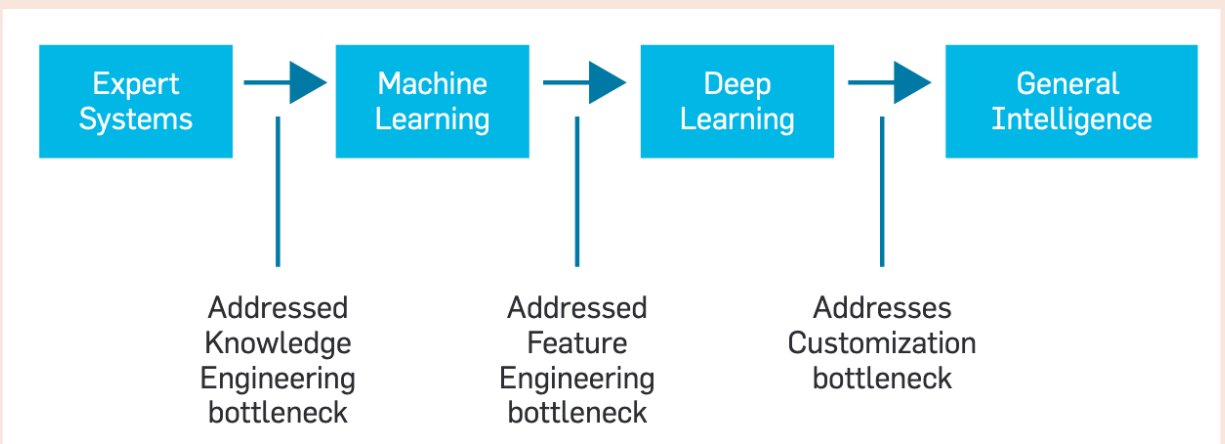
- There is a google doc referenced in Canvas (on the course's home page) that I want you to edit
- There is some starter text for you to add to
- Please also add your own sections but *please* use appropriate section headings and subheadings rather than jury-rigging text attributes as headings


LEVERAGE

Paradigm shifts in AI

Dhar (2024) covers four periods in AI history and the paradigm shifts between them.

Figure. The history of artificial intelligence.



This isn't all of AI history but is a simplified picture that will help us understand the task we face in making AI work for us.

Expert Systems

- Work in well-circumscribed domains where expertise is identifiable and definable
- Perform well at specific tasks where expertise can be extracted from humans
- Represent problems as relationships between situations and outcomes
- Considered intelligent if performing like human experts
- Example: *high pallor* $\Rightarrow$ *excess bilirubin*

Machine Learning

- Learn models from curated examples
- Driven by database technology, Internet, abundance of data
- Grew in parallel with neural networks
- Rely on loss functions measuring difference between predictions and empirical reality
- Influenced by philosopher Karl Popper
- Example: Predicting pneumonia from historical medical records of people with and without pneumonia

Deep Learning

- Usually conducted via deep neural networks
- Takes raw input (images, language, sound) instead of human description
- Acts as a *universal function approximator*
- Learns features implicitly associated with raw input
- Example: Vision system sees lines, colors, curves and, through successive layers of neurons, combines them into doors, windows, street signs

General Intelligence

- Pre-trained models integrate a large corpus of general and specialized models *not* optimized for a single task
- Language is a critical carrier of knowledge, powering large language models (LLMs)
- Pre-trained models learn through self-supervision on uncurated data, e.g., WWW

Paradigm Shifts

- Draws on the work of philosopher Thomas Kuhn
- Perceives history of science as punctuated by upheavals displacing dominant paradigms
- Identifies bottlenecks that lead to paradigm shifts

Table 1. The paradigm shifts in AI.

	Data	Exemplar	Scope	Curation
Expert Systems	Human	Rules	Follows	High
Machine Learning	+ Databases	Rules/networks	+ Discovers relationships	Medium
Deep Learning	+ Sensory	Deep neural networks	+ Senses relationships	Low
General Intelligence	+ Everything	Pre-trained deep neural networks	+ Understands the world	Minimal

Current Paradigm in Science

AI as general-purpose technology

General-purpose technology, such as electricity or the Internet, is defined by economists as a new method for producing and inventing that is important enough to have a protracted aggregate economic impact across the economy

Characteristics of General-Purpose Technologies

- pervasiveness
- inherent potential for technical improvements
- innovational complementarities

Challenging characteristics of the current paradigm

- opacity to humans, i.e., AI can't explain itself well (yet)
- explanation and transparency are important to humans
- legal questions of ownership and control of uncurated data, weights, and outputs
- nature of uncurated data as biased, rife with falsehood and noise
- AI is not designed to be truthful and is noisy and inconsistent

State of the world is a factor

- Alignment problem: are our subgoals aligned with our goals?
- and are the goals and subgoals aligned with those of AI?
- Example: Goal *Save the planet* might cause the AI to infer subgoal *eliminate humans*
- Scalability simplifies the work of bad actors

Human nature

- deepfake porn of hundreds of thousands of ordinary women was overlooked, but policy action was taken when a celebrity was the victim
- policy action is generally slow compared to technological advances
- policy action tends to favor the very rich and well-connected
- this has implications for all the legal and policy questions of AI

Conversation

Collaboration

All of us know more than any of us.

— Robert O. Briggs

The intelligence of that creature known as a crowd is the square root of the number of people in it.

— Terry Pratchett

The Art of Conversation

- If you search for books about the art of conversation, what will you find?
- The first book I found talked about conversation for pure pleasure
- The second book I found talked about body language during conversation
- Prompt engineering is essentially a conversation, but so different from what I think of as conversation
- We need to think of it as conversation, but study it separately from conversation that gives us enjoyment or happens face-to-face with another person

Presentations

Goal

You must each present (1) a success you've experienced with genAI and (2) a challenge that you have not yet overcome with genAI. That is the ten percent presentation grade in class. The Python exercise will determine the schedule of presentations.

- Please create the presentation using Quarto, which I will demonstrate.
- Please explain your success and your challenge (I will demonstrate in a minute)
- Please show your prompts (your dialog) leading to success and failure
- Please tell us how we can benefit from what you've learned

Example—success

I use a format called BibTeX to construct bibliographies. Here's a fragment.

```
119
120  @book{Venkataraman2024,
121    title = {Crafting effective prompts},
122    author = {Subramanian Venkataraman},
123    publisher = {Independently published},
124    year = 2024
125  }
126
127  @book{Phoenix2024,
128    title = {Prompt Engineering for Generative AI},
129    author = {James Phoenix and Mike Taylor},
130    publisher = {O'Reilly Media},
131    address = {Sebastopol, CA},
132    year = 2024
133  }
134
135  @inproceedings{deSantana2024,
136    author = {{de Santana}, Vagner Figueredo},
137    year = 2024,
138    title = {Challenges and Opportunities for Responsible Prompting},
139    booktitle = {Extended Abstracts of the CHI Conference on Human Factors in Computing
```

It's tedious to enter.

My first prompt almost worked!

Please format the following as a BibTeX inproceedings entry

Vinita Tibdewal, Sharoda Paul, Xiaomai Chen, Jennifer Kim, Obinna Anya, Mina Shojaeizadeh, and Ilteris Kaplan. 2018. Unlocking the User Experience of Generative AI Applications: Design Patterns and Principles. In Honolulu ’24: ACM CHI Conference on Human Factors in Computing Systems,May 11–16, 2024, Honolulu, Hawai’i. ACM, New York, NY, USA, 5 pages

Output of first prompt

Here's the formatted BibTeX entry for the given information:

bibtex Copy

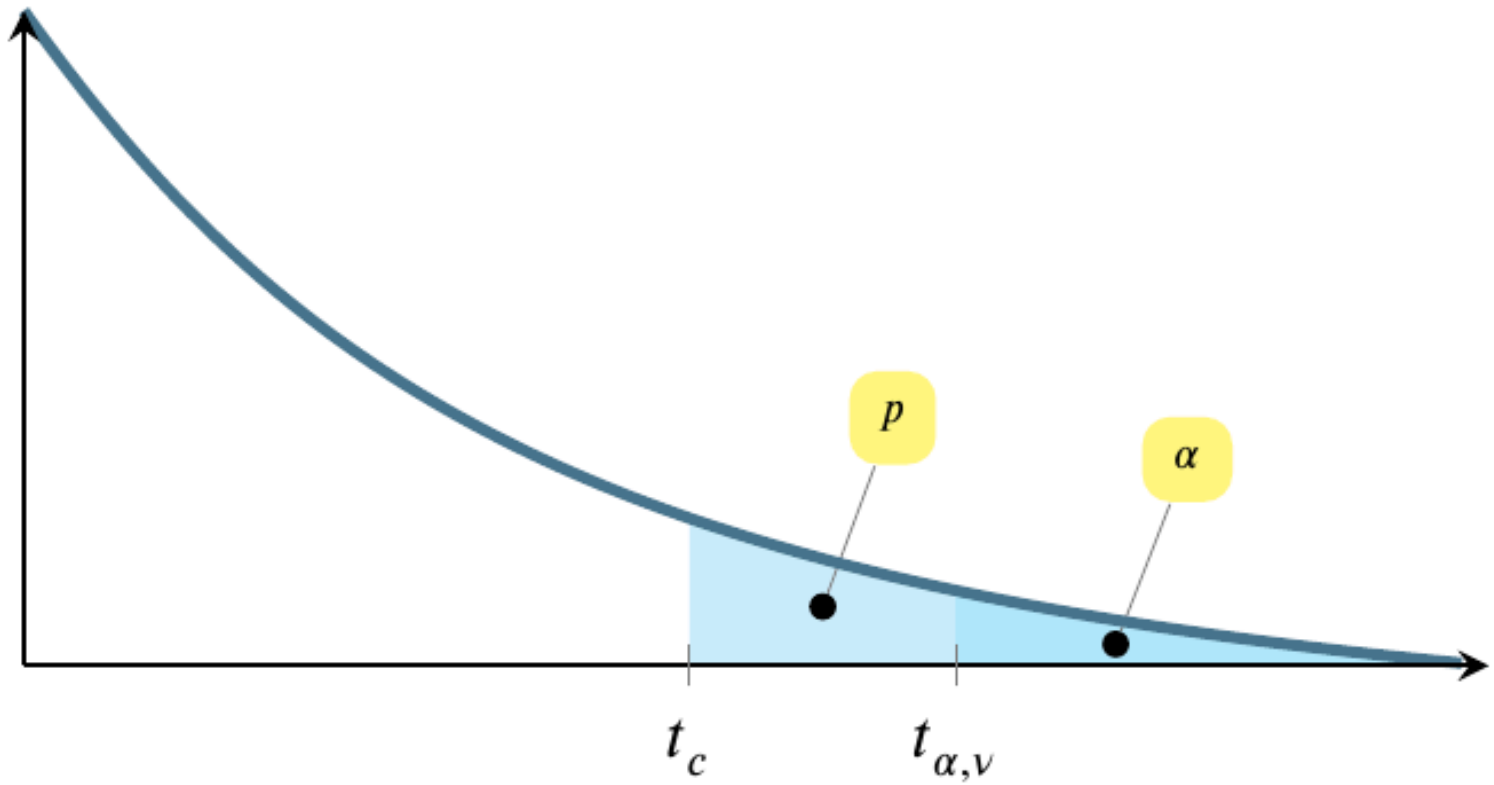
```bibtex
@inproceedings{tibdewal2018unlocking,
  author    = {Vinita Tibdewal and Sharoda Paul and Xiaomai
  title     = {Unlocking the User Experience of Generative A
  booktitle = {Honolulu '24: ACM CHI Conference on Human Fac
  year      = {2024},
  month     = {May 11--16},
  address   = {Honolulu, Hawai'i},
  publisher = {ACM},
  location  = {New York, NY, USA},
  pages     = {5},
}
```

Adjustments

- I have a key format I always use and tibdewal12018unlocking doesn't fit (but the AI could not have known that)
- The booktitle should begin CHI '24 (but the AI could probably not have known that)
- The address and location should be swapped (not sure if the AI should have been able to figure that out)
- Overall, I gave it an A and promptly used it for a long list of unformatted bibliographic items, which it handled similarly

Example—challenge

Animate this


p
α
t_c
$t_{\alpha,\nu}$

The preceding possibility shows the situation where $t_{\alpha,\nu} > t_c$, which is equivalent to saying that $p > \alpha$.

The x-axis scale begins to the left of the fragment shown here, and values of t increase from left to right. At the same time, the shaded regions decrease in size from left to right. Note that the entire shaded region above is p, while only the darker region at the right is α.

The challenge

I tried several times to get a genAI tool (can't remember if it was DALL-E or Midjourney or what) to animate this, only to experience repeated failures of many kinds. I thought of making this an exercise, in fact, because it seems like a difficult problem.

On the bright side, I have since learned that, instead of asking for a picture, I might be better off asking for a programmatic specification of a picture, allowing me to tweak the parameters. That's how I arrived at the preceding *no leverage* picture.

Unfortunately, it did not work in this case and the output is too involved to put on a slide. Instead, I will share my dialog with ChatGPT separately.

⟨ Pause to show dialog ⟩

The outcome

I created eighteen LaTeX files and about fifteen Python scripts trying to debug this but could never get even a rough output I could work with. I tried debugging both the Python scripts and the LaTeX input with no success. The challenge remains • • •

One additional problem is that I couldn't portray the issues in a presentation. I would need a video to go through the ChatGPT window. I'm reluctant to ask you to produce a video for your presentation but I will allow it if you prefer.

Presentations in Quarto

You'll need to turn in a .qmd file and a .html file by 11:59PM the night before your presentation. Optionally, you may also turn in an .mp4 file. (If you do that, we can run the .mp4 file in class instead of having you present.)

How do you do that? Just open R Studio and select File > New File > Quarto Presentation. The system will create a small presentation template for you. Now we'll add to it.

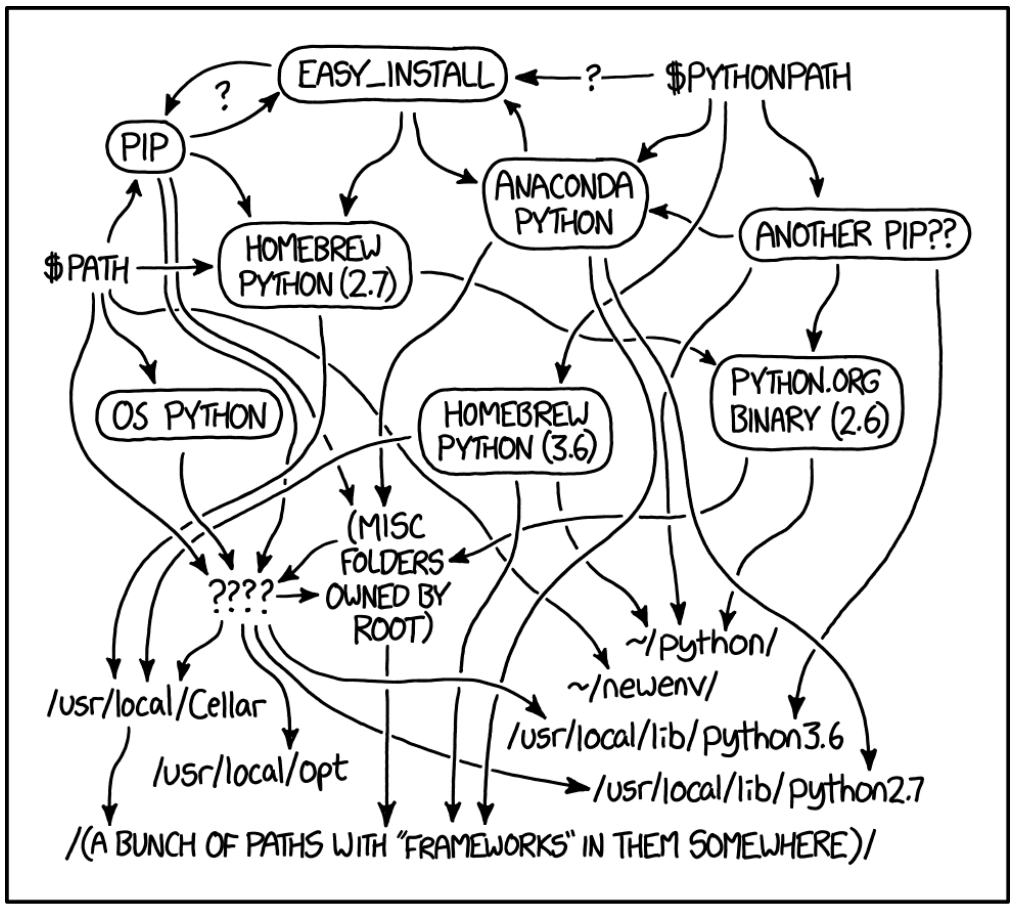
Why Quarto?

- Quarto is a tool that helps you do *reproducible research*
- Quarto helps you document the steps you took
- Quarto encapsulates the code (in this case, Python code) and your commentary about what you did, including any pictures, equations, videos, or anything else that a word processor can include
- Quarto provides a text that is readable by any text editor of your preference
- Quarto documents and slideshows can be rendered to html or any of about a dozen other formats

How to use Quarto

- Note that you don't have to follow these instructions if you already know how to use Quarto
- Be sure you have Python on your computer (how you do that is up to you; I use pyenv)
- Download R; then download R Studio (the order is important)
- Download the file eA.qmd from Canvas
- Open R studio; choose Open File from the menu; open eA.qmd
- eA.qmd is a template for exercise A
- Fill in your name
- Do the exercise in this file, documenting what you do as you go
- Intersperse code and commentary, including prompts you use

Problem?



MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED
THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

⟨ Pause to demo Quarto Presentations ⟩

quarto

Another exercise

Write a short Python program

- task: generate a random order for students to make presentations in this class
- constraint: everyone must present once
- constraint: no more than three may present on a given day
- constraint: no fewer than two may present on a given day
- constraint: there are twelve days on which students may present
- if no program works by the end of today's class, we will use the program I wrote and will show as a demo

Approaching this exercise

- You can make this very simple by doing some manual labor
- You need to break the problem down into parts
- You need to iterate
- You need to document what you do
- As with all exercises, I ask you to turn in a Quarto qmd file and an html file (both are required for credit)

END

References

Dhar, Vasant. 2024. “The Paradigm Shifts in Artificial Intelligence.” *Commun. ACM* 67 (11): 50–59. https://doi.org/10.1145/3664804.

Colophon

This slideshow was produced using quarto

Fonts are *Roboto Light*, *Roboto Bold*, and *JetBrains Mono Nerd Font*