Human AI Interaction

Lecture 3: Thinking in prompts and prioritizing
aidesignclass.org

The story so far

- Journey maps: A start-from-the-user method
- Tech matching: a start-from-tech method

Today:

- Understanding the material of large foundation models, and prioritizing user needs

What are LLMs, how do they work?

- An LLM is an LM - a "language model"
- A language model is a machine-learned simplified/partial representation of human language

Complete this sentence: Emory University is located in _____

Formalizing a bit…

P("Atlanta"| "Emory University is located in") >> P("Jupiter"| "Emory University is located in")

P("Atlanta"| "Emory University is located in") >> P("Earth"| "Emory University is located in")

P("Atlanta"| "Emory University is located in") >> P("the United States"| "Emory University is located in")

For y where P(y | "Emory University is located in") is large

~ loc("Emory University") = y

Similar real-world knowledge

"The average lifespan of a cat is ____"

"In the sentence 'The boy ordered some Coke because it was so hot', 'it' refers to ____"

 "Q: Why is it dangerous to add cold water to a saucepan of boiling oil? A: ___"

PLMs -> ILMs

PLMs are "pretrained language models", which are usually trained to predict the next word or "token".

Instruction tuned LMs (ILMs) go a step further. The take PLMs, and train them further on instructions.

E.g. "Write an essay on the civil war: ____"

Most LLMs you will use in this course are ILMs

ILMs are a better interface

- Easier to use: just say what you want
- Easier to "steer": Many ILMs now understand "Do not do X"
- Many are safety-tuned: they are less likely to say offensive things (but this tuning is not perfect. Be prepared for surprises!)

- Chat ILMs were originally meant for conversations. But they are surprisingly versatile. I suggest you use them as "default"

LLM terminology

LLM instructions = "prompts"

LLM outputs = "predictions" (but outputs is also in common lingo)

LLM temperature = "how consistent do I want my outputs to be?"

0 = deterministic

A few problems with LLMs today

- The output isn't always the same (with temperature > 0)
- The output (e.g. JSON) may not be well-formed
- The output might make up facts
- Small changes in the prompt might make large changes in the output
- LLMs aren't able to do symbolic / arithmetic logic well

Prognosis for problems

- The output isn't always the same (with temperature > 0)
    - Not always a problem?
- The output (e.g. JSON) may not be well-formed
    - Solvable
- The output might make up facts
    - Hard to solve but will get better
- Small changes in the prompt might make large changes in the output
    - Almost solved
- LLMs aren't able to do symbolic / arithmetic logic well
    - Possible to work around this

Designing an LLM experience (copy.ai)

Freestyle

What are you looking to create?

a LinkedIn post

What are the main points you want to cover?

[product details]
[fun facts]
[any instructions]

Choose a tone

🟡 Friendly

Create Copy

Freestyle

What are you looking to create?

a LinkedIn post

What are the main points you want to cover?

Topic: AI powered copywriting
Takeaway: AI is here to empower
you, not replace you

Choose a tone

🟡 Friendly ▼

Create Copy

---

and where it should be used, then sit back and watch the sales roll in!

🔲 Copy   ♡ Save   ⊕ More Like This   ⊗ Remove   x words / y charts

We've all heard the news about how AI can take over our jobs, but we're here to tell you that there's a lot more to this technology than most of us realize.

AI powered copywriting is one of those hidden gems that's going to make your life easier and more productive than ever before. It's a game changer for writers and marketers alike, because it gives you tools that help you do your job better and faster than ever before.

🔲 Copy   ♡ Save   ⊕ More Like This   ⊗ Remove   x words / y charts

AI is here to help you!
We know that there's a lot of uncertainty and probably some

Write a prompt

Freestyle

What are you looking to create?

a LinkedIn post

What are the main points you want to cover?

Topic: AI powered copywriting
Takeaway: AI is here to empower you, not replace you

Choose a tone

🙂 Friendly ▼

Create Copy

https://replicate.com/replicate/llama-2-70b-chat or try with Makersuite

After you play around a bit:

- Why is "tone" a dropdown?
- Do users expect the same response every time for given input?

Write a prompt

and where it should be used, then sit back and watch the sales roll in!

[Copy] [Save] [More Like This] [Remove] x words / y charts

We've all heard the news about how AI can take over our jobs, but we're here to tell you that there's a lot more to this technology than most of us realize.

AI powered copywriting is one of those hidden gems that's going to make your life easier and more productive than ever before. It's a game changer for writers and marketers alike, because it gives you tools that help you do your job better and faster than ever before.

[Copy] [Save] [More Like This] [Remove] x words / y charts

AI is here to help you!
We know that there's a lot of uncertainty and probably some

https://replicate.com/replicate/llama-2-70b-chat or try with Makersuite

(How do you implement "More like this"?)

Why split the task?

- Easier to teach the user what to do
- Greater human control
- Easier to prompt / more reliable

But what should we build?

- One simple technique is to prioritize with RICE
- R = reach (what fraction of your users would use this feature?)
- I = Impact (how much better would it make their lives?)
- E = Effort (How much effort would this take to build?)
- C = how confident are we in our values for R, I, E?

RICE = R*I*C/E (Higher is better)

But what should we build?

- One simple technique is to prioritize with RICE
- R = reach (what fraction of your users would use this feature?) - 1%, 10%, 50%, 100%
- I = Impact (how much better would it make their lives?) - 10%, 50%, 2x, 5x
- E = Effort (How much effort would this take to build?) - 1, 7, 30, 90, 180
- C = how confident are we in our values for R, I, E? - 10%... 90%

RICE = R*I*C/E (Higher is better)

Try it (for Gmail)

Features:

1. Suggest reply to email
2. Suggest rephrasing email to make it more professional
3. Suggest "quick actions" - "Accept invitation", "decline", "send thank you"