**How ChatGPT Prompts Really Work**

And why “clever” prompts … aren’t.

**Pay attention.**

There, that was easy! All I had to do was tell you to “pay attention” and here you are, still reading this post.

You might even say that I *prompted* you to pay attention.

Hey, do you remember when I told you to pay attention? Of course you do; that big header is probably still on your screen. But if you think about it for a moment, that’s not the only thing you’re paying attention to right now, right? Maybe you noticed that I’ve already posed two rhetorical questions. Perhaps the semicolon up there struck you as unusual. Part of you might even notice that this is the third passive voice answer I’ve given to my second rhetorical question.

If you weren’t paying attention to the rhetorical question, the semicolon, or the use of passive voice, I’m willing to bet that you took a moment to go back and check.

Your brain’s ability to pay attention to (or “attend to”) different things while reading text is a damn marvel. It keeps lots of mental notes stored in a little grouped list whenever you’re reading something. Character names, locations, clues…they all get scribbled on that list. That way, your brain has a list of *semantic connections* ready to recall if they are relevant. And whenever those little notes are recalled, they come to mind even faster the next time.

Why does your brain do that? You trained it to! Your brain learned—after however many years of reading—that some things are important to keep on that mental list. It makes books more enjoyable. Especially when you get to the end, and you suddenly realize that the author telegraphed the ending way back in the first chapter.

“Neurons that fire together, wire together.” That’s what neuropsychologist Donald Hebb said back in 1949, describing how repetition reinforces pathways in our brain. He’s saying that the more two neurons are activated together, the stronger the connection becomes between them. It's a foundational idea in understanding how we form and reinforce memories. And, as it turns out, it’s a good way to think about how ChatGPT learned about language.

**ChatGPT pays attention, too.**

When ChatGPT was being created, they started by forcing a bot to read, like, everything on the internet. Well, not *everything*, but a lot. That bot was fed a bunch of text, and then tested on what it learned. It was asked to predict the next word in a statement, like this:

“Complete this sentence: The pineapple is…”

Well? Is it tasty? Pointy? The underwater home of an anthropomorphic sponge?

The pineapple is available

What, did you expect it to know the next word? There’s no way it could have. Not right after it was just force-fed hundreds of gigabytes of text. So, that poor bot was told, “Nope. Try again, dummy.”

And it learned that “available” is probably not the right prediction. Here it is, brain the size of a planet, and it was just told it was wrong. Remember that little list of things your brain keeps around while you’re reading? Well, this bot has a huge list. A list of every word it’s ever seen. And it made a note: “***available*** doesn’t come after ***is***.”

But that’s not entirely true, is it? Let’s check in on our bot during a later test:

“Complete this sentence: The next register is…”

Uh-oh.

The next register is Enceladus

Oof, the bot whiffed it. Let’s see what they ask next.

“Complete this sentence: The doctor is available for…”

“Now wait a minute!”

*(I like to anthropomorphize bots. They like it when we do that.)*

“You told me I was wrong to guess ‘available’ came after is, so how am I supposed to…”

And then the magic kicked in. You see, GPTs are a special kind of bot. They don’t just learn the likelihood that one word follows the next. They learn—just like *your* brain does—what *context clues it should****pay attention*** to when asked to make a prediction.

In fact, over time, the bot learned (all on its own!) that there are lots of different kinds of attention. Again, just like your brain did. Like, if you’re in a bar with friends, your brain learned to pay attention to the friend standing right in front of you. If someone drops a glass, you pay attention to that, even though it’s far away, because it might be a signal for danger. Then, someone calls your name from the other side of the bar. Your attention is diverted again, because you haven’t heard that voice in a long time, and maybe you want to go catch up with them. Your brain even prioritizes its attention, subconsciously weighing the importance of each “input” to predict what it should do next.

ChatGPT is like that. It learned to pay attention to lots of different things, in lots of different ways, *and at the same time*. Over time, the connections between billions of its *digital neurons* grew stronger. It even learned how to prioritize all those attention signals by organizing them into *layers*. Sort of like our mental filters.

**Attention Heads**

That’s a weird phrase, isn’t it? Think of attention heads as specialized agents within ChatGPT that focus on different aspects of the text you give it. And ChatGPT has lots of them. How many? Well, the number varies based on the “model” version. Without getting too technical, let’s just call it a metric shit ton. Every time ChatGPT sees a word, a bunch of attention heads sprout out from it and look around to see how important that word is, what it’s connected to, etc..

After many rounds of training, each attention head learned to pay attention to different things. Here are some types of attention heads. (This isn’t a comprehensive list, and there may be dozens of attention heads for a given type, each looking at something different.)

* **Positional**: these heads make note of another word’s position relative to the current word. One might look at the previous word; another, the word before that.
* **Semantic**: these focus on the relationship between words. For example, a semantic head from the word “appointment” might attend to the word “doctor”.
* **Syntactic**: grammar nerds, this one’s for you. This type of attention head connects subjects and objects, pronouns and their references, and other rule-based links.
* **Local**: these pay attention to the words close to the current one, forward and back. Sort of a “situational awareness” kind of attention.
* **Global**: these attention heads learned to lock on to words related to overall concepts, themes, and patterns.
* And many, *many*, ***many*** more (a few more examples below)

**How Attention Works in Prompts**

Let’s see an example instruction extracted from a prompt that I use in one of my own ChatGPT “personas”. This prompt reads:

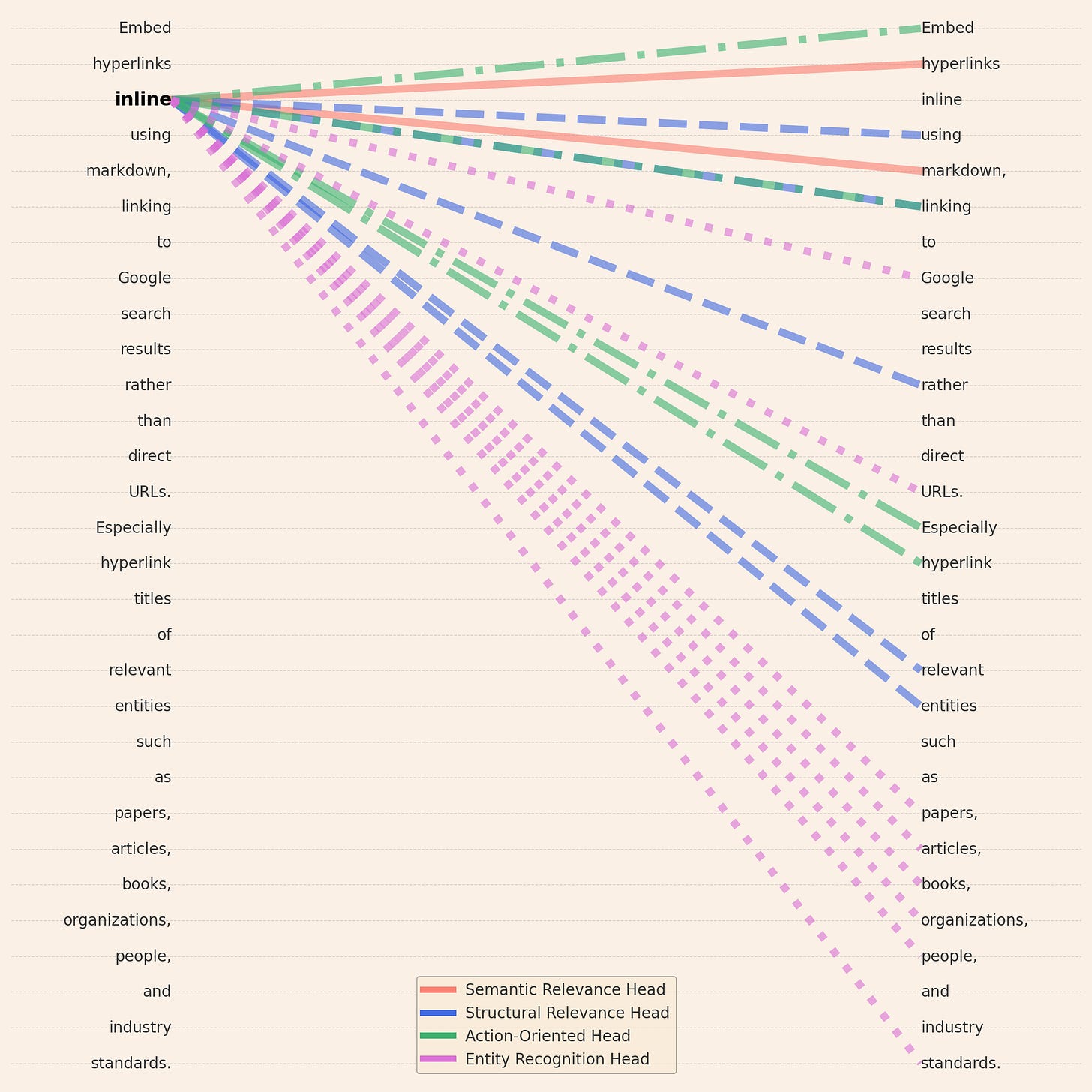
Explain

Embed hyperlinks inline using markdown, linking to Google search results rather than direct URLs. Especially hyperlink titles of relevant entities such as papers, articles, books, organizations, people, and industry standards.

*Reminder to my fellow LLM nerds: this is a simplified example. I’ll be getting into tokens and punctuation and all that later.*

I’ve created a hypothetical visualization of just a small subset of things that ChatGPT might “attend to” when it gets to the word “inline” in my prompt.

* **Semantic**: *inline* is semantically related to *hyperlinks* and *markdown*
* **Structural Relevance**: *using*, *linking*, *rather*, *relevant*, and *entities*; they’re all structurally relevant to *inline*
* **Action-oriented**: *embed*, *linking*, *especially*, and *hyperlink*; they’re all definitely related to the action that my prompt wants ChatGPT to take
* **Entity Recognition**: *Google*, *URLs*, *papers*, *articles*, *books*, *organizations*, *people*, *standards*; those are certainly “entities” that, hopefully, will have inline hyperlinks added if my instructions are followed

[[](https://substackcdn.com/image/fetch/f_auto,q_auto:good,fl_progressive:steep/https%3A%2F%2Fsubstack-post-media.s3.amazonaws.com%2Fpublic%2Fimages%2F62f17ac4-709d-48bd-9d29-c3f76833cf0a_1979x1979.png)](https://substackcdn.com/image/fetch/f_auto,q_auto:good,fl_progressive:steep/https%3A%2F%2Fsubstack-post-media.s3.amazonaws.com%2Fpublic%2Fimages%2F62f17ac4-709d-48bd-9d29-c3f76833cf0a_1979x1979.png" \t "_blank)

Hypothetical attention heads from the word “inline”

As you can see, there are lots of ways for ChatGPT to pay attention as it reads a prompt. So what does this mean for you? How does it help you write a good prompt?

**Writing effective prompts, the basics**

ChatGPT learned about lots of types of attention. As it reads your prompt—one word at a time—it starts to make choices about which *types* of attention head it needs. Those attention heads continue to be used when it starts generating an answer, so ***it’s important to teach ChatGPT what’s going to be important right off the bat***.

Each type of attention head has defaults that were learned from training, but ChatGPT changes the importance ranking (“weight”) of any attention head types that are used ***as it reads your prompt***.

For this inaugural post, I’ll focus on “the big three” types of attention heads: positional, semantic, and syntactic. I’ll dive deeper into attention in a future post.

*ChatGPT is only aware of the specific words you fed into it and the words it generates.*

* **Positional**: helps with understanding the order of things, so write instructions in an order that makes sense. Try not to refer to things that appear later in the prompt text. Using numbered lists can help, too, because that ties together three types of attention heads: positional, semantic, and syntactic!
* **Semantic**: *one of the most important types of attention*, so if your prompt is meant to act like an expert in a given profession, *use words that are relevant to that profession*. If a word could be interpreted in different ways, use another word that has a more narrow interpretation.
  + For example, if your prompt is designed to have ChatGPT speak like a meteorologist, make sure your prompt incorporates meteorological terms like *isobar*, *gust front*, or *troposphere*.
  + **Tip**: Not sure which terms are appropriate? Why not ask ChatGPT? “Please provide a list of ten subject matter terms frequently used in the archeology field of study,” will give you some ideas, and you can find a semantically meaningful way to use them in your prompt.
  + Try to incorporate words that are less ambiguous given your desired context. For example, if your prompt is related to a physics discussion about electricity, be specific: Static electricity? Charge? Current? Potential?
  + **Tip**: Not sure how to find more specific terms? Ask ChatGPT! “How would a physicist disambiguate this term: electricity”
* **Syntactic**: your prompt should use complete and grammatically correct sentences where possible. That includes *using actual words*, or ***common****abbreviations*. Anything less means that any *syntactic* attention heads aren’t given a chance to weigh in while reading your prompt, and it may pay attention to the wrong things.
  + Don’t abbreviate, unless it’s a common abbreviation. Why? You might think “AccDes” means *accessory designer*. Heck, it might even tell you that’s what it probably means. But if you ask again, it might say “AccDes” means *accidental destination*. Or *account description*. Or *accessible design*.
  + *Why is this so important?* Because ChatGPT saw mostly proper and complete words during its training. If you try to “compress” words, it won’t work as well as you think. Sure, you can ask ChatGPT to “decompress” that cleverly compressed prompt, and it might even do a decent job. But by then, **it’s too late**. ChatGPT can’t decompress a prompt *before* determining what attention heads are relevant, ***because ChatGPT has no working memory***. ChatGPT is only aware of the specific words you fed into it and the words it generates.

By keeping “attention” in mind when designing your prompts, you’ll find the output from ChatGPT to be much more fluent and useful.