Welcome back to the next lesson in Module 3: An Introduction to Prompt Engineering.

This time we are going to bust prompt engineering myths. So there are three in particular that we're going to go over and make me pull my hair out sometimes when I see them posted about on Twitter or elsewhere. At the very least, all three of these are not useful if you want to get better at prompt engineering and in some cases actually counterproductive. So we're going to go through these step by step, explain why they're incorrect and how you can think differently about using the technology.

So the first is that ChatGPT is just like Google search. This is just completely false. There are really only two reasons why this statement could be true. One is that right now the primary interface to interact with a Large Language Model is a text box, right? So you're typing some kind of input to get a result. I guess Google search more or less works the same way. So that's understandable that people would think that. We've been trained over the last 20 years to use Google as the gateway to the internet. That makes sense.

And two is on the other side of that search is in some sense the collective knowledge that's been digitized. But in the case of Google, it's more like going to an archive or a library and asking a librarian to pull a bunch of documents from the stacks or the archive. And then arrange them in such a way that I just hand them to you for you to go through and interpret, decide what you want to read.

Whereas ChatGPT Large Language Models, they're doing even something more sophisticated where they're synthesizing information into something that's a lot more personalized. It's a higher level of abstraction and has these different properties that make it very different from using Google search.

So let's dig into what I mean specifically by this.ChatGPT is like Google search. Absolutely not true. Using AI and certainly getting good at prompt engineering is not like using a search engine or getting good at using a search engine like Google.

Why is this the case? Three reasons. The first is when it comes to using AI and doing prompt engineering, more is more. So when you're using Google, you want to fit as few words as you can into your query to try to get the results in the search index. In the search engine result page, each one you can assess what you want to go into and then decide you actually read the source material yourself.

Not what you want to do with ChatGPT. You actually want to provide as much information and as much context as possible. I'll show you some examples. We'll talk about this later. But in particular, if there's one thing that I recommend to every single person who I work with when it comes to using AI, one small simple thing that you can do to become an excellent prompt engineer and just get literally an order of magnitude more value out of these tools immediately, it is to stop typing.

Don't type anymore. Use dictation. The software has gotten quite good. Both Windows and Mac have their own native dictation capabilities. They will get better. They still leave a little something to be desired. On the iOS and Android apps for ChatGPT uses for dictation. This model called Whisper, which is really state of the art. Meaning it's super, super accurate. Like 99.9% accurate. You can mumble. You can talk fast. You can mutter under your breath. And it'll still capture most of what you say, even in a noisy environment. So it's very good.

And then there's a tool that I'm a huge fan of on Mac called Super Whisper, which uses the same model, but it's on your MacBook computer.

Anyway, that technology will get commoditized. Voice will become a more standard way to interface with your computer. And if you can get ahead of that and try to cultivate the habit of using dictation and just speaking naturally rather than typing, you're going to get so much more out of these tools. It's not even funny.

If you just do the math, you can speak three times the speed that the fastest person can type. Let's say two to three times. Two, if we're being conservative. And I'll say it took me about six months to really cultivate the habit of using dictation primarily, not typing anymore and my productivity, subjectively, the way I feel when I'm using computers, it's a lot more natural.

And the quality of the responses that I get from the Large Language Models is so much better because I start every conversation by providing a huge amount of context. And so you can do that in natural language.

You can also copy and paste stuff. One of my favorite things to do is to take the entire transcript of a meeting that you just had and then ask ChatGPT for feedback. With Google, you'd say how to get better at sales or something or sales coaching. Maybe that's the only query you put in and you have to look through this and it's all kind of standard issue stuff. Maybe you say "sales coaching for software as a service companies" or something like that. A little bit more specific.

But with ChatGPT or any of these other tools, you can literally take your entire transcript. Feed it in and then say, "I want you to give me brutal feedback, highlight my blind spots and help me with high levels of specificity, think differently about how I'm interacting with prospects to improve my game as a salesperson or a business development person or whatever you're doing."

You absolutely can't do that with Google. You can do that with ChatGPT, but it requires a lot of input. So more is more. More context means higher quality and more personalization. So more is more leads to personalization, which we just described, but also discursive interactions.

This idea of AI as a discursive medium, meaning you go back and forth. Say you have a conversation with a friend or a colleague, hopefully you're not just broadcasting, talking at them. You're engaging with them in this relational context where you go, you put forward an idea and then maybe they ask you a question in return and you go, "Oh, that's interesting. But what about this?" And so it's this back and forth set of interactions where overall quality of the information and the ideas that are emerging over the course of the conversation. The output, so to speak, of that interaction is getting better and better rather than just a one-off input/output.

We talked about this in a previous lesson around the evolution of human computer interaction. Using AI, like I said, it's more discursive. It's relational. You're going back and forth, which means that you can get somewhere that you wouldn't be able to get to. It's just like a one-off short Google query.

And finally, the Large Language Models allow you to sprinkle some reasoning capabilities on top of the information that's returned. And so an example of this would be, I worked with a client where I worked with the executive team over a course of a quarter. At the end of the quarter, we're doing some reflection on how things went. And the question came up, "Okay, first of all, how did we come to the conclusion that we came to in terms of setting particular goals?" So I went back to, we did like this five-hour workshop at the beginning and took the transcript, fed it through AI, and then just asked it to retrieve that information. And it said, "Okay, this person said this. And these people have these concerns, but we decided ultimately that to choose this metric because of whatever." It was like, okay, that's like standard information retrieval.

Where things get really interesting with these models and Google search obviously can't do this is you can start to ask it as if questions or technically these would be called counterfactuals. And so I put forward the question behalf of the group. If we had not gone with that particular metric for the quarter in terms of goal setting, what might have been alternative metrics that could have been even better? And why wouldn't that have been the case? So to explain its reasoning.

And it came up with some great answers that really simulated conversation with people. Yeah, that's really a good point. And we can think about this process differently in strategy, implications and all this kind of stuff.

So ChatGPT, Large Language Models, AI, absolutely not like Google search. I understand why people think this is the right pattern to adopt when using these technologies, especially for the first time.

And it makes me very sad when I see somebody sit down, they have access to GPT 4, any of these tools, and they type one or two words in and what they get back is just totally generic and not useful. And it's because you need to think completely differently about how you use the tools. So it's not like Google search. Put that out of your head. In fact, do the opposite. More is more.

Okay. The next one. This one just drives me nuts. You see this kind of stuff on social media all the time. This is just like clickbait. Twitter thread. Nonsense, right? A hundred prompts they don't want you to know. The implication is, and this is around the topic of prompt engineering in particular, the way that some people will talk about this is, oh, I have these templates or I have these almost like magical incantations that allow me to use the model so effectively and this is a secret that I have and all this sort of stuff.

On one hand, there is a deep truth there, which is if with high levels of precision, specificity, contextual information. There are extremely specific prompts that you come up with that end up getting the result that you actually want.

The idea that there are these general purpose prompts that are simply going to just work better than what you would provide if you're just trying to tackle the problem that's in front of your face is mostly untrue. There are some general tools that we'll talk about in terms of simulating critical thinking and asking the model to ask you questions and things like that, that you can incorporate into your prompts or a certain little engineering tips.

But here's the thing. Even spending a minute paying attention to that comes at the expense of you just getting hands on with the tools and figuring out how to do it yourself. So prompt engineering is not about memorizing these esoteric incantations, right? There's no spell book that you go to and you read like a grimoire or something. And you're like, "Ooh, I have all the answers." Maybe you go there for a little bit of inspiration and you see what other people have done. But at the end of the day, like if you want to get good at this, or you want to actually get the most out of AI, you need to really focus on the problem that you're trying to solve, the result that you're trying to get, and the context of the real world situation in which you find yourself.

And this is like a deep human thing to do, is we want to reach for kind of an easy answer or a heuristic or a playbook or something like that. And I'm not saying that's all useless, but I will say it's mostly useless in this case. Just because the degree of personalization that you can attain with Large Language Models, if you're really focused on giving it the right inputs, is so much greater than any previous IT wave.

The other thing is, even for the things that do work, so you may have seen there's like on social media, especially every once in a while, there's some new research paper that comes out that comes up with some interesting new technique for improving the quality of simulated reasoning or something like that. So we've seen chain of thought. This is a common pattern you see. Oh, let's think this through step by step to get to the right answer. And that forces the model to break down the problem into steps, assess each one, think it through in some sense, spend more compute doing that. That can, not always, but can improve the quality of certain outputs. You got chain of thought.

Got another one called tree of thought where we're going to have different branches we're going to consider. And we're going to go down and we're going to do this sort of tree search and figure out what are the conclusions we come to and take a vote or use simulated critical thinking to validate the incremental result versus the odds that you're getting to the result that you want.

Those are all incredible discoveries and I think are super interesting. What I'll say practically is, first of all, a lot of those are very expensive, not just in terms of paying for the models or anything like that, but also just in terms of like your time. If you want to implement some of those, first of all, you're not guaranteed to get better results.

And second of all, it increases the complexity of the system that you're building with multiple prompts you want to work in. Use more tokens. Results are not guaranteed. But the more important point is this technology is evolving so quickly that by the time you would put one of these systems into production or something that leverages all these techniques, it's most likely going to be commoditized.

And there's some speculation around the architecture of GPT 4. This is unconfirmed, but it's probably already doing some of this where there's silent thinking on behind the scenes. And a lot of these little techniques will probably just get absorbed back into the foundation. The foundation models or the API-based services that take inputs and provide inputs.

Even if it does provide a little boost, it's not for long. So spend less time focused on these things and getting distracted by carnival barkers talking about prompt engineering on social media. And just get your hands dirty. Okay. Figure it out for yourself what works in your situation.

That leads me to the last point. This is just will always be true. No one knows your job or goal better than you. It may not mean that you're equipped to solve it better than everybody else, obviously. That's why we hire people or work with other folks who have expertise that we don't have, but you know what it is or you can figure out what it is that you need to do. And then call on AI to be helpful in getting you there or explaining certain concepts that you may not be familiar with. Or whatever else has been on ramp to get closer and closer to your goal.

Just like we talked about last time, you pick point B. You try to sprint towards point B and you're probably going to be significantly better off if instead searching around and paying attention to prompt editor and gurus or whatever or signing up for 50M newsletters about AI and trying to figure out like the magical incantations and secrets.

Use the scientific method, okay? You have a result you want to get. You have a hypothesis. Test it. Make contact with reality. See if it's good enough. Does it move the needle on the thing you want to move? Learn from that, and just keep doing that until you hit a wall. And if you hit a wall, then maybe escalate it to an expert, but that's going to be such a better use of time.

On top of that too, I'll just say one other sort of corollary on this is, I've even seen this in some of the foundation model provider's documentation. I'll share my subjective experience on this. When it comes to formatting, I've seen people get super up in arms around using like markdown and all these signs and symbols or whatever. Typing things. It has to be in bullet points and all this kind of stuff when you give a prompt.

If you want something in a certain output like JSON or YAML or something like that, specifying that is super important, giving some examples, whatever. But the rest of it is, in my opinion, just almost complete nonsense. If you're better served by really thinking through the problem, speaking it through a natural language, sometimes rambling for 5, 10, 15 minutes.

If the model can handle that many tokens, get all the information out there. Copy and paste other stuff from relevant documents and just stuff it in. Okay. Don't worry about structure that much. The structure is more for you to remember, "Oh yeah, I did this and put this in. And okay, here's like a headline and all caps or something like that."

Whether you put that in all caps or you put that with a hashtag, octothorpe symbol or something at the beginning, maybe it makes some difference. But it's super incremental and negligible versus just thinking through the problem, getting the good information out there in the first place. So don't get up in arms over formatting and use XML tags and all this kind of stuff.

That's mostly a distraction. Maybe it's useful. Again at the end if you need to break things out just so you can keep tabs on how the prompts are structured, but don't worry too much about it. More is more. Use a scientific method. Test things out. And understand that the underlying. All those little tricks are going to be commoditized so fast it'll make your head spin. Probably by the time you're watching this video, a lot of them already have.

All right, now the third myth. And this one's funny because it is in some sense horseshoe shaped. I'll explain what I mean by that in a second. AI is either going to take or do your job. So the robots are either coming for your job and you're not going to have a job. Or the robots are going to do your job and you don't have to work anymore.

Neither of these things are true. And what's especially funny about this is although those two conclusions seem to lead you to completely mutually exclusive, incompatible future states, the implications for how you act in the present, if you believe one or the other to be true are basically the same, which is a little strange.

And what I mean by that is if you think the robots are going to take your job and there's nothing you can do about it, or the robots are going to do all work and you don't have to work again and you can just kind of sit back and enjoy the ride or whatever, you're not going to do anything with AI.

You're really, yep, this is out of my control and I'm just not going to do anything here. Huge mistake. Not true. I can't predict the future. Maybe in some long run sense, this may be the case. I'm highly skeptical. It's certainly not true in the short term and that is do exist in the present. So you have to make decisions now and that really be in reality.

Like I said, this horseshoe-shaped, it ultimately ends up in the same place. You not using AI. And ironically, actually, although yes, you can delegate a lot of tasks to AI. And so there are ways in which it can handle things that maybe you did handle before without computation, if you're going to do that, it actually forces you to then think at a much higher level.

And the only way to continue really getting value out of the technology is to continue to push forward, think harder and better. So you're using higher level reasoning capabilities. You're thinking much further into the future, trying to play things out. You're considering lots of different perspectives.

You have more attention that you can put into building meaningful relationships. You're going to have to orchestrate all these different pieces and parts and components rather than just focusing on doing the same thing day in and day out. And you get the benefit from it. Not only are you going to, it's not true that you're not going to work and robots are going to do everything for you but you're actually going to have to like work harder and better. The quality and the level of thinking you have to engage in really goes up. And that's a fantastic thing because as an entrepreneur, when you create outsized value, you can also capture a portion of it and grow your business and accumulate capital, which you can put into your projects, whatever else.

But iif you can delegate certain jobs or certainly tasks to AI, you're then going to have to work even harder and move forward on the frontier, become more of an expert. And on that point the ultimate paradox in all this is that if you use AI and you become a master prompt engineer, you will inevitably get to a point where you are dealing with ideas, with information, with experimental data generated from the things that you're doing in reality.

When you're running, let's say entrepreneurial experiments or whatever, doing new things, you're going to venture into pockets of reality are not accounted for in the model. And that's good. Actually, that's the point that you want to get to. And what's a little tricky about that is at that point when you really are super far out on it, AI may not only be not super useful, even as a thought partner, may actually be counterproductive because it might then try to pull you back to its knowledge base. Relate everything you're doing back to a reversion to the mean sort of thing.

It's not always the case. Sometimes it'll work with you. If you find these interesting connections between disparate topics, it'll say, "Oh, that's interesting." And maybe you should think about here's this thinker you've never heard about, all that kind of stuff. But ultimately if you're using AI correctly, it'll make such forward progress in your field, in your business, whatever else that you're going to end up somewhere that the model has never been, or it's never seen data from.

And at that point, you're best off actually just not using AI in some cases at all. And you have to, as I like to say, think naked. So again, robots are not, probably not going to take your job, probably not going to do all your work for you. And even if you use it correctly and you do delegate effectively and so on and so forth, you should expect you have to think harder and better.

And the goal is to ultimately get to a point where you're not using AI at all if you really are on the frontier. It's a different way of thinking about this, but I think it's much closer to the truth. And this has been my experience, I think will continue to be quite true.

All right. Hopefully that helps you understand how to separate fact from fiction, bust some of these myths about prompt engineering, AI, and LLMs more broadly, and arm you with a slightly different way of thinking that will endure the subsequent developments in the evolution of the technology.

Of course, feel free to ask any questions. And I will see you in the next lesson.