Welcome to the next lesson in this module, which is An Introduction to Prompt Engineering.

We're going to talk about how to actually initiate the conversation with your computer. How do you kick things off so you can get the most out of AI and using Large Language Models to accomplish your goals?

Same question. How does one kick off a conversation with the computer? It's kind of simple and not unlike talking to a human, which is good. But also means you want to put a little thought into what you're saying, of course. So when it comes to inputs and initiating the conversation, there are really two components that you want to think about.

So one is how you frame the conversation overall. And this is not unlike if you put an event on somebody's calendar or you invite people to an event. You want to think it through in terms of, so you're using everybody's time effectively. What's the objective? What's the agenda? What's the contextual information that is relevant and will be relevant throughout the course of the conversation?

All these things, right? So that somebody gets an invite. They understand what it is that they're signing up for, why they're there. There's not confusion when you start the conversation. It doesn't have to be a whole lot of clarifying questions that get asked and answered. All that kind of stuff. It's well done, okay? And if you're operating at the executive level, of course, this kind of thing is or at least should be table stakes.

So similarly, when you initiate a conversation that you want to be fruitful and get good results from the computer, you're doing prompt engineering. It's very important that you provide context.

What that means in practice, we'll break this down in a second, but just understand conceptually, that's the case. And then what you want to provide as a compliment to that are specific inputs that you want the team to work with, right?

So let's say for example, you're doing a quarterly objective setting exercise. There is a certain structure that you can follow no matter what, that you want to tell people, “Okay, here's the outcome that we're trying to achieve. Here are the outputs that we're going to create. Here's the agenda. We're going to break this down. Do it this way.”

And that's like the framing, or what's called the system prompt, if you're using the OpenAI models. It's what they call the system prompt. It persists throughout the course of the conversation, But then you also want to provide inputs.

If people are doing an objective setting exercise, they of course have a lot of information in their heads that they're bringing to the table. But then you also might want to make sure that everybody sees the financial statements for the company, sees the results of a survey that you sent to the company to gather information around employee morale or whatever.

Anyway, those are inputs that are going to be used to actually create and synthesize something by the end of the meeting. Working with a model, same kind of thing. So. You want to frame it. You want to provide relevant inputs. And then ultimately you get outputs. So it does the synthesis and the thinking. I'm going to use that term very loosely. But it analyzes and transforms those inputs into something that is close to the result that you want to get. But I've talked about previously, it's not a one-off interaction. This is something that is ongoing and discursive. There's a feedback loop where you will provide it feedback. You may ask the model to assess its results and go through another loop, and that would be called a recursion, technically.

And then there's another technique called retrieval augmented generation, which, again, is just a fancy way of saying you're providing additional contextual information on an interaction by interaction basis when appropriate. So you have the context at the beginning, but then maybe you get a response. An example of this would be like crafting emails for a marketing campaign or something like that. You get the content. Here's the outline of the emails. “Oh, we like this. Okay, now let's move to the next one. For this first one, I want these five points to be covered. Cool.”

It does the draft. You give it a few rounds of feedback. The content's great. And then what you do is you're ready to now get this into the right tone of voice of the sender, and so you find examples of other emails that were really good in the tone of the person, maybe it's you or you're writing on behalf of somebody else. And you pull that into the interaction and you show those examples at that point where you're ready to, you know, sprinkle like the tonality or the voice on top of the emails.

And that would be called retrieval augmented generation. It has additional contextual information that's important. That's basically it. That's how you talk to the computer and you kick it off by giving it the relevant information, having a good idea of what your goal is.

Now the key ingredients, we're going to break each one of these down, is they are the following.

Some are more important than others. They're not all mandatory, but the more of these you have flushed out, the higher quality, the results you're going to get. And the faster you're going to get to your goal.

Okay, so the first one is your goal. Talked about this a million times. I'm going to keep saying it again. Have a good idea of what it is you want to achieve. What is the problem that you're trying to solve? What is the result that you want to get to? And potentially, what are the outputs that you want to get that result? If you have clarity around it, okay. Gt focused on that. Have some clarity around that.

Two is a context. So what is all the relevant contextual information? If you are, for example, drafting CRM notes, you have a certain template that the company uses with specific definitions or an SOP or something like that, just copy and paste that in. Stuff it in and say, “Okay, here's our SOP, right? For reference, it's got to fit this format.” Okay, that's good context. And it's also, that bleeds into instructions a little bit, but it's also contextual information. And then, just like you would explain something to a human, if there are caveats about the situation, or there's other stuff that it should keep in mind, you know, bring that in.

Alright, so then you have a persona. This one is somewhat important. It's definitely becoming less important. I'll explain what this is. Like, you instantiate a person who you would consider to be the best, has the qualities of the person who would be the best individual to handle this task that you're trying to get done. And I'll talk about this in a second more.

So I'm going to skip over that. Instructions. What are the different steps that it needs to follow if you have that codified? And that's clear. Examples. So gold standard examples of other outputs of a similar type are very high quality and close to what you want. A hyper growth mindset. This is actually, I think maybe next to the goal, the most important ingredient. I'll break that down at the end.

And finally, this one's interesting, not always needed, but can be useful, prompts to stimulate reasoning. Simulate, stimulate, and I'll explain the little joke there in a second, but that's the last ingredient.

So you have these seven ingredients. Some are more important than others, but that's what you need to kick off a conversation, have a fruitful conversation.

Okay, your goal. We've talked about this quite a bit. Point B, where are you trying to get to, or at the very least, what is the general direction that you're going in? I'm trying to accomplish X, Y, Z. I want you to work with me to take a transcript of conversation that I just had and help me turn it into a LinkedIn. I don't know, but you know what it is you're trying to do.

Contextual information. What are the pertinent facts? Always more is more. Eventually there are limits to how much information you can stuff in. They're growing and growing and growing. And it's being able to put more context, more data into the conversation will become possible. Claude from Anthropic, I think is, has the biggest context window. You can stuff something like, it's something crazy. I think 50 or 60,000 words worth of stuff, which is like a short novel into each conversation when you kick it off. It'll hold it in its memory and provide you with outputs based on that.

Okay. In the case of what I described about the LinkedIn post, it would be the transcript of conversation that you had. And maybe some details around what it is you're trying to achieve or who you want to reach. Right? So just provide as much as you can. Make sure it's relevant, of course.

And then persona. The general take on this is you say, “Okay, you are an expert, whatever with an IQ of one 20. Even by the time you watch this may have already been commoditized, it'll likely just be like built into the next version of these models that it'll know which expertise to pull in that's going to be best for the job.

And anecdotally, it's one of these things where maybe it increases the quality of the answers. I'm not convinced that this is necessarily the case. I wonder if it's a bit, but. Like placebo effect, but for you as a prompt engineer, like in at least having to think about this, it's almost helpful for you to frame the problem.

So thinking about who you would delegate this to, if you had a human that you were going to delegate this task to. Not a bad exercise. I'm just not convinced that it changes the quality of the results all that much. But you can do this, right? You can instantiate the persona. You're an expert LinkedIn copywriter whose stuff gets clicks every time with an IQ of 120.

Instructions, very clear direction. Some secret sauce is to work with an expert and just record the conversation and ask them in very plain language role play where they're the expert. You're their new analyst or their new hire, and just have them explain. How they do a particular thing end to end? And keep going and asking questions until you really feel like you understand what it is they do. And, then just copy and paste the transcript of the recording and conversation into the prompt. Works really well.

If you already have SOPs or written documentation or something like that, you can use that, but that's it. You provide clear and precise instructions. If you can provide gold standard examples. You want the data or the outputs to be high quality. So good examples of what you're looking for a good real case of this would be like for lead scoring.

So if you have target prospects and you have all this data, let's say you have a giant spreadsheet and you're creating something that's going to interpret all the different fields. I don't know somebody's LinkedIn profile that you got or some data that you have or whatever it is. In addition to saying, “Oh, we're looking for people who fit this profile and this persona and all this kind of stuff,” you can also give the model a bunch of examples where you say, “Actually, these people have been our best customers. These people actually look like the ideal persona.” You could even give the reasons if you want, but you just say, “These are examples of people who are not a good fit” and get positive negative examples. Whatever it is, but it's high quality, like it's pretty clear that those are or not the outputs that you want. Just give that to the model as well as like additional information.

If you're not sure if it's high quality, it's not going to be quite as useful. But yeah, like vetted gold standard examples of what you are looking for, or you can even provide negative ones that'll get you quite a bit of a boost in terms of usefulness of the model.

And then this one's so important. This is not technical at all. This is like a hundred percent just human mindset stuff. If you want to get the most out of talking to your computer and certainly initiating the conversation, you want to have a hyper growth mindset. And what I mean by this is it's this paradoxical combination of two things.

One is an extremely ambitious goal, right? It's by definition, as I was saying before, you're trying to get to point 10 times faster than you've ever gotten there. And a lot of cases that would sound crazy. It would sound impossible. I think we can do this when you have the extreme constraints and the ambitious goal. Okay. Put yourself in a bit of a pressure cooker and see how far you can push it. And then on the other hand, you want to cultivate this complimentary radical humility. Really being grounded, but also very open to the idea that especially if you're an expert, the model may actually come up with solutions, ways of doing things, ways of even thinking about the problem that could be superior or it could be more effective, certainly more efficient than you in terms of getting the result.

I'll tell you a very quick story about this. There was an interesting study came out a couple months ago about radiologists and it looked at AI. Radiologists plus AI, and how accurate they were in doing some sort of clinical scoring or assessment or something like that. I forget. I forget exactly what the task was.

These are highly trained people and counterintuitively what happened was the radiologist plus AI group, which you would have thought would have been the most accurate or the most effective. They ended up, I believe, being the least accurate of all three. And some of the speculation as to why that was the case, cause that's a very counterintuitive finding, was that when they saw something that conflicted with their human judgment in terms of the AI making an assessment of some kind, they immediately discounted it and were just like, nope. Not the case.

Totally understandable. If you're an expert, it makes sense. You're going to have these snap judgments and decisions if you see something that conflicts with the model that can't be right. However, it is entirely possible, and this will increasingly be the case, that even if you've gone really deep into a particular topic. Like, the model may be able to perceive things that you cannot that allow it to make better predictions, even in your field of expertise.

Now, the question of how to discern when to accept or reject the judgment of the AI on these things, incredibly difficult problem. I think it's a big open question. The way that you can potentially handle that is to take a big breath, step back and really work with the AI to understand, “Oh, why did you come to this conclusion?”

Just have a conversation with it. Literally ask it for its reasoning. Why do you think this is the case? You provide your own thing and just have a little bit of an open mind, but that's something to keep in mind. So hyper-growth mindset, ambitious goal, radical humility, those two things together. You're going to be in good shape.

Okay. And then this is optional, but good is prompts to simulate or stimulate reasoning. We've talked about this previously. The AI as a medium, Large Language Models are a discursive medium. So you're talking back and forth. And just if you're having a good conversation with somebody as a thought partner, or to try to get to some answer, there may be times where you ask the other person critical questions, or you ask them to really hit you hard with critical questions, or whatever it is.

And as you enter into this mode where it's moving from like direction to working with it in the stack partnership, which is a little bit more advanced in terms of prompt engineering, you can ask it to work with you in particular ways. Minimum amount of this is once it gives you an output or a result, given your instructions and your prompt and whatever, you can provide your comprehensive feedback.

So read through the result and just say exactly what you think. Try to be relatively specific, clear, concise. Don't hedge. This is something I always think is funny is most people are just like kind people. And if you're giving a human feedback, you want to make sure that it's positive stuff. You say, “Oh, here's some work for improvement.” All that. And the thing is when working with these models, you just want to be super direct, like very precise. Exactly the feedback on what's good, what's not, and why, and whatever else. At least I don't think you're going to hurt its feelings, but it definitely gets better results if you're very specific.

So give comprehensive feedback. You can also ask it to reflect upon its own answer, which is a wild thing. So for it to apply some self reflection and critical thinking, and you can say, “Okay, given your output and the original task that I gave you, I want you to self-reflect, find room for improvement. And then actually come up with suggestions on how you can improve your output.” If it's again, like a LinkedIn post or an email or something, these are easy examples, and you can say, “Oh, I actually like points one, three, and five. Why don't you do another draft with that?” That's pretty cool.

Talked about this one before this is a magical incantation, but you say, let's take a big breath. This came out of a paper recently and it's just hilarious. I'm not sure why this works, but for one of the Google models, apparently it improves quality pretty consistently. But okay, but the point is the thinking it through step by step, you can ask the model to break down the problem, which of course you can do that as well. Thinking through how to get more incremental results to get to the final result that you want to get to, but you can ask it to do that. And it'll say, “Okay, I need to do this and this and this. So let me get this right. And then that can go this and this,” and that may improve the quality.

And then this one's a pretty interesting one. The model may make assumptions or try to give you results with incomplete information. And you can actually ask it to be a little bit more conservative and ask you to provide clarifying information. So this is almost like a human version of retrieval augmented generation, where it's just, if the model says, “Here's what I'm thinking but if you could provide me more information about the prospect in this capacity, or for lead scoring or whatever it is, if you tell it to ask those questions, it will ask those questions to you.”

And then you go, “Oh yeah, that is good contextual information I didn't put in.” And you go retrieve that, or you think about it. Or you say, “I don't have it or something like that.” And then that'll improve the results.

Again, like I said before, a lot of these are going to be commoditized, so it'll probably be built into the model so it gets smarter and smarter to the point that you don't, just like with a human, you don't, people can be trained and have to be educated to think more clearly and more effectively use their reasoning skills at higher levels, but likely a lot of the models will start to have some of these mechanisms built into it, so when you simply engage in a conversation, it'll know in some sense when to jump in, when to adopt different modes of thinking, things like that, so. Anyway, this is some more advanced stuff, but optional and sometimes very useful.

The important point though is again the key ingredients are your goal. What is it you're trying to do? The problem you're trying to solve. All the relevant context you can fit into the token context window, token limit. So as much relevant information as you can. A persona. So if you were to hire an expert to do the task, what are the characteristics they would have? Instructions, so if you have an SOP or you have pre existing guidelines, or you can just ask somebody to articulate it, copy and paste that in. Great. Gold standard examples, definitely positive, I'd say 3 to 5 are good.

You can provide negative ones too, but I found anecdotally just 3 to 5 positive gold standard examples is usually enough to really boost quality. And then a hyper-growth mindset, super ambitious goal, plus radical humility, especially as an expert. And finally, if you want to work with the model, you can do that back and forth through these prompts to simulate or stimulate reasoning, both the model and then yourself. If you're asking it to ask you questions.

That's it on that lesson.

Always feel free to ask us questions and I will see you in the next lesson.