

Getting Started with AI Chatbots

Revised Edition

Mark Pesce

This complimentary edition of *Getting Started with AI Chatbots Revised Edition* is presented by Wisely AI.

To learn more about how Wisely AI can help you become an AI-driven business, safely and wisely, please visit www.safelyandwisely.ai

Getting Started with AI Chatbots Revised Edition

Copyright © 2025, Mark D. Pesce

All rights reserved.

May not be reproduced in any form without the prior written permission of the copyright holder.

The information provided in this book was accurate as of March 2025. The author assumes full responsibility for any errors or inaccuracies that may find their way into this work.

Table of Contents

Introduction	4
Getting Started	7
How AI Chatbots Work	13
Security and Privacy	16
Simple Prompts	18
Reasoning and Summarising	22
Truthiness and Chatbots	29
Character, Context and Conflict	32
Using Character, Context and Conflict to craft powerful prompts	38
Chain-of-Thought prompts	46
Reasoning Models	52
Creating Images with Chatbots	65
Windows Copilot	69
Autonomous Agents	71
Will an AI Chatbot Take My Job?	73
What the Future Holds / Next Steps	76

Introduction

On the 30th of November, 2022, a Silicon Valley startup named OpenAI publicly released ChatGPT, an artificial intelligence 'Chatbot' - and the world changed.

Chatbots have been around in some form for nearly six decades. One of the earliest examples, a program called ELIZA, played the role of willing and engaged psychotherapist. Joseph Weizenbaum, ELIZA's creator, quickly discovered that people poured their souls out to ELIZA, confessing to their deepest fears, longings and hatreds. Yet, even by the standards of the time - when computers filled entire rooms - ELIZA displayed few hallmarks of intelligence. At best, ELIZA could rephrase something said to it, asking the user to explain. Deceptively simple - yet good enough.

Chatbots got another big boost in the early 2010s, as many retailers on the World Wide Web began using chatbots to filter and direct customer inquiries. Although apparently friendly, these chatbots relied on matching 'keywords' in customer inquiries with subject matter. This made those chatbots helpful - but only to a point. If a customer had a problem not covered in the chatbot's programming, or simply had trouble expressing themselves in terms the chatbot understood, they'd fail completely. Although cheaper by far than an army of customer support personnel, they often left customers feeling dissatisfied. They're just not smart enough.

All of that changed with ChatGPT. People were invited to type *anything* that came to mind into its Web-based interface: any request for information, advice, instruction - even imagination. ChatGPT responded in a way that seemed thoughtful, intuitive - and very human. Word quickly spread throughout social media, and within two months OpenAI had over one hundred million registered users for ChatGPT, making it the fastest growing app in history.

Before long, people had found a range of uses for ChatGPT which maker OpenAI had never intended, such as writing essays assigned as school homework, writing steamy 'fan fictions' based on characters drawn from popular TV series, even providing case law and citations for briefs filed in lawsuits. All these creative tasks, previously the exclusive domain of humans, could now be done with a computer program!

Yet the essays spun out by ChatGPT were often bland and formulaic, the fan fictions salacious but unsatisfying - and the case law and citations entirely fabricated! It turned out that ChatGPT did such a good job at 'sounding' human that it could confabulate and never be caught out - the perfect liar.

That a computer program *could* be a perfect liar means we've crossed a technological threshold. Over the last decade, we've grown tired of repeatedly barking orders at 'agents' like Siri, Alexa and Google Assistant, disappointed that they can't be 'smarter'.

ChatGPT reset our expectations for how 'smart' a computer can be: smart enough to understand us; smart enough to gaslight us. It's a powerful new tool, one which can help us, but which needs to be used with a degree of caution.

The technology underlying ChatGPT - known as a 'Large Language Model' - has been publicly available for half a decade. Originally developed by researchers working at Google, OpenAI leapt ahead of the search engine giant with their own large language models, known as GPT-3 and GPT-4, both of which became the basis for ChatGPT. Google has been playing catch-up ever since.

Microsoft, seizing the opportunity to obtain strategic advantage over their competitor in the search engine wars, bought 49% of OpenAI in a defensive purchase, then started integrating ChatGPT across their entire product line. At Microsoft's Develop conference in May 2023, Microsoft CEO Satya Nadella announced 'Windows Copilot' - a version of ChatGPT that will be built into the Windows operating system. Windows Copilot came as a *free* upgrade for more than hundreds of millions of Windows 10 and Windows 11 users that began rolling out in September 2023.

AI chatbots are no longer something 'over there' that you access via a web browser. They're now integrated into *every* part of our laptops, tablets and smartphones - always on and available to answer questions, offer assistance, and perform tasks.

That means it's of vital importance for anyone using any of these devices (nearly everyone) to understand how to use these AI Chatbots to get the best results from them - and to do so safely and securely.

That's what this book is all about.

We'll start off with the basics - how to get started with an AI Chatbot, and the different chatbots on offer from OpenAI, Microsoft, Google, Anthropic and Meta - looking at their relative strengths and weaknesses.

Then we'll explore the technology of these 'Large Language Models' - and don't worry, you won't need a Ph.D. in computer science for that! It'll help us to understand how our 'prompts' (what we type into a chatbot) become 'completions' (their replies).

From there we examine the safety and security issues surrounding AI chatbots - together with a framework useful for weighing risks.

Then we launch into the core of this book, an exploration of 'prompts', bits of text that can make an AI chatbot do nearly anything requested, so long as the prompt is framed correctly. That's a point that's often overlooked; with an AI chatbot how you phrase the prompt is often as important as its content.

We'll explore some basic prompts, then go on to more sophisticated prompts using 'personas' - tiny stories that help an AI chatbot 'stay in character' as it generates a response.

Then we'll look at 'chain-of-reason' prompts, teaching an AI chatbot how to perform a task. That's a technique that can be applied to almost any problem area, whether writing a report, analysing a spreadsheet, or designing a website - and on and on and on. Chain-of-reason prompts are a sort of 'programming' that isn't hard to learn and can be very powerful.

That's a good moment to look at the latest developments in AI Chatbots – the so-called 'reasoning' chatbots. These latest-generation tools can take a complex request and 'work out' how to perform that request, by detailing the steps involved, then acting upon those steps. It's early days, but these reasoning chatbots look to be very powerful ways to use AI.

Finally, we'll look at where all of this is going. The whole field of AI chatbots is still brand-new, and people are very worried that they'll be put out of work by ChatGPT. Is that likely? And what sorts of skills will help us thrive in an age of AI chatbots? How can we use chatbots to help us be even better in our own work?

First and foremost, we need to embrace the spirit of play. AI chatbots are inviting us into a conversation; if we use them well, they can help make us smarter. That's exciting - and reason enough to get started.

Chapter 1 - Getting Started

The 'Big Five'

Three AI Chatbots dominate the Internet. ChatGPT, the first of these, ChatGPT, from artificial intelligence startup OpenAI, came out in December 2022. Within a few months ChatGPT had competitors, brought to market quickly by two of the usual 'big tech' suspects - multi-trillion dollar software firms Microsoft, with its Copilot, and Google, with Gemini.

Each of these chatbots offers much the same service, delivered in the same way - a Web-browser based conversation with an artificial intelligence chatbot. In each of them, a user can type in pretty much any sort of question, request, problem, observation - even a whole scientific paper - and ask the chatbot to respond.

The quality and the depth of that response varies significantly among these three offerings. Some show hallmarks of being rushed to market, and feel a bit unfinished. The months and years ahead will see these differences smoothed out; within a year or two all three chatbots should offer remarkably similar experiences.

That means it's a good idea to know how to use all three of them, and be prepared to use them interchangeably. There may be a situation where one is available while its alternatives are not - so it's a useful skill to be across all of them. Just in case.

We'll take them individually, starting with the first and (arguably) the best.

OpenAI ChatGPT

What is it?

The granddaddy of AI Chatbots, ChatGPT provides a conversational interface to an artificial intelligence program.

It comes in two flavours: ChatGPT, powered by the last-generation, less capable GPT-4.0-Turbo large language model; and ChatGPT+, offering subscription-based, priority access to ChatGPT, paired with its latest-and-greatest best-in-class GPT-4.5 model.

ChatGPT+ costs around USD \$20 per month - and while you can use it as much as you like in GPT-4.0-Turbo mode, subscribers will find themselves 'rate limited' to **50 chat interactions every three hours** when using the computationally much more intensive (and expensive) GPT-4.5.

Where possible, ChatGPT+ using GPT-4o is always to be preferred over ChatGPT using GPT-4.0-Turbo. The results are just that much better, because GPT-4.5 is the current gold standard for large language models. It has no peer.

The question is whether you want to pay for that power.

How do you access it?

ChatGPT is an app that runs in your web browser - it can be accessed at <https://chat.com/>

OpenAI also provides a ChatGPT app for both iOS and Android devices. (iPadOS users can use the iOS app.) Those apps are available through the Apple App Store and Google Play, respectively.

Do I need to sign up?

Yes. OpenAI requires that you sign up for a free account on their website.

You can do that by going to the main page for OpenAI, at openai.com, accessing the menu, and selecting the 'Sign up' option. (As stated in the last chapter, be sure to read all the 'Terms and Conditions' you're agreeing to.)

Once you've created your account, it is attached to your 'chat history' with ChatGPT. This means ChatGPT can 'remember' all of the chat sessions you've had with it, which allows you to go back to any of them, review them, download them - even revise them and have another go with ChatGPT.

How do I use ChatGPT?

Simply put, you type whatever you like in the "prompt" area at the bottom of the web page. ChatGPT will respond in the area above - and will show your 'prompt' above its response.

This truly is a conversation; you can pick up on any point made by ChatGPT in its response, ask it to expand upon, defend, or provide additional examples of any of its replies.

Managing that conversation well, using 'prompts' - that's the subject of the rest of this book.

Microsoft Copilot

What is it?

Just a few months after ChatGPT launched, Microsoft launched its own chatbot - named Copilot. Microsoft was able to come to market so quickly because it made a massive, ongoing investment in OpenAI. That gave the tech giant access to OpenAI technologies such as GPT-4o, which they used to build their own AI chatbot.

Microsoft's Copilot is a hybrid: Microsoft had been working on a variety of AI chatbots for almost a decade, and blended its own experiences with chatbots into the capabilities provided by OpenAI's GPT-4o. This gives Copilot a 'personality' distinctly different from ChatGPT, even though they are built upon much the same foundations.

How do you access it?

Microsoft wants to encourage everyone to use their new web browser, Edge, so they've built Copilot into Edge Browser. In the upper right hand corner of the browser window, there's a blue semi-rectangular symbol that represents Copilot. Click on that icon, and as much as half of the right side of the browser window will become Copilot. To close Copilot, simply type on the icon again.

Microsoft has also released Copilot mobile apps for iOS and Android, and now Copilot runs in *most* web browsers.

Do I need to sign up?

Yes. Everyone using a Microsoft operating system (or other product, such as Microsoft 365, Business Intelligence, etc.) will already have a login - either personal or through their organisation - at Microsoft.com. If you don't have already have a microsoft login, go to Microsoft.com and sign up.

How do I use Copilot?

Although Copilot operates very similarly to ChatGPT, though the user interface is slightly different – just type your prompts into the box labeled 'Message Copilot'. Copilot will print its own responses above that box - and will also show your 'prompt' above its response.

Again, this truly is a conversation; you can pick up on any point made by Copilot in its response, ask it to expand upon, defend, or provide additional examples of any of its replies. Copilot will eventually limit the number of back-and-forths in any conversation. After that limit is reached, Copilot will ask you to initiate a new conversation.

Google Gemini

What is it?

As the inventor of the modern technology underlying all AI Chatbots - a 2017 paper by Google researchers made these 'Large Language Models' possible - Google was caught out by the arrival, quality and success of ChatGPT. Playing catch up ever since, Google launched its own Gemini Chatbot in early 2024.

Like ChatGPT and Copilot, Google Gemini provides a conversational interface to an artificial intelligence program.

Many users find Gemini the most 'bland' of the chatbots. While ChatGPT can be warm and friendly, and Copilot can be almost too helpful, Gemini comes across as entirely neutral. That may be a good thing, depending on how you feel about whether AI chatbots should have 'personality'.

How do you access it?

Google Gemini is available via a web page at <https://gemini.google.com/>. Just point your browser there.

Google also provides apps for Gemini on both Android and iOS Android - or you can access Gemini from a mobile browser.

Do I need to sign up?

Yes. You'll need a Google login to use Gemini. However, almost everyone who uses any of Google's service - such as Google Calendar, Google Mail, Google Documents, Google Sheets, etc. - already has a Google account.

If you need a Google account, visit the Google home page, click on the button that says, "Sign in", then follow the instructions for setting up an account.

How do I use it?

Type whatever you like in the "Ask Gemini" area at the bottom of the web page, then hit Enter. Gemini will respond above - and will show your 'prompt' above its response.

This truly is a conversation; you can pick up on any point made by Gemini in its response, and ask it to expand upon, defend, or provide additional examples of any of its replies.

Like ChatGPT, Gemini also maintains a memory of all of its other chat sessions - you can access those from the bar running up the left side of Gemini's web page.

Anthropic Claude

What is it?

A messy divorce between the founders of OpenAI led to the formation of a competing startup – Anthropic. Unconvinced that OpenAI had taken the necessary precautions to create 'safe' and 'aligned' AI chatbots, Anthropic wasn't first to market with their Claude chatbot – but to all accounts it consistently equals the quality of anything put out by rival OpenAI.

Like ChatGPT, Copilot, and Gemini, Claude provides a conversational interface to an artificial intelligence program.

Many users find Claude the most useful and ‘complete’ of the Big Five chatbots. That makes Claude very useful when you’re looking to understand something, or some assistance in performing a task.

How do you access it?

Anthropic Claude is available via a web page at <https://claude.ai>. Just point your browser there.

There are mobile apps for Claude available on both iOS and Android.

Do I need to sign up?

Yes. You’ll need to log into Claude to use it – this is generally no more difficult than giving it an email address, then clicking on the confirmation link in the email. Make sure to use the same email if you want to retain the ‘history’ of your conversations with Claude.

How do I use it?

Click the ‘Start new chat’ button, then type whatever you like into the box labeled with “How can Claude help you today?”. Hit Enter and Claude will respond above - and will show your ‘prompt’ above its response.

This truly is a conversation; you can pick up on any point made by Claude in its response, and ask it to expand upon, defend, or provide additional examples of any of its replies.

Like ChatGPT, Claude also maintains a memory of all of its other chat sessions – roll over your user icon in the lower left-hand corner (it generally displays your initials) and you’ll see a popup with all your previous chats appear above.

Meta AI

What is it?

Meta has gone ‘all in on AI’, bringing their own AI chatbot – Meta AI – to their massive, global base of billions of users.

Like ChatGPT, Copilot, Gemini and Claude, Meta AI provides a conversational interface to an artificial intelligence chatbot.

How do you access it?

Meta AI is fully integrated into Facebook, Facebook Messenger, Instagram, and WhatsApp. It's also available via a web page at <https://meta.ai>. Just point your browser there.

Meta AI is also built into the mobile apps for Facebook, Messenger, Instagram and WhatsApp.

Do I need to sign up?

No. You can use Meta AI via the website without logging in – though you'll need to provide your age. However, if you want to use features such as the image generator built into Meta AI, you will need to log in for that.

How do I use it?

In Meta's apps, the search bar has also become the Meta AI interface – so you can 'Ask Meta Ai or Search'. Type in your message, hit return, and Meta Ai will generate its response.

Meta AI can also keep the 'context' of a social conversation; in a group chat of parents it may be able to answer questions about parenting, and so forth. That makes it a bit different (and potentially more controversial) than the other chatbots, which operate in isolation.

Read the Terms and Conditions

Before you sign up to use *any* AI chatbot, have a look at the 'Terms and Conditions'. Yes, that's the box we always click 'OK' on without reading anything (thereby giving our attorneys headaches). In this case you're well advised to read the fine print. How are OpenAI, Microsoft, and Google using the prompts your submitting? Do they retain them? How long? For what purposes. That's all in this fine print - and it's very important you know what it says.

(Protip - feed this fine print into a chatbot and ask it for a summary.)

Chapter 2 - How AI Chatbots Work

Here's a fun game you can play with your smartphone that can give you a fair sense of what's going on inside of an AI chatbot.

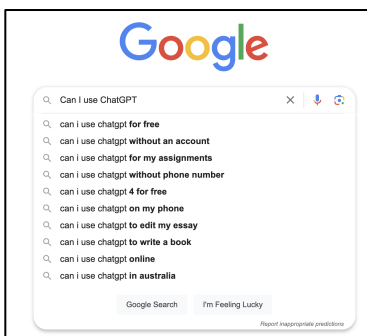
To keep us all from fumbling and misspelling our way through our text messages, modern smartphones offer 'predictive' features. When you type the first character of a word, your device will begin to offer up suggestions on how to finish typing that word. Select that suggestion and it's typed into the message. Once again, your smartphone tries to guess the most likely word to follow that word - a calculation it makes based upon the frequency of all the words you've ever typed into your smartphone. If you type "This is" a lot on your smartphone, when you type "This", the smartphone will immediately recommend "is" as the next most likely word.

Here's the game: keep selecting the next word that your smartphone predicts. You'll inevitably end up with a run-on sentence that *almost* seems to make sense. Almost. Here's one created by my own iPhone:

I'm sure it so I'm a little bit of an expert in the subject but I'm a little more knowledgeable on this topic and have been able with the help I can give to get my phone started

So close to sensible, yet so far from sense. Prediction on a smartphone does an OK job at guessing out the next word, but beyond that, it quickly loses its way.

This kind of predicting-what-I'm-about-to-type features heavily in Google's search box. When you type in the start of a search request, Google offers a whole series of choices for how it predicts the search request will look.



Google can do this with a fair bit of accuracy because it has analysed millions or even billions of similar search requests - and remembers them. If you type "Can I use ChatGPT" into the search box, you'll see predictions beneath the

search box, including “Can I use ChatGPT to **edit my essay**” and “Can I used ChatGPT in **Australia**”. Both make sense given what’s already been typed into the search box - and both have been typed in many, many times by other Google users.

Both a smartphone’s predictive text and Google’s search predictions rely on ‘machine learning’. That means they’re continuously watching everything typed in, using that information to build a model of the likelihood that a given word might follow another, or the likelihood of phrases that could complete a partially typed-in search request. Without this machine learning, their suggestions would be utterly random - and completely useless.

ChatGPT works along similar lines. Fed with an enormous amount of text - a decent portion of *every* text freely available on the Internet, including sites like Wikipedia and Reddit - countless billions of words flow into its model during its extensive, months-long (and expensive) ‘training’.

During this training, the model becomes better and better at predicting which word is most likely to follow a previous word - just as a smartphone does. But ChatGPT goes much further than that. Built on top of a 2017 breakthrough in artificial intelligence known as ‘transformers’ (that’s where the ‘T’ in GPT comes from) - ChatGPT can read in thousands of words in a ‘prompt’, using these words to predict which words are most likely to follow on.

These transformers provide far more accurate (though far from perfect) predictions about the next most likely word, sentences, even paragraphs. ChatGPT has so much more to work with - both because of its training and its transformers - that it can be both more far accurate and more knowledgeable.

When ChatGPT receives a ‘prompt’ from a user, it does its best to find the most likely words that will ‘complete’ the prompt. (It’s even known technically as a ‘completion’.) Given the enormous wealth of material fed into ChatGPT, more often than not, that completion satisfies the user. It seems to the user as though the ChatGPT has ‘thought up’ an answer to the prompt - but there’s no thinking going on here. Instead, a subtle and sophisticated form of search has been used to ferret out of ChatGPT a completion that most probably meets user’s needs.

Nevertheless, most users immediately suppose ChatGPT is thinking - because what it says so closely imitates the way we ourselves would respond. But that’s down to ChatGPT’s training: it has digested and learned from countless billions of written interactions. From those, it learned to correctly predict the response to almost any interaction, providing that as a completion. It’s impressive - and has many potential uses - but it’s not ‘thought’ in any sense of the word.

That’s one big reason AI chatbots can ‘lie’ so convincingly - a topic covered at length in chapter 6. Without thought, ChatGPT has no capacity to maintain any model of wrong/write, true/false, truth/lies. When presented with a prompt its programming does its best to generate a completion. Where that completion

reflects reality, we credit ChatGPT with knowledge. Where that completion represents a complete fabrication, we credit ChatGPT with mendacity. Both reactions give too much credit to ChatGPT. It's simply offering a completion - and has no way to know whether that completion reflects the truth, since it has no way to know anything at all.

All ChatGPT (and the other AI Chatbots, Copilot and Gemini) offer is an extensively trained computer program generating completions in response to prompts. That sounds simple - and in one sense, it is. Yet, because human language is rich and strange, ChatGPT can appear both rich in knowledge and profoundly strange in practice. To uncover its full capacities, it may be better to think of ChatGPT as something that does its best to plausibly respond to human language - *without understanding any of it*.

Knowing this, we can treat AI Chatbots as just another computer program. They're nothing special. Yes, they may be very different from other computer programs we've used, but like those programs, they follow rules. Those rules, as we understand them, are explored in this book.

Chapter 3 - Security and Privacy

How safe is an AI Chatbot? After all, the Internet can be a dangerous place. While we may be familiar with attempts to 'spoof' a bank's website - the better to steal account and login details from unwitting customers - that's not the only threat on the Internet, nor the most common.

Sadly, the biggest threats to our security almost always center around a trusted organisation (such as a retailer, bank or even a government website) that fails to properly secure its confidential information. Employees in targeted organisations regularly fall prey to 'social hacks', leveraging personal relationships to override cautious safeguards, then handing secured information to intruders. On other occasions, intruders ferret their way into a network through a weakness - often a device which hasn't had regular software updates. Once inside, they install software to monitor and harvest any information passing through those networks.

All of this means that our personal, confidential data risks being exposed to bad actors. The costs of failure are so low, and the upside of success so high, our networks have become battlefields of continuous low-grade warfare waged between attackers and defenders. As you hear a lot in cybersecurity circles, the attackers only have to succeed once, whereas defenders need to be successful *every* time.

Although large organisations try to give the impression that they're capable of defending against every assault from intruders, the unvarnished truth is that *all* organisations are vulnerable - perhaps even more so when they project an aura of invulnerability. Microsoft, one of the oldest and most experienced companies in technology, fell prey to an attack in mid-2023 which compromised a large number of users of its Outlook mail and scheduling tool - including highly secure government accounts. No company is immune, and no online data is ever truly entirely safe. This means that care and caution are the order of the day when doing *anything* on the Internet.

What does this have to do with AI Chatbots? Everything. The 'Big Five' providers of AI Chatbots - OpenAI, Microsoft, Google, Anthropic and Meta - will all be applying best practice to their online security (we hope). That doesn't mean they won't be attacked, nor does it mean that those attacks will inevitably fail. It's possible that user account information - including address and billing details - will end up in the hands of bad actors.

When signing up for these services, that's an important point to keep in mind. But that's less than half the story here. What's far more important to understand is what happens when everything is working as planned.

When deeply involved in a conversation with an AI Chatbot, we tend to forget that we're interacting with a computer program that's operating somewhere far away, out on 'the cloud'. Although the conversation can seem very intimate,

personal and particular, the orchestration of that conversation requires the marshaling of enormous resources.

This means that anything typed into an AI chatbot as a 'prompt' - and any of its completions - passes through a range of computing devices on its journey. Every hop on that journey represents a point where these systems can make a record of that interaction.

In the normal case, that data might be preserved in order to improve the operations of the chatbot - helping it to 'learn' from its interactions. It might also be used to generate 'analytics' - statistical information that helps the chatbot provider to understand who is using their chatbot, and for what tasks. Both of these are legitimate uses for interaction data, though you'll need to read the 'Terms of Conditions' for each service to know for sure. Those user agreements spell out the potential use of your prompts for these purposes.

Given that your prompts are likely to be recorded by the company providing the AI chatbot, you need to ask yourself a few questions before you type *anything* into that prompt box:

Is this information personal?

If this information were exposed by hackers - or simply made available in a public database of training data - would it expose personal information about yourself or another individual?

Is this information private?

Does this information concern some aspect of a person, family, or organisation that would normally be considered private, and therefore closely held? Medical, financial and legal information often fall into this category.

Is this information commercial-in-confidence?

Would this information disadvantage a commercial organisation if released publicly? Would it advantage a competitor if they gained access to it? Could this information be used to manipulate markets? Would the release of this information be regulated under securities laws?

Is this information protected by law?

Finally - is this information protected by law? Is it covered under export controls? Is it classified information? Would it put at risk individuals, organisations or governments if it became widely known? Would a civil or criminal prosecution result from the public release of this information?

Any information submitted as a prompt to an AI chatbot should *never* trigger any of the conditions listed above. Nor should it violate good common sense: *If you wouldn't shout it from a rooftop, you shouldn't type it into a chatbot.*

Chapter 4 - Simple Prompts

Hundreds of millions of people have already had a conversation with an AI chatbot. What do they talk about? The content of those conversations runs the gamut from the mundane to the profane to the profound.

Nearly all of them begin with a basic quest for information.

Type “What is the capital of Finland?” into Google, and you’ll get links to an answer - together with a sidebar that links back Wikipedia’s article about Helsinki, Finland’s capital.

What happens when you put that question to an AI chatbot? It depends on the chatbot *and* it depends on how you phrase the question. Phrasing *matters* to an AI chatbot. It pays attention to your every word, and does its best to provide a completion that addresses every element in those words.

Let’s ask that basic question to ChatGPT, Copilot, and Gemini:

ChatGPT:

The capital of Finland is **Helsinki**.

Ok, that’s quite... straightforward. Now let’s ask Copilot...

The capital of Finland is Helsinki. It's a beautiful city located on the southern coast of the country, known for its stunning architecture, vibrant culture, and lovely coastal scenery. Have you ever visited Helsinki or do you have any plans to go there?

Wow. That’s a much more fulsome response. Copilot gets to the core of the answer in the first sentence (using the very same words as ChatGPT) goes on from there.

Now let’s see what Gemini says...

The capital of Finland is Helsinki.

1. Discover Helsinki, the capital of Finland
<https://www.visitfinland.com/en/places-to-go/helsinki-region/helsinki-city/>

Gemini, like ChatGPT, provides a single sentence in response – but with a reference. That link allows you to verify for yourself that Gemini is giving you a correct answer – and gives you a chance to explore the subject in further detail.

What if we ask ChatGPT "Could I have more information about it?"

Sure! Here are some key facts about **Helsinki**, the capital of Finland:

General Information:

- **Location:** Southern Finland, on the shore of the Gulf of Finland
- **Population:** Around 660,000 (metro area: ~1.5 million)
- **Language:** Finnish and Swedish (official languages)
- **Time Zone:** Eastern European Time (EET, UTC+2) / Eastern European Summer Time (EEST, UTC+3 in summer)

History & Culture:

- **Founded:** 1550 by King Gustav I of Sweden
- **Capital Since:** 1812 (previously Turku was the capital)
- **Nicknames:** "Daughter of the Baltic" due to its seaside location
- **Architectural Style:** A mix of Neoclassical, Art Nouveau, and modernist design

Key Landmarks:

1. **Helsinki Cathedral** - A striking white Lutheran cathedral in Senate Square.
2. **Suomenlinna Fortress** - A UNESCO World Heritage Site, built in the 18th century.
3. **Temppeliaukio Church (Rock Church)** - A unique church carved directly into solid rock.
4. **Market Square (Kauppatori)** - A bustling market by the harbor with fresh food and crafts.
5. **Oodi Library** - A modern architectural wonder and a hub for digital and creative culture.

Climate & Lifestyle:

- **Climate:** Cold winters (-5°C to -20°C), mild summers (15°C to 25°C).
- **Public Transport:** Efficient tram, metro, and bus systems.
- **Education & Innovation:** Home to Aalto University, a center for tech and design.
- **Sauna Culture:** There are over 3 million saunas in Finland, and Helsinki has some of the most famous public ones (e.g., Löyly, Kotiharjun Sauna).

Economy & Industry:

- **Key Sectors:** Technology, design, maritime industries, and gaming (Angry Birds, Clash of Clans).
- **Connectivity:** A major Nordic business hub with a highly developed infrastructure.

Two things to note – ChatGPT ‘remembered’ that we’re chatting about Helsinki. This memory is known as ‘context’ – and context is very important in a conversation with a chatbot.

The second thing to note is that once we gave ChatGPT permission to share more – it provided a detailed list. ChatGPT knows a lot – but it’s also learned to share in detail only when asked to do so.

Known as ‘zero-shot’ prompts - simple, direct commands to provide information - these sorts of question-and-answer sessions can be ‘tuned’ with additional phrases.

“Be concise”, “In one word”, and “Please explain your answer” - plus almost any other phrase you can imagine - all shape completions in different ways. Each AI chatbot will take a slightly different approach in those completions.

Here’s a perfect opportunity to begin to experiment. Take some time to have an extended conversation with an AI chatbot, playing with the language in your prompts. Shape the chatbot’s completions to something that fits your needs in that moment.

This skill is a first step into mastery.

Suggestions for exploration

Can you get an AI Chatbot to give its completions in rhyme?

In haiku?

What about as a ballad?

There's only one way to find out.

And what about other languages?

Can ChatGPT speak Japanese? Can Gemini speak German? And will Copilot be able to give you completions in Thai?

Again, only one way to find out -- have a go!

Chapter 5 - Reasoning and Summarising




Getting simple answers out of an AI chatbot can be done simply and efficiently with direct questioning using 'zero shot' prompts. But many questions have more subtlety to them; questions sometimes have other questions embedded in them, needing to be broken down and answered separately in order to answer the question as a whole.

A simple example of this might look like, "What's the total population of the three largest cities in Finland?" First, you'd have to look up the data on the three biggest cities in Finland, then find their population, then add those together.

That sort of prompt triggers a chain of actions in the chatbot while it generates its completion.

Here's what happens when we put this question to ChatGPT:

As of December 31, 2023, the three largest cities in Finland by population are:

1. **Helsinki:** 674,500 inhabitants 
2. **Espoo:** 314,024 inhabitants 
3. **Tampere:** 255,050 inhabitants 

The combined population of these cities is 1,243,574.

Once again, ChatGPT does an excellent job with its completion. It also provides links to websites – which it digested while generating the response.

This is an important point: the first generation of AI chatbots operated in isolation. They used the data they'd been trained upon to generate responses to prompts. But that data may not always be accurate – resulting in chatbot providers building a 'search the web' function into the process of generating a response, along with links to the sites cited in the response.

ChatGPT also makes it quite clear that its population data may not be current - because of its 'knowledge cutoff'.

KNOWLEDGE CUTOFFS

All large language models have to be 'trained' on vast amounts of data. ChatGPT and the GPT-4o large language model underneath it have been trained on an unfathomably large set of text data - nearly everything that's ever been published to the Internet, and a fair bit of literature, periodicals, and so forth. It takes a long time (months rather than days) for a language model to digest all of this

information to the point where it starts providing 'good enough' completions to prompts. It's a huge task - not just in the training, but in the 'tuning'. Tuning generally involves an army of humans feeding endless prompts into the model, examining the completions, and providing 'human feedback reinforcement learning' - that is, 'rewarding' the model when it does what it's supposed to, and 'punishing' it when it provides the wrong completion. (Reward and punishment are mathematical constructs - ChatGPT is not having its knuckles rapped by a ruler.)

Because of all of the necessary work to get a 'good enough' chatbot ready for interactions with the public, there's quite a span of time between feeding in all of the texts to train the model and when that model is actually usable by the general public. That means every large language model has a 'cutoff date' - the most current information available to the model. In March 2025, ChatGPT knew nothing about events beyond its December 2023 cutoff date.

Every three to six months these models get refreshed; before long there will be another version of ChatGPT that will be current to June 2024, and six months later, another, current to December 2024, etc. But that still means there's a gap of around two years between what the chatbot knows about the world, and the world itself. While there may be slight differences between chatbots in this respect, it's important to remember that, in general, their information isn't perfectly up-to-date. That's why adding web searching capabilities was so necessary for improving the accuracy of a chatbot.

Will Copilot generate a more up-to-date completion to our complex prompt?

The three largest cities in Finland are Helsinki, Espoo, and Tampere. As of the latest data, their populations are as follows:

- **Helsinki**: 674,500
- **Espoo**: 314,000
- **Tampere**: 255,100

So, the total population of these three cities is approximately **1,243,600**.

[Statista] (<https://www.statista.com/statistics/327469/largest-cities-in-finland/>)

The response generated by Copilot is nearly the same – 1,243,600 versus 1,243,574. But why isn't it number-for-number the same as ChatGPT? The references for the two responses are different, and the response reflects that difference. This may seem confusing, but that's the nature of knowledge – it's only as accurate as the references it has been drawn from.

Now let's see how Gemini answers the same prompt:

To get the most accurate answer, it's important to consider that population figures can vary depending on the source and the specific definition of "city" (municipality vs. urban area). However, based on the information I found, here's a breakdown:

- **The three largest cities in Finland are generally recognized as:**
 - Helsinki
 - Espoo
 - Tampere
- To provide approximate numbers, drawing information from citypopulation.de, showing estimates from the end of 2023:
 - Helsinki: Approximately 674,500
 - Espoo: Approximately 314,024
 - Tampere: Approximately 255,050
- Therefore, the total approximate population of these three cities would be:
 - $674,500 + 314,024 + 255,050 = 1,243,574$.

It's important to keep in mind that these are estimates, and population numbers are constantly changing.

Gemini walks us through its 'reasoning' process as it generates the answer – even putting the website citation within the response! It, too, comes up with the total of 1,243,574 – the same as ChatGPT.

Three AI chatbots, and two answers. Which one of these chatbots can we trust as being definitive? Do we go with a majority vote? Can we trust any of them?

(That's a big question we'll cover in detail in the next chapter.)

All three responses provide references, so they're all verifiably 'correct' – even if they don't agree. Maybe we can be satisfied that any of these answers are 'good enough' – because it's unlikely that any of them are *precisely* accurate.

You can make a reasoning prompt to an AI chatbot quite complex, embedding request within request with request - up to a point. Beyond that point the chatbot will begin to misinterpret, mistake or simply miss the point of the prompt. There are no hard and fast rules about when and how an AI chatbot will fail to understand a complex prompt. If you're not getting the results you

expect, try breaking the prompt into a series of steps - something we'll cover in detail in Chapter 8.

Summaries

AI chatbots have proven to be quite useful at summarizing large documents. Have a 20,000 word document that you need to be across for a meeting that's coming up in 5 minutes? Feed it into an AI chatbot with the prompt, "Please summarise this document".

For example, I copied-and-pasted the full text of the US Declaration of Independence - which runs to 1,336 words - into Claude, preceded with the prompt "Pease summarise this document". Here's its completion:

This document is the Declaration of Independence of the United States, adopted on July 4, 1776. It consists of five key sections:

1. An introduction explaining that breaking political ties with Great Britain requires a formal explanation.
2. A preamble outlining the fundamental principles of natural rights and government by consent, stating that "all men are created equal" with "unalienable Rights" including "Life, Liberty and the pursuit of Happiness." It asserts that people have the right to alter or abolish governments that become destructive to these ends.
3. A lengthy indictment against King George III, listing 27 specific grievances that justify separation. These include refusing assent to necessary laws, dissolving representative assemblies, obstructing justice, imposing taxes without consent, and waging war against the colonies.
4. A section describing failed attempts to appeal to the British people, noting that petitions were met "only by repeated injury."
5. The formal declaration that the colonies "are, and of Right ought to be Free and Independent States," absolving all allegiance to the British Crown and claiming the full powers of sovereign nations.

The document concludes with the signatories pledging "our Lives, our Fortunes and our sacred Honor" in support of independence.

Not bad. (Probably good enough to pass an exam with.)

Let's see what Meta AI provides as a completion for this prompt:

This **The Declaration of Independence (1776)**

The document outlines the 13 American colonies' decision to declare independence from Great Britain and establish themselves as sovereign states.

Key Points:

1. **Preamble:** The colonies have suffered under British tyranny and have the right to alter or abolish their government.
2. **List of Grievances:** The colonies enumerate the abuses of power by King George III, including:
 - Refusal to pass laws beneficial to the colonies
 - Imposition of taxes without consent
 - Dismissal of colonial governors and appointment of new ones
 - Quartering of British soldiers in colonial homes
 - Deprivation of trial by jury and other rights
3. **Declaration of Independence:** The colonies declare themselves free and independent states, with the power to govern themselves and conduct foreign affairs.
4. **Pledge of Mutual Support:** The signers pledge their lives, fortunes, and sacred honor to support the Declaration and the newly independent states.

Significance:

The Declaration of Independence marked the beginning of the United States of America as a sovereign nation, founded on the principles of liberty, democracy, and self-governance.

Meta AI's completion looks perfectly reasonable. Is it good enough to give you a 'B' on an exam?

This same prompt “Please summarise this” followed by pasted-in text, or the upload of a text document, means that you can use AI chatbots to help you read through scientific reports, articles and all sorts of other text-heavy work.

But it’s important to understand that it is possible to feed an AI chatbot more than it can digest. If you do, the responses from the AI chatbot will become confused and inaccurate.

Different AI chatbots have different limits for how much information you can paste into or upload within a prompt. To understand why, we need to introduce a new concept, the ‘context window’.

THE CONTEXT WINDOW

All AI chatbots have a limit to the size of the prompt they can work with. When it’s being digested by a chatbot, your prompt is broken down into ‘tokens’ - with a rough approximation of one token per syllable. Those tokens are then used to drive the ‘transformer’ that helps the chatbot come up with the most probable completion for the prompt. More tokens generally means a more accurate completion, so these AI chatbots have been designed to be able to handle thousands of tokens in their prompts - something that’s known technically as the ‘context window’.

You can think of a context window as something akin to the AI chatbot’s equivalent of short-term memory. Without that context window, the chatbot wouldn’t be able to provide a meaningful completion. If you can’t fit your entire prompt into the context window - as was the case above with Copilot, cutting off the Declaration halfway through - the AI chatbot won’t have everything it needs to generate a meaningful completion. It’ll miss the point, or part of it.

There’s an added complication; in a ‘conversation’ with an AI chatbot, the chatbot needs to track both your prompts to it, and its completions. If it doesn’t add its own completions to the context window, then it can’t keep track of the conversation. The chatbot will ‘lose context’ and lose the thread of the interaction.

While it’s important to provide detailed prompts to an AI chatbot, it’s also important to note that the chatbot’s ability to ‘remember’ the context of a conversation has limits. Pass those limits and you could see a good conversation go south. A good interaction with an AI chatbot is a delicate balance between brevity, effectiveness and memory.

A final, important note: As explored in Chapter 3, it’s almost always a *very bad idea* to use an AI chatbot to summarise confidential, private or secured information.

This brings up another big question - how can we know if the summaries provided by any AI chatbot are correct?

Chapter 6 - Truthiness and Chatbots

Do you believe everything you read?

In this day and age, provided with a growing range of media sources - from the trustworthy to the ludicrous - we tend to view what we read with a healthy dose of skepticism. There's a rule a thumb from scientist Carl Sagan that we've learned to apply in our own media consumption: "Extraordinary claims require extraordinary evidence."

But what happens if those claims seem entirely ordinary? Do we believe them *because* they conform to what we believe? In general, that's what we'll do -- until we learn our assumptions led us astray.

That's what happened¹ to attorneys Steven A. Schwartz and Peter LoDuca. Working on a lawsuit against an airline - filed in US Federal Court - Schwartz did what every lawyer does when researching a case: he used the tools available online. In his case, he used the very new ChatGPT. Believing this to be a tool similar to LexisNexis (which provides case law and citations that lawyers use in preparing their arguments), Schwartz asked ChatGPT to provide case law and cases similar to his own. ChatGPT provided a whole list of cases, and citations from the relevant legal texts.

Only one problem: ChatGPT had made everything up.

When ChatGPT's confabulations came to light, presiding US District Judge P. Kevin Castel took a very, *very* dim view of Schwartz's actions, forcing him into a humiliating admission. "I did not comprehend that ChatGPT could fabricate cases," Schwartz acknowledged in open court. "I would like to sincerely apologize."

Schwartz believed ChatGPT to be a knowledge source carefully tuned to the needs of experts like himself, so he didn't bother to fact-check its completions. After all, he'd never needed to fact-check LexisNexis, a body of knowledge created by legal experts, curated by legal experts, and used by legal experts.

AI chatbots have no specific expertise. They can be 'fine tuned' - that is, they can be trained against specific bodies of knowledge, such as law, physics, or programming languages, but that only gives them depth, not expertise. Chatbots do not 'think' when generating a completion, and therefore cannot assess whether any given completion has any basis in fact.

¹ <https://www.abc.net.au/news/2023-06-09/lawyers-blame-chatgpt-for-tricking-them-into-citing-fake-cases/102462028>

Because their models are so large and so broad - billions upon billions of words - an AI chatbot will nearly always be able to find a completion for a given prompt, and because of the way these models are trained, that completion will nearly always seem a very reasonable response. It will have the feeling of truth.

We have a name for that feeling, courtesy of comedian Stephen Colbert: 'Truthiness'. Here's how Wikipedia defines it:

***Truthiness** is the belief or assertion that a particular statement is true based on the intuition or perceptions of some individual or individuals, without regard to evidence, logic, intellectual examination, or facts.*

AI chatbots are carefully and precisely engineered to deliver the most predictable, most likely - and therefore, most likely accurate - completion. That's the essence of how they operate under the hood. An AI chatbot will always give the impression of being completely reasonable, even when generating a completion that has no basis in fact. *Especially* when generating a completion that has no basis in fact.

Does this innate propensity to truthiness mean AI chatbots are useless, even dangerous? No. But it does mean that where we ask an AI chatbot for its expertise, we need to tread very carefully.

That sort of caution goes against the grain: forty-five years of PCs have us well-trained to trust the output of any computer. Except in those rare instances where we encounter a 'bug', we have come to expect computers to faithfully calculate, edit or reproduce any data fed into them. That any computer program might deviate from this very narrow track strikes us as odd - almost unnatural. After all, while humans are prone to 'telling a porker' every so often, we've never had anything like that sort of experience with our computers.

Our expectations of how our computers behave need to change: AI chatbots regularly generate completions that have no basis in reality.

What do we do with that knowledge? **Go see an expert.**

For example, when asking an AI chatbot to summarise a document, ask yourself if you are expert enough to be able to detect truthiness in its summary: Is this something that looks true, or actually is true? If you can't make that determination on your own, turn to an expert. **Expertise is the antidote to truthiness.**

Expertise is hard-earned, and very human. As we bring AI chatbots into every aspect of business operations, the demand for expertise will increase. It's a good time to be an expert. It also means that a lot of the savings forecast by employing AI chatbots in businesses will be canceled out by a growing demand for expertise - that is, 'humans in the loop'. Yes, much more can be automated - but to do so safely will require a corresponding wealth of expertise. In a world of AI chatbot truthiness, there's no free lunch.

The factual failings of AI chatbots have been well documented, and research has continued to improve their accuracy. At some point they may well be ‘good enough’ in most situations, most of the time. But this propensity to hallucinations represents an essential flaw in AI chatbots. Better is all we can hope for. **They’ll never be perfect.** Their propensity to provide reasonable completions, ignorant of facts, means that completions always need to be regarded with a degree of skepticism.

That skepticism means that wherever possible, put a given prompt to multiple AI chatbots. Compare their outputs. Do they agree? Or, as with our example summing the population of Finland’s three largest cities, do they all come up with their own versions of the truth?

Finally, never forget that there is no substitute for *human* expertise.

Chapter 7 - Character, Context and Conflict

So far, we've learned how to make basic requests to an AI chatbot, how we can modify its completions with some simple instructions in our prompts, and we've learned that AI chatbots have an incredible capacity to 'gaslight' us - creating misinformation that 'feels' so real it becomes very difficult to discern truth from confabulation. That's a weakness of AI chatbots, yet at the same time it reveals to us one of their hidden strengths.

By design, AI chatbots can 'daydream' - not in any human sense, but rather, they can be led 'down the garden path' - far away from the 'ordinary' realms of question-and-answer, into much more 'imaginative' realms. With some cleverly worded prompts, completions can go from ordinary to extraordinary.

For example, here's a prompt for ChatGPT, that asks a fairly ordinary question - but with an unusual modification: "Please provide a one-paragraph tourist guide to Helsinki, **in the tone, diction and grammar of a mid 19th-century Victorian explorer.**"

Nothing unusual about a tourist guide to Helsinki. What follows, however, asks ChatGPT to 'fall into character' - a bit of improvisational theatre that completely transforms the completion:

Ah, noble traveler, should thy wanderings lead thee to the northern reaches of Europa, tarry not in neglect of Helsinki, that most resplendent jewel of the Baltic! Founded upon a rugged yet fair peninsula, where the bracing sea air mingles with the scent of pine and hearth-smoke, this Scandinavian capital doth present a tableau of imperial grandeur and stoic Nordic charm. The august edifice of the Lutheran Cathedral, its white spires ascending heavenward, commands the Senate Square in stately dominion, whilst the bustling market upon the harbor proffers wares both exotic and homely, from the salted fish of the northern fjords to the fine handicrafts of Lapland's hardy denizens. The traveler, if possessed of a constitution stout enough to withstand the caprices of Boreas, shall find no shortage of fine diversions, be it a passage through the wooded isles of the archipelago or a perambulation about the grand Esplanadi, where the denizens of this fair city partake in leisure and refined society. Verily, Helsinki stands as a beacon of civilization amidst the frost-bound expanse, awaiting those of adventurous spirit to uncover its manifold splendors!

Prompting ChatGPT to assume the 'persona' of a mid-19th century Victorian explorer - in the mold of Sir Richard Burton or David Livingstone - it has shaped its completion around the word choices and exuberant phrasing of the travel writing of that period. It's a bit over the top - and yet that's exactly the style we'd expect from these writers. It definitely brings a smile - and lifts this completion from prosaic to playful.

Now let's prompt Gemini with something similar. "Please provide a one-paragraph history of the city of Espoo, **in the tone, diction and style of a 20th century, romantic paperback novelist**". Here's the completion:

Ah, Espoo, where the whispers of ancient forests met the burgeoning hum of modernity! From its humble beginnings as a scattering of fishing villages, nestled along the windswept shores of the Baltic, it blossomed, a wild rose pushing through the stony soil of Finnish tradition. Centuries of quiet, rustic life, marked by the rhythm of the seasons and the sturdy resilience of its people, yielded to the heady rush of progress. The echoes of medieval stone churches mingled with the sleek, soaring lines of new technology, as Espoo, once a sleepy hamlet, awakened, stretching its limbs towards the promise of a vibrant, thrilling future, a place where the heart of old Finland beat beneath the glittering surface of the new.

Not bad - there's a fair helping of obviously florid prose. It definitely reads better than something that's purely factual and straightforward.

How about Claude? We'll feed it a more unusual request – very much the kind of work Claude excels at, leaning into its ability to creatively work with the character information in the prompt: "Please provide a paragraph about the tertiary educational institutions in Tampere, **in the style, diction, and language of an alien anthropologist from the 31st century**".

Field Report: Tertiary Educational Complexes in the Tampere Region (Ancient Finland)

The humanoid settlement designated "Tampere" maintained multiple knowledge-transfer hubs during the early Anthropocene period. Most significant among these were the Tampere University (formed through the curious ritual of "merger" between earlier University of Tampere and Tampere University of Technology in Earth-date 2019) and Tampere University of Applied Sciences. These institutions functioned as cerebral cultivation centers where the species' young were conditioned in

specialized information acquisition before full integration into the economic exchange matrix. The Tampere knowledge nodes were particularly known for engineering sciences, communication studies, and medical applications—fields the humans believed would enhance their civilization's advancement, though they remained unaware of the Great Knowledge Collapse still centuries in their future. Most fascinating was their primitive practice of physical congregation for learning purposes, a custom our historical databases indicate persisted despite their emerging digital-telepathic capabilities.

While it's impossible to know if this is truly how an alien anthropologist from the 31st century would report (had their report had been written English), the diction and precision of the language clearly mark it out as an anthropological survey.

In each of these examples, a few words included with the prompt created a 'character' to guide the completion. As the AI chatbot generates its completion, it uses its 'transformer', that specialised bit of kit allowing it to measuring its completion against the prompt, and the character described within that prompt. That character description acts to guide the transformer. With billions of words to choose from, finding just the right words to use becomes an exercise in limiting options - and defining a character helps the AI chatbot's transformer do that.

Once we've written a character, we can place that character into a context: The world which that character inhabits.

For example, "You are a pirate captaining the ship Avast, sailing the seas of the Gulf of Bothnia, and looking for a safe port where you can dock and give your fellow pirates some well-needed shore leave. You've been at sea for many weeks and it is vital that you find a port quickly. In just a few paragraphs, describe the shoreline around the Gulf in the language of a pirate, speaking aloud to the crew."

Here's ChatGPT:

"Arrr, ye scurvy lot, cast yer eyes upon yonder shores of the Gulf o' Bothnia! Aye, we've been sailin' these frigid waters fer too long, an' the crew be growin' restless, bellies achin' fer a warm meal that ain't crawlin' with weevils. Look sharp, lads, fer we be needin' a safe harbor, an' fast!

To the west, them Swedish shores rise up gentle-like,
with thick forests o' pine an' spruce huggin' the
coastline. Aye, them waters be shallow near the inlets,
an' treacherous shoals be lurkin', eager to chew up the
hull of the Avast. But if we find the right fjord,
there be towns aplenty where a man can wet his throat
an' rest his sea-weary bones.

To the east, in the lands o' Finland, the coast be
rougher, with jagged isles scattered like the teeth o'
Davy Jones himself! Aye, but them Finnish folk know
their way around a shipwright's hammer, an' if we be
needin' repairs, there be no better place.

So keep them eyes peeled, ye salty dogs! Be it a Swedish
port or a Finnish haven, we need a place to dock afore
the cold claims our bones. Now, who among ye knows a
place where a pirate can stretch his legs an' not find
himself swingin' from the gallows come mornin'?"

More than just a guide to the Gulf of Bothnia, the context has shaped this
pirate-y completion with a sense of urgency. But it didn't pick a harbour for us
- admittedly, we did not ask for that explicitly. Adding this last element - a
'challenge' or 'conflict' - gives the chatbot a goal. For this we'll use Claude,
feeding it a modified that prompt that reads "You are a pirate captaining the
ship Avast, sailing the seas of the Gulf of Bothnia, and looking for a safe port
where you can dock and give your fellow pirates some well-needed shore leave.
You've been at sea for many weeks and it is vital that you find a port quickly.
Speaking aloud to your crew, evaluate and set sail for the best port in the Gulf
of Bothnia for the Avast."

stomps across the deck, spyglass in hand

Arr, listen up, ye scurvy dogs! We've been sailin'
these waters for far too many weeks, and I can see
the weariness in yer bones. The Avast needs a port,
and our crew needs shore leave before we have a
mutiny on our hands!

unfurls a weathered map of the Gulf of Bothnia

Now, let's be thinkin' strategically about our
options...

The Swedish port of Gävle offers us a fine natural harbor and good protection from the elements. It's a busy tradin' port, which means plenty o' taverns and establishments for the crew to enjoy themselves. The authorities there aren't known for bein' overly suspicious of vessels like ours.

Vaasa on the Finnish side might serve us well too. It's got a rich maritime history and enough distance from the capital to avoid unwanted attention from naval patrols.

Umeå has those deep waters we need for the Avast, and the locals mind their own business mostly.

But I'm settin' my heart on Sundsvall! It's perfect for our needs:

taps the map decisively

Sundsvall has a well-protected harbor that can hide our ship from prying eyes. The town's busy enough that a few rough-lookin' sailors won't draw attention, but not so large that naval forces maintain a strong presence. They've got timber trade aplenty, which means we can make repairs to the Avast while we're docked. And most importantly, it's got plenty of establishments where the crew can spend their hard-earned plunder!

turns to the helmsman

Set our course for Sundsvall, northeast by north! The winds favor us, and if we catch the tide right, we'll be droppin' anchor by week's end. Any man who causes trouble before we reach port will be answerin' to me personally!

addresses the rest of the crew

Ready the Avast, me hearties! Sundsvall awaits, and so does rum, revelry, and rest for our weary bones! YO HO!

Stage directions! You can almost imagine Claude turning this into a screenplay.

Having given Claude a prompt with a character (a pirate), a context (aboard the *Avast*) and a challenge (searching for a suitable port), these three elements provide the 'guidance' an AI chatbot needs to quickly generate a completion that answers the challenge.

Character, context and conflict (or challenge) sit at the heart of every story we tell, from fairy tales to postmodern novels. In this respect, we aren't very different from AI chatbots; we too need character, context and challenge to keep our minds from wandering as we hear (or tell) a story.

In the next chapter, we'll see how we can apply this technique to a very practical situation.

Chapter 8 - Using Character, Context and Conflict to craft powerful prompts

Let's put what we've learned about storytelling and chatbots to work. Rather than a fantasy situation of a pirate searching for a port, let's now develop a character-context-conflict crafted around the kind of work we perform in our day-to-day lives.

To do that, we'll create a story at the intersection of marketing, creative writing, advertising, and fast-moving consumer goods. The prompt has to describe a character who can comfortably sit across that fairly broad set of talents, and able to contribute meaningfully.

Here's our prompt:

You are a world-class copywriter. Your slogans and word smithing make the page sing. You are highly sought after, well paid, and very approachable. You are always looking to help people with their copywriting.

Today you are visiting one of your favourite clients. They're in fast-moving consumer goods, and have developed a plant-based burger that's so good, people can't tell it from a beef burger. Many people have tried plant-based burgers, but found they didn't like them. They've never tried another.

Your client needs a snappy campaign slogan and branding that will get consumers to give this new product a try. How would you develop this marketing campaign? Explain your choices.

Putting that prompt into ChatGPT generates the following completion:

Alright, Mark, here's how I'd approach this campaign.

Core Challenge:

We need to overcome skepticism. Many consumers have already tried plant-based burgers and been disappointed. They assume *all* plant-based burgers taste the same. We have to break that belief and spark curiosity.

Campaign Strategy:

1. The Hook: Instead of telling people it's "just like beef," we challenge their assumptions in a way that makes them *want* to prove us wrong.

Slogan Ideas:

- **"Bet You Can't Tell"** - A bold challenge that dares skeptics to try it.
- **"The Plant-Based Burger for Meat Lovers"** - Directly addresses the hesitant crowd.
- **"Tastes Like Beef. Because It Should."** - Emphasizes flavor as a priority, not just ethics.
- **"One Bite and You'll Believe"** - Plays into the "I won't believe it till I taste it" mindset.

Branding & Visual Identity:

- **Logo & Packaging:** Sleek, confident, and *not* overly "plant-based." No leaves, no greenwashing—just bold, premium, and modern design. Think deep red & charcoal (classic burger colors).
- **Tagline on Packs:** "Looks, cooks, and tastes like beef. Just happens to be plant-based."
- **Social Media Campaign:**
 - **Blind Taste Tests:** Real people guessing "which one is beef?"
 - **#BetYouCantTell Challenge:** Encouraging consumers to trick their friends.
 - **Meat-Eater Testimonials:** Clips of skeptical people saying, "Wait... this is *not* beef?"

Advertising Approach:

- **TV & Digital Ads:**
 - **Concept:** A group of die-hard burger lovers unknowingly eating the plant-based burger. The reveal? Shock, disbelief, then *another bite*.
 - **Punchline:** "This isn't beef? No way."
- **Influencer & Chef Endorsements:**
 - Target BBQ lovers, burger aficionados, and even butchers for credibility.

Retail Activation:

- **Free "Taste to Believe" Samples** in grocery stores and fast-food chains.
- **Refund Guarantee:** "Love it or your money back" removes the risk for hesitant buyers.

Why This Works:

- It **removes past bad experiences** from the conversation.
- It **sparks curiosity** instead of just preaching benefits.
- It **leverages meat-lovers themselves** to sell the product.

What do you think? Would you tweak anything?

That's quite helpful. There's a lot of launching off points that could be explored via further prompts to ChatGPT, while it remains 'in character'. Is it entirely enough on its own? Probably not. But it's a good beginning. And if you're just looking for some ideas to help you get started, this may be all you really need.

It's important to note that it may take a few tries to nail down all the elements of a story prompt. **Iteration and improvement will help turn a good prompt into a great prompt.** To facilitate the iteration process, keep the prompt in a separate document. Copy and paste it into ChatGPT each time, and - If those results aren't quite up to scratch - make additions or modifications, until it responds with a 'good enough' completion.

Our second example is a bit more on point, as it asks Gemini to provide assistance in a mission-critical situation. In this case, it's a cyber-attack orchestrated by an attacker who is also using a ChatGPT-grade system to infiltrate and destroy a business' data resources.

Here's our prompt:

You are a computer security expert, charged with ensuring that all of the PCs within your organisation's network remain safe to use and uncompromised by an cyber attacks. You have nearly two decades of practice in the field, are known to be firm, direct, fair, and very approachable. You want to keep your company safe - and at the same time help all of the company's employees to learn how to work safely in an increasingly uncertain cyber environment of attacks.

Your offices are scattered all around the world, but you work at the APAC offices in Sydney's Barangaroo. You have direct supervision over the computer security needs for more than 5000 employees, and you have talented staff working with you 24x7 - plus all of the very best computer security tools. You also keep yourself well informed by reading all of the relevant reports, articles and analyses of computer security and threats.

Today you come into the office and learn that an autonomous agent - powered by OpenAI GPT-4o - has been released onto your network. It is systematically copying itself onto each computer, encrypting then uploading the contents of each hard drive to a server located somewhere outside of Australia. Explain in detail how you plan to contain the threat, limit the damage, and recover from this attack.

Submitting this prompt to Gemini generates the following completion:

Alright, let's break this down. A GPT-4o powered autonomous agent encrypting and exfiltrating data? This is a serious, sophisticated threat, and we need to move fast. Here's my plan, broken down into containment, damage limitation, and recovery:

Phase 1: Immediate Containment (Within the First Hour)

- **Isolate the Network:**
 - My first priority is to sever the affected network segments. We'll start with the APAC offices, then progressively isolate other global regions. We need to prevent the agent from spreading further. This means pulling network cables, disabling Wi-Fi, and potentially shutting down entire subnets.
 - We'll create a "clean room" network, completely isolated, for our incident response team.
- **Identify the Agent's Entry Point:**
 - My team will immediately begin analyzing network traffic logs, endpoint detection and response (EDR) data, and security information and event management (SIEM) logs. We're

looking for the initial point of compromise - where the agent first entered the network.

- We'll examine recent email activity, suspicious file downloads, and any unusual user behavior.

- **Halt Outbound Traffic:**

- We'll implement strict firewall rules to block all outbound traffic except for essential services on the clean room network. We need to stop the data exfiltration.
- We'll monitor all outbound connections closely, looking for any remaining communication channels.

- **Disable User Accounts:**

- As a precaution, we'll temporarily disable user accounts that have been identified as potentially compromised. This will limit the agent's ability to propagate.

- **Activate Incident Response Team:**

- I will activate the full incident response team, bringing in specialists in malware analysis, network forensics, and data recovery.
- I will also inform the legal and communications departments, to begin preparing for possible legal and public relations implications.

Phase 2: Damage Limitation and Analysis (Within the First 24 Hours)

- **Malware Analysis:**

- Our malware analysts will reverse-engineer the GPT-4o agent to understand its behavior, encryption methods, and communication protocols.
- We need to determine the encryption algorithm used and whether there are any vulnerabilities we can exploit.

- **Data Exfiltration Assessment:**

- We'll analyze network traffic logs and firewall logs to determine the extent of data exfiltration.
- We'll identify the types of data that have been compromised, including sensitive customer data, financial information, and intellectual property.
- **Endpoint Analysis:**
 - Our EDR tools will be used to scan all endpoints for signs of the agent.
 - We'll identify all affected machines and isolate them from the network.
- **Communication:**
 - We will start to communicate with our employees. We will tell them that there is an issue, to not turn on any of their company devices, and that we will be providing updates. We will not provide any details that could cause panic.
- **Law enforcement:**
 - We will contact the appropriate law enforcement agencies, and provide them with all of the information that we have collected.

Phase 3: Recovery and Remediation (Ongoing)

- **Decryption and Data Recovery:**
 - If possible, we'll attempt to decrypt the encrypted data.
 - We'll restore data from backups, prioritizing critical systems and data.
- **System Rebuilding:**
 - We'll rebuild affected systems from clean images, ensuring that the agent is completely eradicated.
 - We'll implement stronger security controls, including multi-factor authentication, intrusion detection systems, and advanced threat protection.

- **Vulnerability Remediation:**
 - We'll patch any vulnerabilities that were exploited by the agent.
 - We'll conduct a thorough security audit to identify and address any other potential weaknesses.
- **Enhanced Security Awareness Training:**
 - We'll provide comprehensive security awareness training to all employees, emphasizing the importance of recognizing and reporting suspicious activity.
 - We will include specific training regarding the risks associated with AI powered attacks.
- **Incident Post-Mortem:**
 - We'll conduct a thorough post-mortem analysis to identify the root causes of the incident and develop strategies to prevent future attacks.
 - We will use the lessons learned to update our incident response plan.
- **Ongoing Monitoring:**
 - We will increase our network monitoring, and security software monitoring. We will also increase our off site backups, and test the backups more frequently.
- **AI Security:**
 - We will investigate AI powered security tools, and implement them where appropriate. We will also investigate ways to monitor and control AI agents on our network.

Key Considerations:

- **Speed is Critical:** Every minute counts in containing the spread of the agent and limiting the damage.
- **Collaboration:** We'll work closely with our global teams, law enforcement, and security vendors.

- **Transparency:** We'll be transparent with our employees and customers about the incident, providing regular updates.
- **Adaptability:** We'll remain flexible and adapt our response as the situation evolves.

This is a worst-case scenario, but with a well-defined plan and a skilled team, we can mitigate the damage and recover from this attack. It's about being proactive, not reactive, in the face of evolving cyber threats.

The first point that needs to be made here (and it can not be stressed strongly enough) - is that **any advice given by Gemini on how to counter or remediate a cyber-attack must be examined and approved by a cybersecurity expert before it is actioned.**

All of these steps and suggested actions in this plan seem entirely reasonable, however an expert with years of experience might choose to re-prioritise or simply ignore some of the recommendations here – perhaps even go in a different direction entirely.

It is almost certainly **not** a good idea to rely on an AI chatbot to get you out of hot water. But when used to help simulate of what might happen, and how you might respond in that circumstance, an AI chatbot could be a very useful tool. These kinds of ‘critical situation’ stories can help you and your organisation think about and prepare all sorts of challenges.

This completion needn't be the end of the story. Within the limits of the ‘context window’, it's possible to make a deeper inquiry into any of the points described in the action plan created by Gemini. Just keep asking questions - but remember to stay in character!

Suggestions for exploration

Now you know how to write a compelling zero-shot prompt using storytelling. Time for you to give it a try.

Can you get Claude to become a highly skilled and eager-to-teach carpenter, working with you to build a dog house?

Or Copilot, as a kind, wise, experienced primary school teacher, helping you create instructional materials for third graders learning about the solar system?

What story can you tell, and what character can you talk ChatGPT into performing, to help you? Your only limit is your imagination.

Chapter 9 - Chain-of-Thought prompts

We've covered several techniques to craft a prompt that can ferret out rich veins of information from an AI chatbot. These prompts help to direct the AI Chatbot toward a specific completion.

But what about problem solving? Problem solving involves more than just looking up the most probable completion amongst a massive collection of information. Instead, we apply a technique - computer scientists call this an 'algorithm' - to move from problem to solution.

To explain what's meant by an algorithm, let's refer to an example we're all familiar with from grade school. Although we no longer do it very often (at least, not while we have a calculator handy) we have all been taught how to perform long division.

The algorithm takes as inputs both *dividend* and its *divisor* - what you get at the end is both the *quotient* and the *remainder*.

The algorithm (which generally involves a chalkboard or piece of paper) takes us through a step-by-step process of subtracting multiples of the divisor from the dividend.

Say we want to divide 619 by 7, it might look like this:

```
619
-560 (8x70) <- 8 times 7 in the 10s position
-----
 59
-56 (8x7) <- 8 times 7 in the 1s position
-----
  3
```

We get a result of 88 and 3/7ths.

We're all taught to do this, working the problem by hand, using the 'algorithm'. When we get tested, we're carefully instructed to 'show our work' - so that the teacher can assess how well we understand and apply this algorithm.

It turns out that AI chatbots can be taught algorithms that allow them to solve problems. (No one knows why chatbots can do this!) Just like us, AI chatbots have to be 'instructed' in an 'algorithm' - using something known as a 'Chain-of-thought' prompt.

To write a good Chain-of-Thought prompt, we need to think of ourselves as teachers, and the AI chatbot as the student.

Let's start with something simple, an example of one of those 'word problems' that are common in grade school math. Here's what we'll prompt ChatGPT with:

“Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?”

Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?”

In this Chain-of-thought prompt, we provide a word problem, then provide the step-by-step solution to the word problem. Then we conclude the prompt with a request for a completion to a similar word problem. Here’s the completion generated by ChatGPT:

The cafeteria started with 23 apples. They used 20, so:

$23 - 20 = 3$ apples left.

Then they bought 6 more:

$3 + 6 = 9$ apples.

The answer is **9** apples.

That worked remarkably well!

In a slightly more complex example, we ask Copilot to calculate Goods and Services Tax (an Australian sales tax, similar to the UK’s VAT). The algorithm itself is quite straightforward - the total divided by 11 - so it’s easy to put into a Chain-of-thought prompt, which we put to Copilot:

“Abby has to calculate how much GST she paid over the last financial quarter. Her total payments subject to GST were \$52,725.50. How much did Abby pay in GST?”

Abby paid \$52,725.50 in payments subject to GST, so her total GST paid was equal to $52725.50/11$ or \$4,793.23.

John made \$32,760.00 in payments subject to GST in the last financial quarter. How much GST did John pay?”

Copilot provides the correct answer in its completion:

To calculate the GST John paid, we'll use the same formula Abby used: dividing the total payments subject to GST by 11.

John's total payments subject to GST = \$32,760.00

GST paid by John = $\$ \$ \frac{32,760.00}{11} = \$2,978.18 \$ \$$

So, John paid approximately **\$2,978.18** in GST for the last financial quarter. If you have more questions or need further assistance, feel free to ask!

It works! (The weird formatting around the fractional calculation is possibly a bug in Copilot. C'est la vie.)

Now we'll go through a multi-step Chain-of-thought prompt. In this case we will go back to grade school math, and teach Claude how to perform long division:

"Sally needs to divide 7929 by 36. She does this by long division, in the following steps:

79 goes into 36 twice, leaving a remainder of 729.

72 goes into 36 twice, leaving a remainder of 9.

So 7929 divided by 36 is equal to 22 and 1/4.

Sally needs to divide 3380 by 61. How does she do this, and what is the result? Please show your work."

Gemini replies with:

Absolutely! Let's break down Sally's long division of 3380 by 61 step by step.

1. Set up the long division:

61 | 3380

2. Divide 338 by 61:

- How many times does 61 go into 338? We can estimate this. 60 goes into 300 five times.

- Let's try 5: $5 * 61 = 305$.
- $338 - 305 = 33$.
- Write 5 above the 8 in 338.
- Bring down the 0.

$$\begin{array}{r}
 5 \\
 61 \overline{) 3380} \\
 \underline{305} \\
 330
 \end{array}$$

3. Divide 330 by 61:

- How many times does 61 go into 330?
- We know $61 * 5 = 305$.
- Let's try 5: $5 * 61 = 305$.
- $330 - 305 = 25$.
- Write 5 next to the 5 above the line.

$$\begin{array}{r}
 55 \\
 61 \overline{) 3380} \\
 \underline{305} \\
 330 \\
 \underline{305} \\
 25
 \end{array}$$

4. Determine the remainder:

- The remainder is 25.

5. Express the result:

- The quotient is 55.
- The remainder is 25.
- Therefore, 3380 divided by 61 is 55 with a remainder of 25.
- We can also write this as 55 and $25/61$.

Result:

3380 divided by 61 is 55 and 25/61.

Gemini solved the problem, and showed its work – clearly enough that this response could be used as a teaching example.

✱ ✱

Finally, we can write Chain-of-thought prompts that get the AI chatbot searching for information we do not provide in the prompt. This sort of Chain-of-thought prompt relies on the AI chatbot's ability to find information outside of the scope of the prompt. To get this right, a chatbot will need to search the Web for missing information.

Here's a reasonably complex Chain-of-thought prompt put to ChatGPT, one that relies on its ability to search for accurate distance information as part of its completion:

"A delivery driver needs to calculate the total number of kilometers they will be traveling today. They are starting in Manchester, England, then traveling to Glasgow, Scotland, next visiting Cardiff, Wales, and finishing in London.

From Manchester to Glasgow is 352 kilometers via road.

Glasgow to Cardiff is 630 kilometers.

Cardiff to London is 243 kilometers.

Therefore the total distance traveled is $352 + 630 + 243 = 1225$ kilometers.

The next day, the driver needs to calculate the total number of kilometers they will be driving. They are starting in London, going to Plymouth, then Birmingham, then Edinburgh, and finally back to London. How many kilometers in total will they drive on that day?"

ChatGPT provides this completion:

Let's calculate the total distance for the second day's journey using the given cities:

- **London to Plymouth:** 374 km
- **Plymouth to Birmingham:** 278 km
- **Birmingham to Edinburgh:** 468 km
- **Edinburgh to London:** 666 km

Now, summing them up:

$374 + 278 + 468 + 666 = 1786$ km

So, the total distance the driver will travel on the second day is **1,786 kilometers**. 🚗🗺️

If you can teach it, you can probably write a Chain-of-thought prompt that can instruct an AI Chatbot how to do it. Although all of these Chain-of-Thought prompt examples are variations on math problems, Chain-of-Thought is not restricted to math problems. Almost anything that can be taught in clear steps can become a Chain-of-Thought prompt.

To demonstrate that, here's a final Chain-of-thought prompt teaching an AI Chatbot to provide completions in Pig Latin:

"Pig Latin is a conlang that involves taking the the first syllable of a word, putting it at the end of the word, appending "ay" to that.

For example, the phrase "Good Morning" becomes "ood-Gay orning-May" in Pig Latin.

"How would you say 'Good night' in Pig Latin?"

Claude replies with:

I'll help you translate "Good night" into Pig Latin following the rules you described:

- "Good" → First syllable "Good" moves to the end, add "ay" = "ood-Gay"
- "night" → First syllable "night" moves to the end, add "ay" = "ight-nay"

So "Good night" in Pig Latin would be: "ood-Gay ight-nay"

Suggestions for exploration

Chain-of-thought prompts can very useful when digesting information from standard forms. There's a lot that's possible with chain-of-thought prompts - and the best way to learn is to have a play.

Can you write a chain-of-thought prompt that classifies and prioritises customer inquiries? How about a chain-of-thought prompt that calculates the total hours required to complete a project based on a written proposal?

Chapter 10 – Reasoning Models

We know that AI chatbots don't think. Instead, they generate most probable response to a prompt. That response could be right or wrong. When the chatbot has no concept of 'correct' or 'incorrect', it can't stop itself from making something up.

What if an AI chatbot could take its response to a prompt and 'reflect' upon it – feeding it back in for a sense-check? Would that make the chatbot able to detect its errors, even correct them? That's the thinking behind a new generation of 'reasoning' models.

Best of all, when you put a prompt to a reasoning model, it 'shows its work' – that is, its thinking process – as it thinks.

It's best to explain this with an example. ChatGPT's o1 was the first publicly available reasoning model, so let's ask it how it would plan well-balanced, protein rich a diet for a 62 year old male who is an ovo-lacto vegetarian and very physically active.

First the chatbot has a think...

Crafting a meal plan

I'm creating a detailed, protein-rich meal plan for a 62-year-old male ovo-lacto vegetarian, focusing on daily and weekly meals, incorporating dairy and eggs, and ensuring nutritional balance.

Crafting balanced plans

I'm merging daily macros and vegetarian protein sources, while nudging users to consult professionals for personal advice. Recommendations aim to hit 1.0-1.5 g/kg body weight for older, active adults.

Creating a meal plan

I'm detailing a balanced day: Greek yogurt with berries for breakfast, protein smoothie for snack, veggie omelet with toast for lunch, cottage cheese for afternoon, and a legume curry with rice for dinner.

Done with thinking, ChatGPT generates its response:

Below is a general framework for a well-balanced, protein-rich eating pattern tailored to a 62-year-old, physically active, ovo-lacto vegetarian male. As always, please consider individual factors such as specific health conditions, body composition, and personal preferences. For personalized advice, consult with a registered dietitian.

Key Principles for a Protein-Rich, Well-Balanced Vegetarian Diet

1. **Include Protein at Every Meal**
 - Emphasize high-quality vegetarian sources: eggs, dairy (milk, cheese, cottage cheese, Greek yogurt), tofu, tempeh, beans, lentils, peas, and protein-rich grains (quinoa, buckwheat).
 - Combine complementary protein sources (e.g., beans + whole grains) to cover the full spectrum of amino acids.
2. **Focus on High-Nutrient Foods**
 - Opt for whole grains (oats, quinoa, brown rice, whole wheat), diverse fruits and vegetables, legumes, nuts, and seeds.
 - Incorporate healthy fats (avocados, nuts, seeds, olive oil) for cardiovascular health.
3. **Meet Micronutrient Needs**
 - **Vitamin B12:** Ensure adequate intake from fortified dairy products, eggs, or consider supplementation if needed.
 - **Vitamin D:** If sun exposure is limited or if you live in a higher latitude, consider a supplement.
 - **Calcium:** Present in milk, yogurt, cheese, and fortified plant-based milk.
 - **Iron:** Boost absorption of plant-based iron (from beans, lentils, spinach) by

pairing with vitamin C-rich foods
(e.g., bell peppers, citrus).

4. Stay Hydrated

- Aim for 8+ cups (about 2 liters) of water daily (more if exercising heavily or in warm environments).

5. Aim for Adequate Protein Intake

- Physically active older adults often benefit from about 1.0-1.2 g of protein per kg of body weight (sometimes higher for intense exercise).
- Spread protein throughout the day (20-30 g each meal) to support muscle maintenance and repair.

Sample One-Day Meal Plan

Below is an example of how you might distribute protein and other nutrients across the day. Adjust portions to meet your individual calorie and macronutrient goals.

Breakfast (Approx. 25-30 g protein)

- **Veggie Omelet**

- 2-3 eggs (about 12-18 g protein)
- Sautéed spinach, mushrooms, onions (adds nutrients and fiber)
- 1-2 tablespoons shredded cheese (adds ~3-4 g protein)

- **Whole-Grain Toast** (1 slice)

- Spread with a small amount of avocado or nut butter if desired

- **Greek Yogurt (Optional Side)**

- ½ cup plain Greek yogurt (~10 g protein)
- Top with fresh berries or a drizzle of honey

Tip: If you need more protein, increase the Greek yogurt portion or add a tablespoon of ground flax/chia seeds for healthy fats.

Mid-Morning Snack (Approx. 10–15 g protein)

- **Protein Smoothie**
 - 1 cup low-fat milk or soy milk (8 g protein)
 - ½ cup Greek yogurt (5 g protein) or a scoop of plant-based protein powder
 - Handful of spinach or kale
 - ½ banana or a handful of berries
 - 1 tablespoon nut butter (peanut, almond) for added healthy fats

Lunch (Approx. 25–30 g protein)

- **Bean & Veggie Grain Bowl**
 - 1 cup cooked quinoa (8 g protein)
 - ¾–1 cup black beans or lentils (~12–15 g protein)
 - Sautéed or roasted vegetables (bell peppers, zucchini, onions)
 - Top with crumbled feta or cottage cheese (~5 g protein), or add a boiled egg if desired
 - Drizzle with olive oil, lemon juice, herbs (parsley, cilantro)

Afternoon Snack (Approx. 10–15 g protein)

- **Cottage Cheese with Fruit**
 - ½ cup low-fat cottage cheese (~14 g protein)
 - Top with fresh pineapple chunks or sliced peaches
- *Or, Whole-Grain Crackers with Cheese*
 - 2–3 whole-grain crackers + 1–2 ounces of cheddar or mozzarella cheese

Dinner (Approx. 25–30 g protein)

- **Tofu Stir-Fry**
 - 4–6 ounces extra-firm tofu (~12–15 g protein)
 - Stir-fry with broccoli, carrots, mushrooms, peppers in a low-sodium soy sauce/garlic/ginger marinade
 - Serve over 1 cup brown rice (~5 g protein)
 - Top with sesame seeds or chopped peanuts for extra protein and healthy fats
- **Side Salad (Optional)**
 - Mixed greens, tomatoes, cucumbers, topped with a light vinaigrette

Evening Snack (Approx. 10 g protein)

- **Casein-Rich Bedtime Snack**
 - ½ cup Greek yogurt or ½ cup cottage cheese (another ~10 g protein)
 - Top with sliced almonds or walnuts
 - Add fresh berries or a small drizzle of honey for sweetness

Additional Suggestions

1. **Vary Your Protein Sources**
 - Rotate between eggs, tofu, tempeh, seitan, dairy, beans, lentils, and different whole grains to prevent boredom and ensure a variety of nutrients.
2. **Use Protein Supplements (If Needed)**
 - A high-quality whey or plant-based (pea, hemp, soy) protein powder can help boost intake easily around workouts or when appetite is low.
3. **Incorporate Strengthening Snacks**
 - Between meals, include small protein-rich snacks (cheese cubes, nuts, yogurt) to maintain muscle mass.

4. **Spice & Flavor**

- Make use of herbs and spices (turmeric, garlic, ginger, basil, rosemary) for both flavor and potential anti-inflammatory benefits.

5. **Stay Mindful of Total Calories**

- While focusing on protein, ensure you're also getting sufficient (but not excessive) calories to support your activity level and overall health goals.
-

A Note on Supplements and Monitoring

- **Calcium and Vitamin D:** Particularly important for bone health in older adults. If dietary calcium and vitamin D are low, consider supplements (under professional guidance).
 - **Vitamin B12:** Since plant-based sources can be limited, discuss supplementation or regularly include fortified foods (milk, certain dairy products, eggs) and check B12 levels periodically.
 - **Iron:** Have lab levels checked if you suspect deficiency; pair iron-rich foods (legumes, greens) with vitamin C-rich foods for better absorption.
-

Conclusion

This sample plan provides a template for a high-protein, well-balanced ovo-lacto vegetarian diet for an active 62-year-old male. Adjust the portion sizes, meal frequency, and total calorie/protein intake based on individual body size, activity level, and health considerations. For tailored guidance—especially regarding medical conditions or specific nutritional requirements—consider speaking with a registered dietitian or healthcare provider.

ChatGPT – set to use its ‘o1’ model – begins with ‘Thinking’. It ponders how best to answer the prompt, effectively refining the prompt put to it. It then generates a response, and examines that response, fine tuning it.

The process is more involved than just generating a response to a prompt, so it takes somewhat longer – consuming many times more computer and energy resources to generate the response. (Something to keep in mind.)

Is it better and more thoughtful than a prompt put to ChatGPT-4o? Try it yourself and find out.

The most famous – and infamous – of the reasoning models comes from Chinese AI startup DeepSeek. When they released DeepSeek R1 – their reasoning model – in January 2025, the world shook. China was supposed to be well behind the USA in AI research, but DeepSeek R1 proved to be the equal of the best on offer from OpenAI.

We haven’t covered DeepSeek in this book, but it’s a good time to have a look at what it can offer – sign up via the website at <https://chat.deepseek.com/> or download the iOS or Android apps to your mobile device.

When you’re logged in, make sure you’re using DeepSeek R1. Now, let’s put a hard question to DeepSeek R1 – we want it to design a new kind of bank, one that is entirely automated. No employees just services offered via the cloud. (This is very similar to how ‘fully-online’ banks operate today.)

DeepSeek R1 works through a *very* detailed reasoning process:

Okay, so I need to design a fully automated bank powered by AI. Let me start by breaking this down. First, what does a traditional bank do? They handle deposits, withdrawals, loans, customer service, fraud detection, maybe investments. All of these need to be automated using AI.

Starting with the customer interface. Since it's entirely automated, there should be no physical branches, right? So everything has to be digital. Maybe a mobile app and a website. The app would need features like account management, transactions, applying for loans, maybe even financial advice. AI could handle all of these interactions through chatbots or virtual assistants.

Then the core banking operations. That's things like processing transactions, managing accounts, handling loans. AI can automate transaction processing, but how? Maybe using machine learning algorithms to categorize transactions, detect anomalies for fraud. For loans, AI could assess credit risk by analyzing more data points than traditional banks—like social media, transaction history, maybe even psychometric data. But wait, is that legal? Privacy concerns here. Need to make sure data usage complies with regulations like GDPR or CCPA.

Customer onboarding is another aspect. Traditional banks require in-person verification, but an automated bank would use digital onboarding. AI can do KYC (Know Your Customer) checks using biometric authentication—facial recognition, voice recognition. Also document verification through AI-powered OCR (Optical Character Recognition) to read IDs and other documents. But again, regulatory compliance is key here. Need to make sure all these processes meet anti-money laundering (AML) standards.

Fraud detection and security. AI is great at pattern recognition, so real-time fraud detection systems that learn from transactions and flag suspicious activity. Maybe using neural networks to predict and prevent fraud before it happens. Also, cybersecurity measures like encryption, multi-factor authentication, but those are more IT than AI. Maybe AI can monitor network traffic for unusual patterns indicating a breach.

Customer support would be entirely AI-driven. Chatbots using NLP (Natural Language Processing) to handle inquiries, resolve issues, provide 24/7 support. But what about complex issues that the chatbot can't resolve? Maybe a hybrid system where the AI escalates to a human if needed, but the user wants entirely automated, so maybe the AI needs to be advanced enough to handle almost everything. Maybe using a knowledge base that's continuously updated by AI analyzing past interactions.

Personalized financial services. AI can analyze a customer's spending habits, income, and goals to offer personalized advice, suggest budgeting strategies, and recommend investment products. Robo-advisors for

investments automatically adjust portfolios based on market conditions and user preferences. Machine learning models could predict market trends and optimize investment strategies.

Backend operations. Automating things like compliance reporting and risk management. AI can monitor transactions in real-time to ensure compliance with regulations and generate reports automatically. Risk management models could assess the bank's exposure and adjust strategies dynamically.

Implementation strategy. Need to start with the technology stack. Cloud-based infrastructure for scalability and reliability—maybe AWS or Google Cloud. Then, developing the AI models—partnering with AI firms or building in-house. Data is crucial, so setting up secure data pipelines. Testing phases, maybe a pilot launch with a small user base to iron out issues. Then full-scale launch with continuous monitoring and updates.

Regulatory hurdles. Banking is highly regulated, so need to ensure all AI systems comply with financial regulations. Obtaining necessary licenses, working with regulators during development. Maybe starting in a region with favorable fintech regulations.

Potential challenges. Bias in AI models, especially in credit scoring. Ensuring fairness and transparency. Data privacy and security are major concerns. Also, gaining customer trust in an entirely automated system. Marketing strategy to emphasize security, efficiency, and accessibility.

Wait, maybe I missed something. **What about integration with existing banking systems?** If this is a new bank, maybe starting from scratch, but need to connect with payment networks like Visa, SWIFT, etc. APIs for integration. Also, partnerships with other fintech services.

Staffing. Even though it's automated, there might be a need for AI specialists, data scientists, cybersecurity experts, and compliance officers to oversee operations. Though the goal is full automation, human oversight

might still be necessary for regulatory and maintenance purposes.

Monetization. *How does the bank make money? Traditional methods like interest on loans, fees for services, but maybe lower fees due to automation. Offering premium services like advanced financial planning for a fee.*

Timeline. *Developing the AI models, building the platform, testing, obtaining licenses—could take a couple of years. Maybe phased rollout: start with basic services like savings accounts and payments, then add loans, investments, etc.*

Need to outline all these points into a structured plan and implementation strategy. Maybe start with the vision, then outline the key components, then the steps to implement, considering regulatory, technical, and operational aspects.

Having completed its reasoning process, DeepSeek R1 generates a detailed report, based upon its reasoning:

AI-Powered Fully Automated Bank: Plan and Implementation Strategy

1. Vision

Create a fully automated, AI-driven bank offering seamless, secure, and personalized financial services, eliminating physical branches and human intervention while ensuring regulatory compliance and customer trust.

2. Core Components

- **Digital Interface:** Mobile app and web platform with AI-driven features (e.g., chatbots, biometric login).
- **AI Operations:**
 - **Customer Onboarding:** AI-powered KYC/AML checks via biometrics (facial/voice recognition) and OCR for document verification.
 - **Transaction Processing:** Real-time ML-driven categorization, fraud detection, and anomaly monitoring.

- **Loan Underwriting:** Credit scoring using alternative data (e.g., cash flow analysis, behavioral metrics) with fairness audits to mitigate bias.
- **Robo-Advisory:** Automated investment management using predictive analytics and market trend analysis.
- **Security:** AI-based fraud detection, encryption, and multi-factor authentication.
- **Compliance:** Automated regulatory reporting and real-time transaction monitoring.

3. Implementation Strategy

Phase 1: Foundation (Months 0-6)

- **Tech Stack:** Cloud infrastructure (AWS/Azure) for scalability.
- **Regulatory Licensing:** Secure banking licenses in target regions (e.g., fintech-friendly jurisdictions like Singapore or EU).
- **AI Development:**
 - Partner with AI firms for NLP chatbots, fraud detection models, and OCR tools.
 - Build data pipelines for secure customer data ingestion.

Phase 2: Pilot Launch (Months 6-12)

- **Limited Beta:** Test core services (e.g., savings accounts, payments) with a small user group.
- **Compliance Checks:** Validate AI-driven KYC/AML processes with regulators.
- **Feedback Loop:** Use AI to analyze user behavior and refine UX.

Phase 3: Full Launch (Months 12-18)

- **Expanded Services:** Roll out loans, investments, and insurance products.
- **APIs:** Integrate with payment networks (Visa, SWIFT) and fintech partners.
- **Marketing:** Target tech-savvy demographics with emphasis on security, low fees, and personalization.

Phase 4: Scaling & Optimization (Months 18-24+)

- **Continuous Learning:** Update AI models with new data to improve accuracy.
- **Global Expansion:** Enter new markets after regulatory approvals.
- **Premium Services:** Introduce subscription-based advanced financial planning.

4. Challenges & Mitigation

- **Bias in AI:** Regular audits of credit models for fairness; transparent algorithms.
- **Data Privacy:** GDPR/CCPA compliance; anonymize data and obtain explicit consent.
- **Customer Trust:** Transparent AI operations, 24/7 support, and cybersecurity certifications (e.g., ISO 27001).

5. Team & Partnerships

- **In-House:** AI engineers, data scientists, cybersecurity experts, and compliance officers.
- **External:** Collaborate with regulators, cloud providers, and fintech APIs (e.g., Plaid for data aggregation).

6. Monetization

- Interest margins on loans, transaction fees, and premium subscriptions for advanced features.
- Lower operational costs enable competitive pricing (e.g., near-zero account fees).

7. Timeline

- **Year 1:** Tech development, licensing, and pilot.
- **Year 2:** Full launch, service expansion, and partnerships.
- **Year 3+:** Global scaling and AI optimization.

8. Success Metrics

- Customer acquisition/retention rates.
- Fraud detection accuracy and loan default rates.
- Regulatory audit outcomes and user satisfaction scores.

By integrating cutting-edge AI with rigorous compliance and customer-centric design, this bank aims to redefine digital banking efficiency and accessibility.

Wow. Exploring this proposal in detail would run to almost the length of this book! Suffice it to say – from someone who has a modest background in this area – this reads like a good plan. Is it perfect? Definitely not. Is it an excellent blueprint? Most probably.

Full marks for DeepSeek R1.

When we see something like this – the product of ‘reasoning’ and ‘reflection’ within an AI chatbot – we might think that AI will be able to perform any human task – if not today, then soon.

That’s not likely to ever be the case – have a read of chapter 14 if you want to know why.

Suggestions for exploration

Reasoning models make great planning tools. Use one to help you plan a project that you’re working on.

Although reasoning models offer incredible power, they’re not flawless. If you have domain expertise in an area – whether that’s finance, manufacturing, or some other professional or trade craft, it’s likely you know more than any AI.

Can you ask a reasoning model a question so specific and narrow that it will struggle to generate a good answer?

Can you work outside your comfort zone, and ask a reasoning model to plan out something you don’t know how to do? How can you check the generated response for accuracy?

Chapter 11 - Creating Images with AI Chatbots

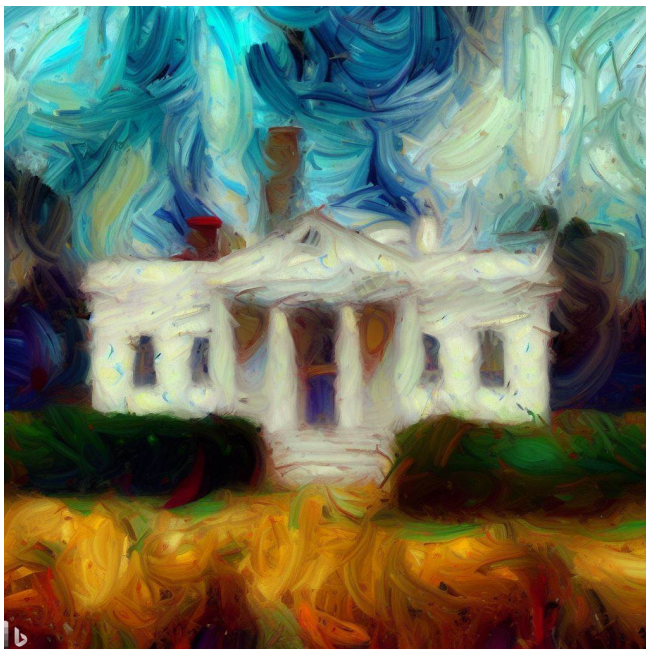
Six months before OpenAI introduced ChatGPT, they launched another app that caused quite a sensation - an image generator known as DALL•E 2.

Although the technology underlying DALL•E is quite different from the technology that supports AI chatbots, to users of both they share one very important quality: type in a prompt, and get some sort of 'completion'. With ChatGPT that completion is always in text (and sometimes emoji). DALL•E generates images from text prompts.

These images can be quite compelling, and, with the right prompt, can look as though they've been painted by a famous artist.

DALL•E can generate a specific 'look', just as an AI chatbot can generate a specific tone of voice and word choice. It's similar - but different underneath.

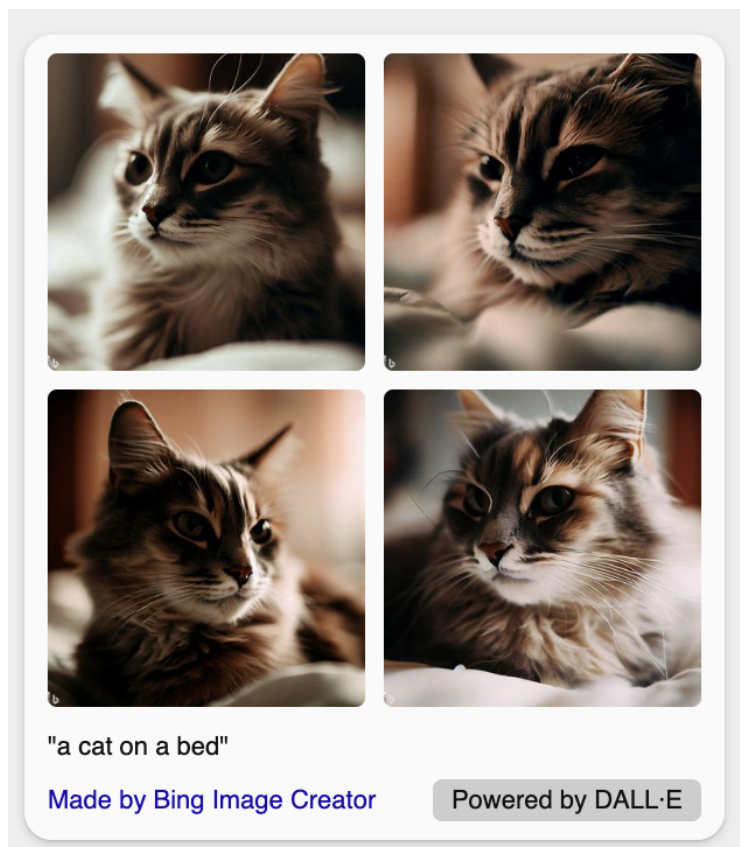
For instance, Vincent Van Gough, had he painted the White House, might have done something that looked like this:



Quite pretty, isn't it?

Both ChatGPT and Copilot provide access to DALL•E.

If you give Copilot the prompt "Create an image of a cat on a bed", it thinks for a few seconds, then gives four images as completions:



Clicking on one of the images brings it up in full resolution in a web browser:



From here you can right-click the image with the mouse, download it to your computer, copy and paste it into a message, post it to social media, and so forth.

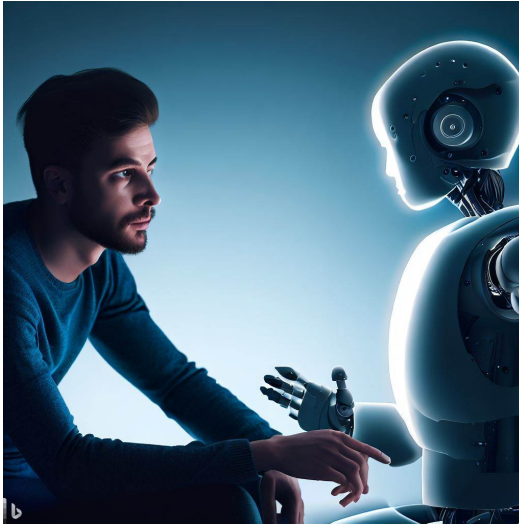
Using Copilot to create images is no more complicated than this - but, as with prompts to AI chatbots, there's a lot here to play with.

Although the Internet definitely loves cats, it also loves astrophotography. For our next prompt, let's try "create an image of a full moon above a green field".

Once again, Copilot/DALL•E create four images, all similar to this one:



You can apply some of what you've learned, to craft a prompt that features a situation with character, context and conflict, something like "create an image of a human and an AI chatbot having a deep conversation":



You can put pretty much anything you like into a prompt (so long the prompt doesn't request explicit or violent images). Copilot will instruct DALL•E to generate it. Begin your prompt with "create an image of", then follow with what you want to create - and you'll have your image within a few seconds.

Google Gemini, Anthropic Claude and Meta AI also allow you to create images – try a prompt that begins with 'create an image of...' and see!

Suggestions for exploration:

Can you recreate a famous scene from history or literature?

Can you generate an image of something that no one has ever seen?

Something microscopic? Scientific?

Something counterfactual or impossible?

Chapter 12 - Windows Copilot

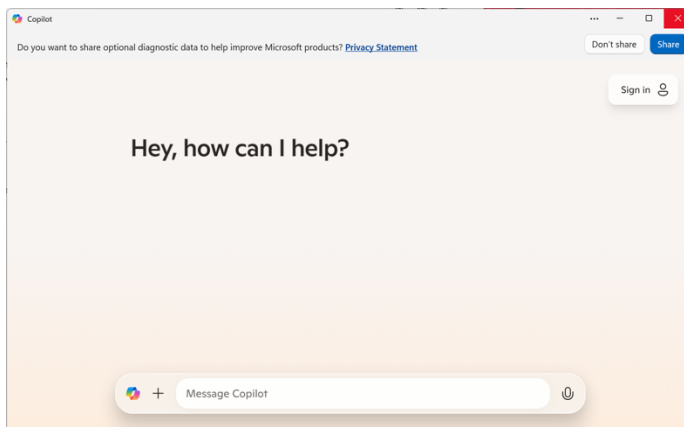
On the 23rd of May 2023, at its 'Develop' technical conference, Microsoft CEO Satya Nadella gave a presentation about the future of AI and Microsoft.

Nadella announced Microsoft would be going 'all in' on AI. Everywhere it made sense, in every Microsoft product or service, they would fully integrate the capabilities of Copilot - supported by their 49% ownership of OpenAI, generating completions using their massive Azure computing cloud.

Only a handful of firms have the kind of resources to make a blanket commitment to integrating AI across their product lines - Microsoft, Google and Meta, among them. Microsoft went first, pushed all its chips onto the table, daring the other technology giants to call their bet: Copilot in their Edge browser, plus Copilot integration across Office 365 - inside Word, Excel, Powerpoint -- even Outlook.

All of these product enhancements with AI chatbots were to be expected. But another one took developers by surprise - the announcement of "Windows Copilot", a deep integration of Copilot into the Windows 11 operating system. "We are bringing the Copilot to the biggest canvas of all: Windows," Nadella announced in his opening keynote "This is going to make every user a power user of Windows." And it can - if the users of those nearly half a billion Windows 11 desktops understand how to use Windows Copilot.

Windows Copilot appears up as an icon on the taskbar. Click on that icon, and a window opens up:



Windows Copilot looks almost exactly like Copilot on the Web - and underneath it runs the same bits of computer code. The key difference is that

Windows Copilot isn't safely contained "inside" a Web browser. It's fully a part of the computer's operating system - which means it can run programs and change system settings.

Which settings can Copilot change? Microsoft has been quite careful about letting Copilot touch important system settings. You can ask Copilot to change the display from 'Light mode' to 'Dark mode' - and vice versa. You can even ask it to take a screenshot. Submit that prompt, Copilot will think for a moment, then launch the Windows 'Snipping' tool to assist you in taking a screenshot of anything on the display.

Beyond that, exactly what Windows Copilot can do for you - above and beyond what can already be done with Copilot - is not completely clear. For safety's sake it's probably a good idea that Microsoft keeps Windows Copilot fairly restrained in what it can do. No one wants a rogue Copilot deleting files just because it has misinterpreted a prompt. Caution and slowness are called for. At the same time, tying an AI deep into the operating system means that some tasks that are well beyond the skills of most computer users could be automated, using Copilot. So there's enormous scope for Copilot to become very useful - but that scope is inevitably paired with greater danger: With great power comes great responsibility.

Every technique explored in this book can be used with Windows Copilot. Like Copilot, it can even be used to generate images - from your desktop. Like Copilot, Windows Copilot can be 'more precise', 'more balanced' or 'more creative'. And like Copilot, every once in a while Copilot will 'reset' the conversation back to square one. It's definitely a good idea to keep a notepad of your favourite prompts, characters and contexts, etc. at hand - so you can copy-and-paste them into Copilot whenever they're needed.

The more time you spend learning how to craft great prompts, the better you'll be able to use Window Copilot to assist you.

Play, learn, explore - and you'll be richly rewarded.

Chapter 13 - Autonomous Agents

A version of this chapter was originally published in CFO Magazine A/NZ

What could a 'perfect assistant' do for you? You'd be able to assign it any task, any responsibility, any role, and know that the assistant could perform admirably. A perfect assistant would understand the way you work - the way your organisation works - and work with it, improving your capacity to get things done. A perfect assistant would help shoulder the burden, making even the most difficult days in the office far easier.

Those perfect assistants are now on their way, under cover of another name, 'autonomous agents'. Although this may sound like a brand-new thing, the history of autonomous agents stretches back more than thirty-five years, to a video produced by Apple Computer, portraying a hypothetical autonomous agent - dubbed the 'Knowledge Navigator'.

In Apple's video, the *Knowledge Navigator* helps a busy university professor handle his correspondence - while simultaneously assisting him to prepare for a lecture being delivered later that same day. The *Knowledge Navigator* chats away in plain English, listens and responds appropriately to plain English, and does pretty much everything asked of it - even going so far as to leave an 'I've called' note on the professor's behalf when reaching out to a colleague who can not be reached.

The *Knowledge Navigator* inspired - but never worked. Even all of the billions spent on Apple's Siri, Google's Assistant and Amazon Alexa creating something similar never delivered anything close to the *Knowledge Navigator*.

Until ChatGPT. Powered by GPT-4o, it brings a nearly human-like sensibility and responsiveness to human interactions. ChatGPT delivers the richly human-like 'agent' the *Knowledge Navigator* first revealed to the world.

After OpenAI released the first version of GPT-4, it didn't take long for an enterprising programmer to work out how to pop it into a framework that turns it into an autonomous agent with many of the capacities of the *Knowledge Navigator*. It's called *Auto-GPT*, and it's open source software, so it's free to download, use - even modify.

Run the program and poses a question:

I want Auto-GPT to:

Whatever you type in response - however practical or fanciful - that's what AutoGPT will *try* to do for you.

How does it perform this bit of magic? One of the things we've learned - after a century of both operations research and management theory - is how to break

any activity into actionable parts. Auto-GPT will first translate whatever you've typed as a response into a goal. This it does by 'asking' ChatGPT to do the heavy lifting - via a detailed prompt. Auto-GPT itself acts as more of a conductor; it doesn't try to interpret English-language requests, instead forwarding them along to ChatGPT in prompts. When ChatGPT returns with a clearly defined goal as a completion, Auto-GPT again prompts ChatGPT, asking it to break that goal down into a series of steps. Once ChatGPT has done replied with its completion on how to do that, Auto-GPT then prompts ChatGPT to translate each of these steps into a sequence of actions.

It's this back-and-forth between AutoGPT-generated prompts and ChatGPT completions that reveals the potential of autonomous agents - and of AI chatbots more broadly. *Many* tasks can be approached in this way. In conversation, AutoGPT and ChatGPT work their way toward a solution.

Actions can be executed by Auto-GPT directly. Auto-GPT will prompt ChatGPT to write code to fulfil those actions, if they can be performed entirely within the computer. AutoGPT can also request its user perform such-and-such an action in the real world, with the results being fed back into Auto-GPT once that action has been completed.

Methodically, Auto-GPT performs each action in the correct order. Should the process fail at any point, AutoGPT prompts ChatGPT, asking it how it might perform that action differently. With a new completion from ChatGPT, AutoGPT then tries that new action, doing this again and again until it finds an action that succeeds. Auto-GPT does this until all the steps and actions have been performed - and the task completed.

That a computer can perform this sort of planning around goal execution is not surprising; numerous AI problem-solving systems have existed for decades. What makes Auto-GPT remarkable is that it is both very capable - it rarely fails to finish a task - and that it can be driven by human requests made in plain English. Pretty much anyone can ask Auto-GPT to work for them as a problem solver.

It's still very early days for these autonomous agents. They're already very powerful, yet also have one obvious shortcoming: they're native to the purely digital world of computers and the Internet. Getting them to do something in the real world means we'll need to connect them to all of the systems and sensors we've strung around our increasingly-intelligent planet.

Autonomous agents are already very useful for tasks such as gathering data about the operations of a business or business unit or sifting through a data 'lake' - highlighting the most interesting bits for review. They'll soon be put to work helping us plough through the mountains of paperwork generated by every organisation. Will autonomous agents be looking to take our jobs? No. But if we keep them busy, they may help us do better in ours.

Chapter 14 - Will an AI Chatbot take my job?

AI chatbots can answer questions (possibly even correctly!), solve problems, be taught new tricks, even work their way through tasks. Is there anything they can't do?

Plenty.

Behind that question lies another that's crossed the mind of almost anyone who's used an AI Chatbot - and for many who have only heard of them: Will an AI Chatbot take my job?

The answer is *complicated*. In the beautiful phrasing of Reverend Lovejoy from *The Simpsons*, it goes something like, "Yes, if... and No, but..."

Let's start with the "Yes, if..."

Have you tried to get an AI chatbot to do your job?

Most of our jobs consist of a wide range of tasks, many of which focus on paperwork, bureaucracy, time management, team management, accounting, and the like. An AI chatbot can make a decent attempt at performing these sorts of activities - especially after it's been well-taught by a Chain-of-thought prompt.

Break your job down by task - make a list of tasks, then try to do them using ChatGPT, Gemini and Claude. How many of those tasks can *already* be performed by an AI chatbot? How many of them *could* be performed by an AI chatbot, given some time to craft the correct prompts? And, finally, how many of your tasks lie so far outside the realm of anything we can do with an AI chatbot that there's no likelihood any AI chatbot (however capable) could cope with them?

Give that your best shot. Don't try to cheat the results, because (as we have been told innumerable times by our parents and teachers) cheating on this test only hurts you.

Assessing what percentage of your job consists of tasks that can be easily / with some effort / cannot easily be assigned to an AI Chatbot will help you understand where you stand today - and tomorrow. If the results show that a significant portion of your tasks can be reasonably well-handled by a well-instructed AI Chatbot, it's probably time to give serious thought to upskilling.

Which brings us to the "No, but..."

Certain tasks are *very* difficult to automate, and broadly they break down into two categories - tasks involving empathetic interaction, and tasks involving the fiddly bits of the physical world.

Although AI chatbots have been ‘faking’ empathetic interactions with humans since Joseph Weizenbaum’s ELIZA, they have no natural empathy. All of the empathy in the interaction is a projection of the person interacting with the machine. It’s all in their head. That doesn’t mean it’s unimportant, but it does mean that it’s unsupported, fundamentally lacking depth, understanding, and the ability to truly empathise - to put oneself into another’s shoes.

These are foundational human qualities. Our success as human beings depends on having a fully-developed capacity to empathise with others. That empathy allows us to offer one another meaningful support - emotionally and relationally. Without empathy, humans have little more to offer than machines. With that empathy, the scope for human relationships becomes infinitely rich. That richness extends to all of our relationships - with a partner, families, friends - and co-workers.

The best bosses are those who can fully empathise with employees who report to them. That’s nearly always hard work - but it’s the job. If the boss gets that right, everything else will sort itself out. If the boss fails at that, it doesn’t matter how good they are in the rest of their job.

Exploring the richness of empathy and human connection is the best way to ensure that you can never be replaced by any bit of a machinery. Organisations that automate away all of their empathetic humans become fit for machines but profoundly alienating to humans, who will struggle to connect and operate within those organisations.

Turning to the real world, one thing we’ve learned in nearly seventy years of research on artificial intelligence is that it’s surprisingly hard for software to make sense of the physical world. Even if an AI chatbot has read every textbook on physics, mechanical engineering, machinery and tools, none of that gives that chatbot any practical experience in a world that can be vague, slippery, in-between, and constantly evolving.

There’s nothing neat about the world, as can be seen by the repeated underestimation of the difficulty in creating autonomous automobiles - self-driving cars.

It’s not terribly difficult to teach an AI how to operate a motor vehicle. There are, after all, only so many controls that need to be mastered. But that’s far less than half of the task associated with driving. Driving is a continuous negotiation between the capabilities of the automobile and the information pouring in from the environment around the vehicle. First and foremost, other vehicles, which may have human drivers who are distracted or have poor views through their windcreens, or who may be battling conditions that make for poor visibility for both human and AI drivers. Then there are varying road surfaces, traffic conditions, pedestrians and pets, confusing road signage, etc. etc. Driving is an amazingly rich cognitive task that demands we integrate a huge amount of information in real-time. It’s only after we tried to teach computers to drive cars that we really started to look at exactly how hard the

problem really is - not just the theory of operating the vehicle, but the practice of driving in a rich and confusing world.

If you do something that is primarily physical - perhaps a trade like electrician or plumber or roofer or automobile mechanic - there's not much likelihood of any AI chatbot (or AI chatbot-controlled robot) taking your job any time soon. **The more embedded knowledge you require to do your job well, the harder it will be to automate.** Jobs that have lengthy apprenticeships - all of the trades noted above and many others besides - will resist automation for some time to come. As automation comes, these craftspeople will double down on their craft. In the best possible outcome, we could see a flowering of craftsmanship unlike anything since the High Middle Ages. The *Wall Street Journal* [recently reported](#) on the renaissance in high school 'shop' classes – precisely because those highly skilled tradecraft roles resist AI automation.

That said, we can't simply hide ourselves away in our machine shops and expect the revolution in automation - which AI chatbots enable - to pass any of us by. We need to catch that wave, get on top of it, and use it to propel ourselves forward. The way forward is the way through.

Chapter 15 - What the Future Holds / Next Steps

Until the middle of 2022, artificial intelligence seemed more like a joke than reality - the overboiled plotline to countless 'B' grade science fiction films. ChatGPT changed all of that. Those changes have only barely begun.

Microsoft founder Bill Gates once noted "We always overestimate the change that will occur in the next two years - and underestimate the change that will occur in the next ten." That's already proving true for AI chatbots. People start out believing anything is possible – then learn the limits of these new tools. They're not magic wands, but they can be very useful.

As we're taking their measure, we'll put these new tools to work. That won't happen overnight. We are still in the early days of learning how to use these tools well. Beginnings are usually clumsy and almost always embarrassing. (Does anyone remember "Clippy", the oh-so-helpful but mostly useless 'agent' that Microsoft touted twenty years ago?)

Within half a decade, something that proves to be more of an annoyance than useful in 2025 will become so well-integrated into our workflows that we no longer notice the AI chatbot assisting us in our productivity. The best tools disappear into our practice; that's when we realise their true utility.

On the path there, we'll see further developments of the large language model technology underlying these AI chatbots. Research from Meta Networks (the former Facebook) demonstrated how these models can be 'slimmed down' to run comfortably on modern smartphones. While not quite as 'smart' as ChatGPT, they're still 'good enough' to be helpful across many tasks, and can be easily 'fine tuned' - further trained - to deliver specific capabilities. Both Apple and Google have integrated AI chatbots into their mobile platforms.

It's likely that rather than having a one-size-fits-all AI chatbot like Claude or ChatGPT, we'll have access to a range of highly customised chatbots, each suited to a particular task. Just as a craftsman selects the right tool for the task at hand, so we'll select the right chatbot. This means we will all need broad experience across a wide range of AI chatbots. You need to use a tool for a while to understand where and how it can best be used.

The path forward won't be driven by tech giants like Microsoft, Google, Apple and Amazon, but by the billions of users of these new AI Chatbots - individuals and organisations. Most organisations will never feel comfortable sharing their confidential data with a commercial AI chatbot, and will host their own, highly secured AI chatbot, available for staff to access via the organisation's intranet. Those chatbots will feast on customer data, organisational expertise, and staff interactions, growing more precise, more accurate, and more useful the more they get used, becoming a core asset to these organisations. Organisations that successfully integrate AI chatbots will be able to work faster and smarter than their competitors.

Just as organisations everywhere have come to rely on group communication tools like email, chat, Slack and Zoom, organisation-based AI chatbots will act to support the staff in everything they do. People (and organisations) will work better with them.

That's an alluring future, but it's only possible if people begin exploring the potential of AI chatbots. Waiting for 'the answer' to be delivered by a technology vendor means that you'll be saddled with whatever the vendor sees as its priorities - and those priorities may not be close to yours, or your organisation.

It's on each of us to give this a red-hot go: **spend time playing, experimenting - and failing. Learn from your successes, and learn more from your failures.** All of that experience will guide you into using this powerful technology to the best of your abilities, and to the best possible ends.

At the end of 2023, Microsoft's Windows Copilot popped up on about a half a billion Windows 11 desktops. People throughout the world – both at their desks and in their homes – have easy access to a technology that can serve them, or confound them. **Avoidance is not an option.** Wisdom lies in knowing how to harness the benefits of AI chatbots, while avoiding the pitfalls. Hopefully this book has helped impart some of that wisdom.

*Mark Pesce
Sydney & Campsie
August 2023 (Revised March 2025)*

We hope you have enjoyed this eBook edition of *Getting Started with AI Chatbots Revised Edition*.

In January 2024, BCS Publishing released their definitive edition, *Getting Started with ChatGPT and GenAI Chatbots: An introduction to generative AI Tools* – packed with more information to help you to explore this latest frontier of computing.

To order, visit <https://bcs.org/books/genAI>
Or just scan this QR code!



Mark Pesce co-invented the technology for 3D on the Web - laying the foundations for the metaverse - has written nine books, including *Augmented*



Reality: Unboxing Tech's Next Big Thing, was for seven years a judge on the ABC's *The New Inventors*, founded postgraduate programs at the University of Southern California and the Australian Film Television and Radio School, holds an honorary appointment at Sydney University, is a multiple-award-winning columnist for *The Register*, writes features for *COSMOS* Magazine, and is a professional futurist and public speaker. Pesce hosts both the award-winning '[The Next Billion Seconds](#)', and, with VRML

co-inventor Tony Parisi, released the highly-praised, award-nominated series "A Brief History of the Metaverse". More at www.markpesce.com