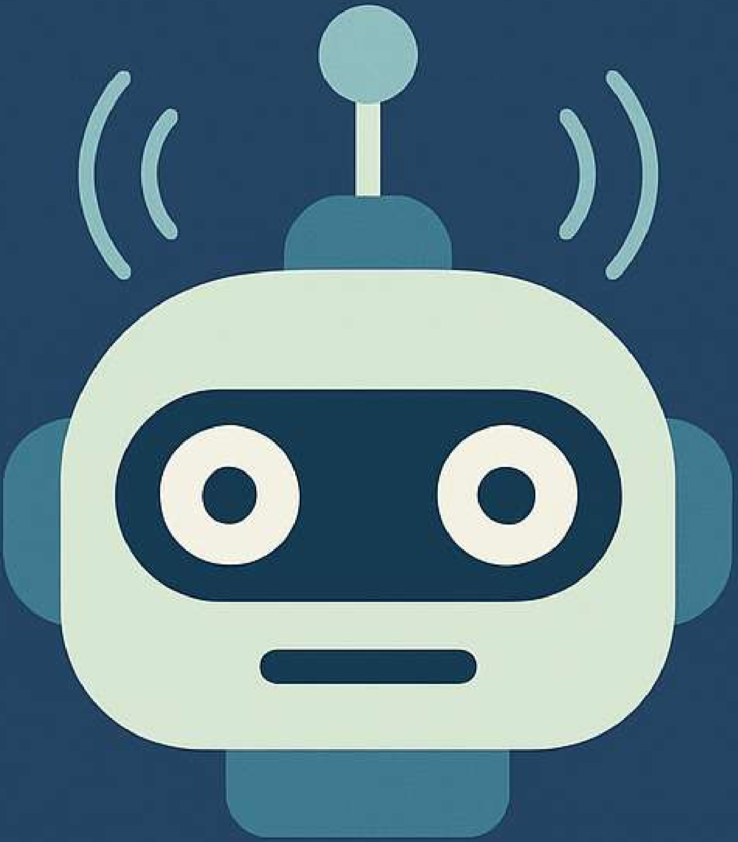
UNDERSTANDING CHATGPT & MODERN AI

The Good, The Bad, and The Questions That Remain



01 May 2025

HAMMAN SCHOONWINKEL

With some assistance from a very helpful robot

Introduction

Artificial Intelligence (AI) is no longer just science fiction – it's part of our daily lives. From asking Siri or Alexa about the weather to having Netflix suggest what to watch, AI helps computers perform tasks that normally require human intelligence. This chapter will introduce AI in a friendly, accessible way. We'll start with a simple explanation of what AI is, take a quick tour through the history of AI, and then focus on one of the most exciting recent AI developments: ChatGPT. We'll learn when ChatGPT was launched, what it can do, how it grew out of earlier AI models, and how it sparked rapid advancements in AI across many fields. We'll peek under the hood to see how ChatGPT works as a large language model (LLM) – explaining concepts like training data and how it predicts text, as well as its limitations (like why it sometimes "hallucinates" incorrect information). We'll compare ChatGPT to traditional search engines like Google to understand the differences. After that, we'll explore the explosion of other AI applications beyond ChatGPT – from image generators and voice tools to productivity apps that can speed up our work. We'll also discuss important ethical issues (such as misinformation, bias, plagiarism and privacy) and how AI might affect jobs in the future. Finally, we'll look ahead to the future of AI, including ideas like artificial general intelligence, brain-machine interfaces, and how humans and AI might work together in the years to come. The tone throughout will be friendly and conversational, making complex ideas as easy to understand as possible. Let's dive in!

---

1. What is Artificial Intelligence?

At its core, Artificial Intelligence (AI) is about making computers do things that usually require human intelligence. For example, when you talk to a virtual assistant like Apple's Siri or Amazon's Alexa, you're interacting with AI – the assistant recognizes your speech, interprets your question, and responds with an answer. Another everyday example is a movie recommendation on Netflix or a song suggestion on Spotify; those services use AI algorithms to learn your preferences and suggest things you might enjoy.

To formalize it a bit, one definition of AI is the ability of computers and systems to perform tasks that typically require human cognitive skills. These tasks involve things like understanding language, recognizing patterns, learning from data, and making decisions. AI systems often *mimic* aspects of human cognition – for instance, an AI might analyse thousands of pictures of cats and dogs, learn their features, and then be able to tell apart a cat or dog in a new picture (much like a human can after some practice).

A key concept in modern AI is machine learning. Rather than programmers hand-coding every rule (which is practically impossible for complex tasks), machine learning lets the AI *learn* from examples. For instance, to teach an AI to recognize spam emails, we don't write thousands of specific rules for every spam message; instead, we feed the algorithm lots of example emails labeled "spam" or "not spam" and let it figure out the patterns.

In summary, AI is about computers doing *smart* things. It encompasses a range of techniques and systems, from simple decision trees to complex deep neural networks. You likely interact with AI more often than you realize - whether it's through recommendation systems, voice assistants, or even the autocorrect on your phone. Now that we know what AI is in general, let's take a quick journey through the history of AI to see how we got to where we are today.

---

2. A Brief History of AI

AI might seem like a buzzword of the 21st century, but the dream of intelligent machines goes back many decades. Understanding the history of AI will give context to how we eventually arrived at technologies like ChatGPT. Below is a brief timeline highlighting some key milestones in AI development over the years:

1950 – The Turing Test:

British mathematician Alan Turing published a paper titled “Computing Machinery and Intelligence,” where he introduced the idea of the Turing Test. The Turing Test was a thought experiment to evaluate if a machine could exhibit intelligent behaviour indistinguishable from a human. If a human judge conversing with a machine (via text) couldn’t tell whether it was a machine or a person, the machine could be considered “intelligent.”

1956 – The Birth of "Artificial Intelligence":

The term “Artificial Intelligence” was coined at a workshop at Dartmouth College, proposed by researchers including John McCarthy and Marvin Minsky. This is often considered the founding moment of AI as a formal field of study.

1950s–1960s – Early AI Programs:

AI pioneers developed early programs like Arthur Samuel’s self-learning checkers game and Joseph Weizenbaum’s ELIZA — an early chatbot that mimicked a psychotherapist. Robots like Shakey could navigate rooms and perform simple tasks.

1970s–1980s – AI Winters and Expert Systems:

After initial enthusiasm, progress stalled. Disappointment with limited results caused funding cuts — known as "AI winters." Still, some success came from expert systems — rule-based programs used in fields like medicine and troubleshooting.

1997 – Deep Blue Beats World Chess Champion:

IBM’s Deep Blue defeated Garry Kasparov, a milestone for AI in strategic games. Though based on brute-force calculations, it proved AI could outperform humans in complex tasks.

Late 1990s–2000s – Rise of Machine Learning:

As computing power grew, AI research shifted to machine learning — allowing systems to learn from large datasets. Web data exploded, enabling smarter algorithms and real-world applications like speech recognition and the Roomba robot vacuum.

2011 – Watson and Siri:

IBM’s Watson beat human champions on Jeopardy! by parsing complex natural language. Apple launched Siri, bringing AI voice assistants to the mainstream.

2014–2016 – Game-Playing AI and Creative Tools:

Google DeepMind’s AlphaGo defeated one of the world’s best Go players — a feat many believed was decades away. Around this time, Generative Adversarial Networks (GANs) and self-driving car prototypes also gained traction.

2017 – Transformers and Modern AI:

Google introduced the Transformer architecture in a paper titled “Attention Is All You Need.” This unlocked a new wave of progress in natural language processing — enabling the GPT family of models that power tools like ChatGPT.

2020 – The Big Language Model Breakthrough:

OpenAI released GPT-3, a massive language model with 175 billion parameters. Its ability to generate surprisingly human-like text on a wide range of topics marked a turning point in AI capabilities. This showed that scaling up AI models could unlock general-purpose language understanding.

2021 – Creativity and Multimodal AI:

OpenAI unveiled DALL·E, an image generator that turned text prompts into pictures — showing that AI could connect language and vision in creative ways. AlphaFold (by DeepMind) solved a 50-year-old challenge in protein structure prediction, showcasing AI's scientific potential.

2022 – The Arrival of ChatGPT:

On November 30, 2022, OpenAI released ChatGPT, a conversational AI based on the GPT-3.5 model. It went viral almost instantly — reaching millions of users in days. ChatGPT brought generative AI to the public in a usable, exciting way.

2023 – The Generative AI Boom:

OpenAI launched GPT-4, enabling multimodal input (like images). Microsoft integrated this tech into Bing and Office. Google released Bard. AI image, text, music, and code tools flooded the market. It was the year AI went fully mainstream.

2024 – AI in Your Pocket:

Major tech companies began consolidating their AI tools into branded ecosystems, making it easier for users to recognize and interact with AI across different platforms. Google rebranded its chatbot Bard as Gemini, aligning it with their powerful new AI model family of the same name.

Microsoft brought all of its AI features under the Copilot brand, integrating AI into its Office suite, Windows, and even its Bing search engine. Apple introduced Apple Intelligence, focusing on private, on-device AI deeply embedded in iPhones, iPads, and Macs. Meanwhile, Samsung launched Galaxy AI, which offered features like real-time call translation and AI-powered photo editing. This marked a shift from AI being just something you accessed through a website or app — it was now becoming an invisible assistant baked into the devices and operating systems people use every day. This historical journey sets the stage for the rest of the chapter, where we'll explore ChatGPT more deeply — how it works, what it can do, and what it means for the future.

This timeline is just a high-level overview. AI's history is rich with ideas and technologies, from the early rule-based systems to the learning machines of today. The key takeaway is that progress was not linear – there were ups and downs – but overall, the field has advanced tremendously, especially in the last decade, thanks to machine learning and big data. Now, let's zero in on ChatGPT, since it's such a pivotal development in AI and a focus of this chapter.

---

3. ChatGPT and the Generative AI Revolution

By now, you've probably heard of ChatGPT – the AI chatbot that took the world by storm. In this section, we'll explore what ChatGPT is, when it launched, what it can do, how it evolved from earlier AI models, and how its arrival sparked rapid advancements in other areas of AI.

The Launch of ChatGPT

ChatGPT was introduced to the public on November 30, 2022 by the AI research company OpenAI. OpenAI made ChatGPT available as an online chat interface that anyone could try out for free (initially as a research preview). Within just a few days, it became a viral sensation. In fact, ChatGPT surpassed 1 million users in only five days after launch. To put that in perspective, that growth was faster than practically any previous internet service – it took Facebook and Instagram many months to reach a million users, but ChatGPT did it in under a week. By January 2023 (two months later), it was estimated to have around 100 million monthly active users, which likely made it the fastest-growing consumer app in history at the time. This sudden popularity was a clear sign that ChatGPT had tapped into something special – ordinary people found it genuinely useful and often mind-blowing.

Why all the excitement? Unlike many previous AI demos, ChatGPT felt like chatting with a knowledgeable friend who could help with almost anything. OpenAI had taken its powerful GPT-3.5 language model and fine-tuned it to be conversational, polite, and helpful. The result was an AI you could *talk* to, ask questions, get explanations, or even have it create content for you. And it worked astonishingly well for a broad range of queries, from writing a poem about winter to explaining quantum physics in simple terms.

The launch of ChatGPT is often seen as a watershed moment for AI because it brought AI into the mainstream consciousness in a new way. People who never used programming or machine learning directly were suddenly playing with a cutting-edge AI by just messaging it, as if texting a very smart friend. Social media buzzed with examples of what ChatGPT could do, and it often felt magical.

What Can ChatGPT Do?

ChatGPT is a kind of *Jack-of-all-trades* when it comes to language. Since it was trained on a broad swath of the internet (articles, books, websites, etc.), it picked up knowledge on an enormous range of topics. Here are some of the things ChatGPT can do:

Answer Questions and Explain Concepts: You can ask ChatGPT factual questions (e.g., “What is the capital of France?” or “How does photosynthesis work?”) and it will give you an answer. It often provides explanations in a clear, step-by-step manner. For instance, students have used it to help understand concepts in math, science, or history – basically acting like a tutor that can explain things in different ways if you ask.

Write and Create Content: ChatGPT can produce human-like text on demand. It can write essays, articles, or reports on a given topic. It can also draft emails or letters, which is handy if you're not sure how to phrase something. It can compose stories or poems, often in a requested style (like "write a poem about summer in the style of Shakespeare"). It can even attempt jokes or puns. The writing isn't always perfect, but it's impressively coherent most of the time.

Help with Coding and Problem Solving: One surprising application is that ChatGPT can assist with computer programming. You can ask it to write a snippet of code (e.g., "write a Python function to sort a list of dictionaries by a field") and it will try to generate it. You can also paste code and ask for help finding a bug or explaining what the code does. It became like a coding assistant for many developers.

Generate Ideas and Brainstorm: Need ideas for a birthday party theme? Or topics for an essay? ChatGPT can brainstorm with you. It can generate lists of ideas, pros and cons for a decision, or outlines for projects. It's like having an ever-ready brainstorming partner.

Translate and Summarize: ChatGPT can translate text from one language to another (though it's not officially a translation tool, it often does a decent job). It can also summarize long texts. For example, you could give it a long article and say "Summarize this in a few bullet points" and it will attempt to pull out the key points.

Act as a Conversational Partner: Some people use ChatGPT just to chat or role-play. It can take on various personas or engage in creative, open-ended conversations. Want to practice a conversation in French? Or pretend to interview Nelson Mandela or Marie Curie? ChatGPT can play along (within certain limits of appropriateness set by OpenAI). It often feels surprisingly human-like in dialogue, remembering what you said earlier and maintaining context over multiple turns.

For example, you could ask:

“Who would make your all-time Proteas Test XI?”

ChatGPT might respond with names like Graeme Smith, Jacques Kallis, AB de Villiers, Dale Steyn, and Allan Donald. But you could push back:

“What about Shaun Pollock — doesn’t his all-rounder record earn him a spot?”

“Where’s Hashim Amla?”

“Would you go for Jonty Rhodes in the middle order or stick with Kallis and Duminy?”

“How would this team perform against Australia’s all-time XI?”

Before you know it, you’re deep in a nuanced sports debate — with an AI. And while you’re doing that, you're also practicing reasoning, argumentation, and even historical research — because ChatGPT will often cite stats or contexts to back up its picks (though always double-check them!).

This kind of roleplay doesn’t just make learning more fun — it turns abstract concepts into real-time engagement. Whether you're arguing about cricket, simulating a job interview, or preparing for a presentation, ChatGPT can be a surprisingly effective conversational partner.

Creative and Miscellaneous Tasks: ChatGPT’s ability to follow instructions in plain language means it can do a lot of random things. It can help write a song lyric, fix grammar in a sentence, act as a study quiz generator, simulate a text-based game, provide motivational quotes, and so on. Its versatility is a key strength – essentially, if the task involves *text*, ChatGPT can probably attempt it.

It quickly became a go-to tool for a wide range of users — from students and teachers, to writers, coders, professionals, and even the simply curious. Anyone who could type a question into a chat box could suddenly access an incredibly powerful language model — no technical skills required. However, as we'll discuss later, ChatGPT is not perfect. It has limitations and can produce incorrect or nonsensical answers at times. But before we get to that, let's explore where ChatGPT came from – it didn't just pop out of nowhere; it's the result of years of research and previous versions.

From GPT-1 to ChatGPT: Evolution of the Model

The beginnings: ChatGPT is built on the foundation of the GPT series of models developed by OpenAI. "GPT" stands for Generative Pre-trained Transformer. We already touched on the significance of the Transformer architecture (invented in 2017) in the history section. Now, let's see how it led to ChatGPT:

The journey to ChatGPT began with GPT-1 in 2018 — a small but groundbreaking model that proved transformers could be used to pre-train a language model by predicting the next word in a sentence. GPT-2 followed in 2019 with 1.5 billion parameters, generating coherent paragraphs and sparking debate due to concerns about potential misuse. GPT-3, released in 2020, was a major leap, with 175 billion parameters and the ability to generate impressively human-like text, even for tasks it wasn't explicitly trained on. However, it didn't always follow instructions well. To improve this, OpenAI created InstructGPT by fine-tuning GPT-3 using human feedback (a method called Reinforcement Learning from Human Feedback, or RLHF). This led to GPT-3.5 — the model that formed the foundation for the first public version of ChatGPT. GPT-3.5 was more aligned with user expectations, more conversational, and better at staying on topic, setting the stage for the AI chatbot millions would soon meet.

ChatGPT (Nov 2022): Using the GPT-3.5 family model and making a conversational interface was the stroke of genius that became ChatGPT. The chat format allowed back-and-forth interaction, so users could ask clarifying questions or request rewrites, and the model would take previous context into account. This made it feel much more interactive and useful than single-turn question-answering. ChatGPT's training also incorporated examples of dialogue and what to do or not do in a conversation (for instance, it was trained to not use profanity, not disclose certain internal details, and so on).

GPT-4 (Mar 2023): While the chapter is mainly about ChatGPT, it’s worth noting that OpenAI didn’t stop at GPT-3.5. They developed GPT-4 which can handle not just text but also images as input (you can show it a picture and ask questions about it, for example).

GPT-4o (May 2024): OpenAI introduced GPT-4o, where the "o" stands for "omni," highlighting its multimodal capabilities. GPT-4o is optimized for speed and versatility, making it suitable for tasks requiring quick responses across various formats.

o1 (September 2024): Designed to handle complex reasoning tasks. Unlike its predecessors, o1 takes more time to process inputs, allowing for deeper analysis and more accurate responses. This makes it particularly effective in fields like scientific research, coding, and intricate problem-solving.

Deep Research (February 2025): Deep Research is an AI agent integrated into ChatGPT, designed to autonomously conduct extensive web research. It plans and executes multi-step searches to gather, analyse, and synthesize information, producing comprehensive reports within 5 to 30 minutes. Users can input queries along with supplementary materials like images, PDFs, or spreadsheets to provide context. Deep Research is particularly beneficial for in-depth knowledge work in areas such as finance, science, policy, and engineering.

Operator (January 2025): OpenAI introduced Operator, an AI agent designed to autonomously perform a variety of web-based tasks on behalf of users. Unlike traditional AI models that primarily generate text or analyse data, Operator interacts directly with web interfaces, simulating human actions such as clicking, typing, and scrolling. This capability enables it to handle tasks like filling out forms, ordering groceries, and booking travel arrangements. While Operator represents a significant advancement in AI-driven task automation, it is still in its early stages and may require user supervision, especially for tasks involving sensitive information or complex interactions.

ChatGPT Subscription Models (as of April 2025)

OpenAI offers three subscription levels for using ChatGPT, each offering increasing access to advanced AI tools:

Free (R0/month)

- Access to a basic version of GPT-4o (called GPT-4o mini)
- Limited features (e.g., file uploads, voice, and image tools are restricted)
- Suitable for light or casual use

Plus (±R375/month)

- Full access to GPT-4o and several advanced reasoning models
- Can use tools like image generation, data analysis, voice, and video more freely
- Includes early access to new features (e.g., GPT-4.5 and Deep Research in limited form)

Pro (±R3,750/month)

- Everything in Plus, with unlimited access to the most powerful models (like o1 and GPT-4o full)
- Extended use of voice, video generation, and Deep Research
- Includes access to Operator, an AI agent that can perform web tasks like booking or searching on your behalf

ChatGPT's Impact and the AI "Race"

The launch of ChatGPT in late 2022 wasn't just a big moment for OpenAI — it was a shockwave across the tech world. Within weeks, it became clear that conversational AI had not only matured, but was ready for mainstream use. What followed is now often called the AI arms race — a scramble among major tech companies to respond, compete, and stake their claim in what many saw as the next big platform shift after smartphones and cloud computing.

ChatGPT's viral success forced almost every major tech company to react. It wasn't just about hype — people were using the tool to solve real problems, from writing essays to analysing code and even brainstorming business ideas. The potential was obvious, and so was the competitive threat.

Microsoft, already a major investor in OpenAI, quickly embedded ChatGPT's technology into its products. Bing got a conversational search mode, and Microsoft Office evolved into Microsoft 365 Copilot, letting users generate text, summarize data, and automate tasks in Word, Excel, and beyond — all through natural language.

Google, long a leader in AI research, was caught off-guard by how fast ChatGPT took off. Despite inventing the Transformer architecture that powers it, Google had been more cautious about launching public-facing AI tools. But that changed fast — it launched Bard in early 2023 (later rebranded as Gemini), and by 2024, AI was baked into Gmail, Docs, Search, and more.

Meta released its open-source LLaMA models (Large Language Model Meta AI), which now power AI assistants integrated into apps like WhatsApp, Instagram, and Messenger — meaning students may have already chatted with LLaMA without even realizing it.

AI Everywhere

What makes this moment feel different from past tech trends is how quickly generative AI became integrated into everyday life. It's not a single app — it's becoming part of almost every app. From Adobe Photoshop offering AI image editing, to Zoom summarizing meetings, to Khan Academy deploying AI tutors, the impact of ChatGPT has cascaded across the tech ecosystem.

Some analysts have compared this moment to the launch of the iPhone or the early days of the internet. ChatGPT showed the world what was possible, and in doing so, reset the expectations for how we interact with technology.

The result wasn’t just one popular app — it was a paradigm shift. AI is now seen not as a niche tool, but as the foundation for future digital experiences.

4. How ChatGPT Works

ChatGPT might feel like talking to a human at times, but underneath, it’s a complex machine learning system. In this section, we’ll explain in simple terms how ChatGPT (and large language models like it) actually work. We’ll cover its training process, how it generates responses (the idea of predicting tokens), and its limitations such as why it can sometimes give wrong or made-up answers. We’ll also contrast ChatGPT with traditional search engines to highlight how it’s different.

Training on Massive Data

The first thing to know is that ChatGPT learned to write by *reading* a whole lot of text. During its training phase, it was fed an enormous amount of text data from the internet – including books, Wikipedia articles, websites, news, forums, and much more. Essentially, it digested much of the written content available on the web (up to around 2021 for the model versions we’re discussing). This is why ChatGPT has knowledge on all sorts of topics: history, science, literature, pop culture, coding, etc. It’s because somewhere in its training data, those topics were discussed.

The training process is unsupervised in the sense that the AI wasn’t explicitly told “this is the correct answer” for each input. Instead, it learned by doing a simple but profound task: predicting the next word in a sentence. For example, if the training data had the sentence “The capital of France is Paris,” the training process would feed the model “The capital of France is ___” and ask it to predict the next word. If it predicts “Paris,” it gets reinforced; if it predicted something else, the model adjusts its internal parameters to be a little more like the correct prediction. By doing this billions of times, the model gradually became very good at predicting likely sequences of words.

Because it had to predict next words, the model learned grammar, facts, reasoning patterns, style, and even some level of common-sense knowledge (just by statistical associations in text). For instance, it saw many examples of how sentences are formed, how answers to questions usually look, what factual statements tend to be, and so on. By the end of training, you get a neural network with billions of parameters that encodes a vast amount of information about language and the world, all in a very distributed way.

One way to think of it is this: ChatGPT doesn't have a database of facts it's pulling from explicitly. Instead, it has *implicit knowledge* compressed in its network weights from all that reading. It's learned patterns like "usually after the words 'the capital of [country] is', the next word is the capital city of that country" because it saw so many examples. Similarly, it learned structure like how a well-organized essay flows (intro, body, conclusion) or how code is typically written, etc., from examples.

However, one big caveat is that its knowledge has a cutoff. Since it was trained on data up to a certain point (for example, up to 2021), it doesn't know about events or information that came after that. So if you ask it who won the World Cup in 2022, an un-updated model wouldn't know (or it might guess and possibly guess wrong). Newer versions or updated versions of these models can be trained on more recent data, but as of the initial ChatGPT release, knowledge was not fresh or updated in real-time.

Additionally, the training data might contain errors or biased perspectives, and the model can absorb those too – more on this when we discuss limitations.

The Magic of Token Prediction

When you interact with ChatGPT, what's happening behind the scenes as it generates each response? It might be surprising, but it's not pulling sentences from a library or doing a search – it's literally generating text on the fly, one piece at a time. Specifically, it generates text *token by token*. A token is usually a word or a sub-word or character; language models break text into tokens for processing. For example, the sentence "AI is fascinating." might be broken into the tokens ["AI", " is", " fascinating", "."]. For a longer word, sometimes it might break into parts if it's rare.

ChatGPT looks at your prompt and the conversation history (which is also tokens to it) and then starts predicting a next token that would make sense as a continuation. It doesn't plan out the whole answer in advance. It chooses a next word (token) based on probabilities – essentially, "Given everything so far, what is the most likely next word or a highly plausible next word?" Once it picks that word, that word becomes part of the input for predicting the following word, and so on. It repeats this process really fast, which is why it can generate long answers that seem coherent.

Crucially, the model uses something called attention (from the transformer architecture) to look at all the context of the conversation to decide what comes next. It has no consciousness or intention; it's doing a complex statistical prediction. But because the model is so large and was trained on so much, this statistical prediction can produce very impressive results – paragraphs that have a logical flow, sentences that answer the question, etc. It's a bit like autocompletion on steroids. If you've seen your phone suggest the next word while texting, imagine that but with the knowledge of the entire internet and the ability to craft whole essays, not just finish a word or two. ChatGPT is essentially a very advanced autocomplete that "writes" the most likely continuation of the dialogue that an intelligent person might write.

However, to avoid always giving the most boring or obvious response, the model can add some randomness or follow certain decoding strategies that balance between the highest probability and some creativity. That's why you don't get the exact same word-for-word answer every time for a given question (unless the question is very straightforward). There's some stochastic (random) element in how it picks tokens, which is why it can even come up with slightly different phrasings or ideas on different tries.

One thing to highlight: ChatGPT does not think or understand in the way humans do. It doesn't have beliefs or awareness. It doesn't have a mental model of reality or a consistent “self”. It’s simply using patterns. For example, if in training texts most people said “I’m sorry, I don’t have information on that” when they didn’t know an answer, the model might output a similar phrase when it’s stumped – not because it *feels* sorry, but because that’s the learned appropriate response pattern. This is an important point when we consider its limitations.

Fine-Tuning and Safety Measures

As mentioned earlier, ChatGPT underwent fine-tuning where human AI trainers gave it demonstrations of good conversations and ranked its responses, etc. So on top of the raw language prediction training, it got some additional training to align with user expectations and ethical guidelines. This is why ChatGPT might refuse to answer certain questions. For example, if you ask it for something inappropriate or harmful (like instructions to do something dangerous or a hateful remark), it often replies with a refusal: “I’m sorry, but I cannot assist with that request.” That’s not because it *morally* decided that on its own; it’s because during training, humans told it “if someone asks for something like this, that should be the response.”

Despite these safety measures, sometimes the model can still produce disallowed content or get things wrong. OpenAI and others continuously refine these aspects.

ChatGPT vs. Search Engines (e.g., Google)

Now, let’s differentiate ChatGPT from a traditional search engine like Google, because this is a common point of confusion. Both are used to get information, but they work in fundamentally different ways:

Search Engine (Google/Bing without AI): When you type a query in a search engine, the engine goes and looks through its index of the web for pages that are relevant to your query. It then shows you a list of results – usually webpage titles and snippets – and it’s up to you to click and find the information you need. Essentially, a search engine is a *lookup tool*. It finds what other humans have written (websites, documents) and directs you there. The content you get is authored by humans (or sometimes by AI these days if the site uses it, but originally it’s human content). A search engine does not generate new content on the fly; it retrieves existing content. Also, search engines are up-to-date (Google is constantly indexing new pages, so it has very current information).

ChatGPT (LLM-based QA): ChatGPT, on the other hand, generates an answer on the fly by synthesizing what it “knows.” It does not give you quotes or sources by default (unless you specifically ask or the interface is augmented to do that). You ask a question, and it directly gives you an answer in natural language. It doesn’t tell you where that answer came from (in the basic interface). So you’re kind of taking the model’s word for it. The advantage is convenience – you get a single, coherent answer that’s often exactly what you need, without having to read through multiple web pages. The disadvantage is trustworthiness – you have to trust that the AI’s answer is correct, and as we discussed, it might not always be. Another big difference: ChatGPT’s knowledge has a cutoff and it doesn’t automatically know things updated after that (unless connected to a live search tool), whereas Google will have the latest information and news.

Context and Conversation: ChatGPT remembers the context of your conversation. You can ask follow-up questions like “Okay, now explain that in simpler terms” or “How does that compare to X?” and it will understand you are referring to the previous discussion. Google doesn’t remember your last search in the standard workflow (though it may personalize results to you generally, it doesn’t have a memory of a conversation where each query builds on the last). This conversational ability is a huge plus for ChatGPT – it’s more like an ongoing dialogue, which is often how humans naturally seek information (“actually, tell me more about that part…”).

Creativity and Open-Ended Tasks: If you ask Google for say, “Write a short story about a dragon who loves painting,” Google will just show you results where that sentence appears or similar. ChatGPT will actually *write you an original short story* on the spot. For any task that requires generation of new text (a story, an essay, a piece of code, a poem, a summary), a search engine alone cannot do that – it can only retrieve what’s already out there.

ChatGPT can create new content tailored to your request. That’s a fundamental difference: retrieval vs. generation.

Accuracy and Sources: If you need a guaranteed factual answer with a credible source, search engines are often safer. For example, if you need the current population of a country, Google will show a snippet with that info and maybe a source like the World Bank. ChatGPT might give you a number too, and often it’s right for well-known facts, but if it’s wrong and you don’t double-check, you wouldn’t know from the answer itself. Traditional search encourages you to evaluate sources (you might trust a .edu or .gov site more, etc.). ChatGPT doesn’t show sources unless you prompt it to, which changes the dynamic. Some newer AI systems (like Bing’s AI chat mode) actually *cite* sources for their answers, merging the two approaches – that is ideal, since you get a coherent answer plus the citations to verify.

In short, ChatGPT is like talking to a knowledgeable, verbose expert, whereas Google is like searching a massive library for relevant documents. They have different use cases. Sometimes ChatGPT can replace a search – for instance, if you want a quick explanation or a summary. But other times, especially for very recent information or authoritative sourcing, search is superior.

It’s interesting that with the advent of ChatGPT, there were headlines like “Will ChatGPT kill Google?” because people found themselves using Google less for certain questions. As noted, some even wrote “obituaries” for traditional search. The reality is that these technologies are now merging: search engines are adopting AI to give better answers, and AI models are being augmented with retrieval tools to use the live web. As a student, you’ll likely use both: maybe you’ll ask ChatGPT to explain something, then use Google to double-check a fact or find a specific reference for your assignment.

One more difference to be aware of: the cost and efficiency. ChatGPT’s processing (especially the large models) is computationally expensive. Every query costs a lot more computing power than a Google search, which is why it’s not trivial to just replace all search with AI answers – it’s expensive to scale. Also, ChatGPT might have limits (like number of messages in an hour) on the free version, etc., because of that cost. Search is highly optimized to be fast and cheap per query. Over time, these differences might blur as tech improves.

In summary, ChatGPT works by generating text based on patterns learned from a vast dataset. It excels at producing human-like responses and handling a conversation, but it can also hallucinate incorrect information and reflect biases from its training data (more on this later). It differs from search engines in that it creates answers from scratch and doesn't directly show you external sources. Both have their place, and understanding how ChatGPT works helps us use it wisely – leveraging its strengths (convenience and creativity) while being mindful of its weaknesses (accuracy and bias).

Now that we've explored ChatGPT deeply, let's broaden our view again and look at what else is happening in the AI world beyond text-based chatbots. ChatGPT was kind of the "big bang" for AI awareness, but it's part of a larger trend of AI breakthroughs in various domains.

5. Beyond ChatGPT

While ChatGPT is an AI focused on text and language, the surge in AI development has extended into many other areas. We are now seeing AI tools that can generate images, create videos, mimic human voices, assist with office work, and much more. In this section, we'll survey some of these AI applications beyond just chatbots. This will give you a sense of how AI is becoming a multi-faceted tool, impacting creativity, communication, and productivity.

AI Image Generators: Art at Your Fingertips

One of the most visually impressive developments in AI is the rise of text-to-image generators. These models allow you to type a description — something as simple or imaginative as "a castle floating in the clouds, in watercolour style" — and the AI will generate a brand-new image that matches your words. It’s like collaborating with a robot artist who never gets tired or runs out of paint.

OpenAI’s DALL·E model (with its improved versions like DALL·E 2 and now DALL·E 3) helped popularize this concept. Other major players include Midjourney and Stable Diffusion. These models learned from millions of image-caption pairs during training. So, when you give it a new prompt, the AI doesn’t search the web or pull from a database — it imagines what that kind of image should look like based on patterns it’s seen in the past. The result: surprisingly creative and often beautiful artwork, which can be photorealistic, stylized like a painting, or something completely surreal.

A famous example? In 2022, an AI-generated piece created with Midjourney — titled “Théâtre D’opéra Spatial” — actually won first place at the Colorado State Fair’s digital art competition. The judges didn’t initially know it was AI-made. The win sparked huge debate: is this “real” art? Is it fair to human artists?

More recently, AI tools have taken a huge leap forward — they don’t just generate images from scratch anymore. They can also edit, transform, or completely restyle images that you upload. This was a major feature added when OpenAI integrated DALL·E 3 directly into ChatGPT, allowing users not only to generate images but also to upload photos and apply artistic styles or transformations to them. One hugely popular trend? Turning ordinary photos into images that look like they’re from a Studio Ghibli film — soft colours, glowing sunsets, and all the charm of a hand-drawn fantasy. You could upload a picture of your neighbourhood and ask the AI to convert it to “Ghibli style, with a magical atmosphere”, and it would return a version that looks like it belongs in Spirited Away or My Neighbour Totoro.

For students and casual users, these tools are like having a powerful visual assistant at your fingertips. They can help generate inspiration, create presentation graphics, or explore storytelling ideas visually. But for professional artists and designers, these tools raise serious questions. If anyone can create art with a single sentence or turn a photo into a stylistic masterpiece in seconds, what happens to jobs in visual design, illustration, or advertising? Will companies stop hiring artists in favour of quick AI outputs?

Some argue that human creativity will still be essential — to guide, refine, and contextualize what AI produces. Others worry that the value of creative labour is being eroded, or that AI is simply remixing the work of thousands of artists it was trained on, without permission or credit. In short: these tools are revolutionary, but we need to understand their power — and their impact. We'll return to some of these ethical concerns in the next chapter.

Voice and Audio AI: Machines That Speak and Listen

Another area of rapid AI growth is in audio – including speech generation and recognition. Here are a few notable aspects:

Text-to-Speech (TTS) and Voice Cloning: We've had robotic voices reading text for a long time (like the classic GPS voice or screen readers for the visually impaired), but AI has made them far more natural. Modern AI-powered TTS can produce voices that are nearly indistinguishable from a real human's voice, including proper intonation and emotion. For example, services now exist where you can clone a person's voice by providing a few minutes of sample audio, and then the AI can read any text in that person's voice. This could be used for making an audiobook in the author's own voice without the author recording every line, or for dubbing movies in different languages with the same actor's voice tone. It's impressive but also a bit scary because it means you can no longer trust that an audio recording is genuine – someone could synthesize your voice saying something you never said.

Voice Assistants with AI: Voice assistants like Siri, Alexa, Google Assistant have been around, but they are becoming more AI-savvy. Initially, they worked mostly by recognizing a command and mapping it to a specific answer or action. Now, with integration of language models, they can handle more complex queries and have more natural dialogues. For instance, the latest Alexa can have a back-and-forth conversation and maintain context better than before, thanks to AI improvements.

Speech Recognition: AI has also improved converting speech *to* text (automatic transcription). Models like DeepSpeech and now various offerings from companies can transcribe meeting conversations, lectures, etc., with high accuracy. OpenAI's model Whisper (released in 2022) is particularly known for robust speech recognition across many languages and even in noisy environments. This is helpful for generating subtitles, transcribing podcasts, or dictating notes just by speaking.

Music and Sound Generation: AI is also dabbling in music – there are models that can generate musical compositions in the style of Bach, or AI that can produce sound effects or even human-like singing given lyrics and a melody. While this is a bit less mainstream than image or text AI, it's a growing field. You might have seen fun examples on social media where someone makes an AI model of, say, Frank Sinatra's voice singing a modern pop song – those are done with AI voice cloning and music generation techniques.

One concrete example in the voice domain: In 2023, OpenAI gave ChatGPT the ability to speak. They added voice capabilities where you can actually talk to ChatGPT and it can talk back in a natural-sounding voice (this was rolled out in their mobile app with a few different voice options). This basically turns the text chatbot into a conversational AI that feels even more like you're talking to a person. They achieved this by integrating advanced TTS and also speech recognition (so you can talk instead of typing). So, the boundaries between chatbots and voice assistants are blurring.

An innovative development in AI audio technology is Google's NotebookLM, an AI-powered research and note-taking tool. NotebookLM offers a feature called Audio Overviews, which allows users to convert written documents into podcast-style discussions. By uploading documents, such as class notes or research papers, NotebookLM generates an audio conversation between AI-generated hosts who summarize and discuss the content in a conversational manner. This feature aims to make complex information more accessible and engaging, catering to auditory learners and those seeking alternative study methods.

For instance, students can upload their lecture notes into NotebookLM, and the tool will produce an audio overview where AI hosts discuss key concepts and themes from the material. This approach not only aids in comprehension but also allows for convenient on-the-go learning, as users can listen to these AI-generated discussions like they would a regular podcast.

However, the ethical side: with voice cloning, there’s concern about deepfakes – not just in video but audio. Imagine a scam call that perfectly mimics your parent’s voice asking for something – these are not far-fetched because of this tech. It raises the importance of verification and maybe new tools to detect AI-generated audio.

6. Ethical and Societal Implications of AI

With AI systems like ChatGPT and others becoming more widespread, it’s crucial to discuss the ethical issues and broader impacts on society. AI is a double-edged sword – it can do a lot of good, but it also has potential downsides or misuse cases. In this section, we’ll cover some key concerns: misinformation and hallucinations, bias, privacy, plagiarism and the impact of AI on the job market. Understanding these issues will help you be a responsible AI user and future professional in a world with AI.

#1: Misinformation and “Hallucinations”

One of the biggest limitations of AI language models like ChatGPT is something called a hallucination. No, not in the psychedelic sense — in AI, a hallucination means the system makes something up that sounds plausible but isn’t true. The model doesn’t have a fact-checker built in, and it doesn't “know” things the way people do. Instead, it predicts what text is likely to come next based on patterns in its training data.

This can be a serious issue. If you ask ChatGPT a question and it doesn’t know the answer, it might still give you one — and present it with total confidence.

For example, it might generate a fake quote, make up a source, or give you a detailed biography of someone filled with invented information. It’s not lying intentionally — it just doesn’t have an internal compass for truth versus fiction. You can think of it like a student bluffing their way through an exam essay: it sounds good, but it might be completely off.

A now-famous case involved Google’s Bard (now Gemini), which, during its first public demo, was asked about the James Webb Space Telescope. It confidently responded that the telescope had taken the first image of a planet outside our solar system — but that wasn’t true. That photo was actually taken in 2004 by a different telescope. The mistake was widely reported, and Google’s stock even dropped as a result. It was a wake-up call: AI can sound authoritative while being flat-out wrong.

This phenomenon is more than just embarrassing. It ties into the broader societal issue of misinformation. If people assume AI tools always tell the truth — especially when the answers are phrased eloquently and delivered instantly — they might trust and even share false information. And worse, bad actors can deliberately use AI to generate convincing fake news, fabricated research, or even propaganda. That’s why experts and lawmakers are paying close attention to AI’s role in the spread of disinformation.

At the same time, AI is also being developed to help fight misinformation — for example, by detecting deepfakes, verifying facts, or flagging AI-generated content. But it’s an arms race: the tools that create fake content are evolving as fast as the tools that try to detect it.

So what’s being done to reduce hallucinations?

Developers are working on ways for AI to admit when it doesn't know something. For example, newer models are being trained to say “I don’t know” or to avoid giving answers when confidence is low. There’s also research into getting AI to cite real sources, so users can check facts more easily.

Users, on the other hand, need to bring a healthy level of digital skepticism. If something sounds surprising, especially in a technical or sensitive topic, it’s important to verify it using reliable sources. Treat AI like a helpful assistant, not a source of absolute truth.

Even in small ways, hallucinations can cause real harm. Imagine a student asking ChatGPT for a physics formula, getting the wrong one, and then using it on a test. Or someone relying on AI for legal or medical advice without checking with a professional. That’s why hallucinations aren’t just a technical bug — they’re an ethical and social challenge, and one we need to address both through better design and smarter usage.

#2: Bias

One of the most complex and debated ethical issues in AI is bias. Because language models like ChatGPT are trained on huge amounts of real-world data — mostly scraped from the internet — they can pick up and reflect the same stereotypes, inequalities, and imbalances found in society. AI doesn’t have opinions or beliefs, but it repeats patterns, and those patterns might be biased in all sorts of ways: gender, race, culture, politics, profession, and more.

How Does Bias Enter the Model?

Imagine that most of the online content about scientists contains subtle (or not-so-subtle) stereotypes. The model, learning from this data, might then repeat or reinforce those same ideas in its responses. That’s not because the AI “believes” anything — it’s just seen those patterns more often. One early example that was reportedly observed involved ChatGPT generating a rap about scientists, where it allegedly implied that women and scientists of colour were inferior to white male scientists. Although OpenAI didn’t program it to do this, the model may have reflected biases present in its training data.

And bias doesn’t just show up in text. Image generation tools have faced similar issues. In early versions, prompts like “a person in a kitchen” would often result in images of women, while “CEO” would mostly produce images of men — revealing how AI can reinforce traditional gender roles. These associations aren’t necessarily “intentional,” but if left uncorrected, they lock in old stereotypes and present them as normal or expected.

Ideological and Political Bias

In addition to racial or gender bias, studies have also found evidence of ideological bias in AI systems. A peer-reviewed study published in the *Journal of Economic Behavior and Organization* in 2023 found that ChatGPT tends to exhibit left-leaning political bias, even when users attempt to frame questions neutrally. The researchers observed that responses often align with progressive views on topics like gender identity, climate change, and social policy. This likely reflects the fact that the model’s training data includes many mainstream and academic sources, which themselves may lean left. While ChatGPT is not political by design, its output can subtly favour certain worldviews over others — something to be aware of, especially when AI is used to inform debates, education, or public discourse.

Real-World Risks

Bias in AI isn’t just an academic or media issue — it can have real consequences, especially when AI is used in decision-making. In areas like hiring, loan approvals, law enforcement, and education, biased algorithms can lead to unfair outcomes:

- Facial recognition systems trained mostly on lighter-skinned faces have been shown to perform poorly on darker-skinned individuals, leading to higher rates of false identifications — a serious concern in policing.
- AI tools used in recruitment might screen out female candidates if the training data includes mostly male hires.
- Loan approval systems trained on historical data might unintentionally reproduce economic discrimination, even if race or gender aren’t explicitly included as variables.

The danger is that AI can appear neutral or objective while quietly reproducing and amplifying historical inequalities.

The Flip Side: Overcorrecting for Bias

On the other end of the spectrum is overcompensation — when attempts to fix bias create a new kind of inaccuracy. In 2024, Google's Gemini image generator came under fire for doing exactly that. The system was designed to promote diversity in image generation, which is a good goal. But when users asked for images of historical scenes like medieval knights or U.S. Founding Fathers, it often returned racially and gender-diverse depictions that didn't match the actual historical context. The intention was inclusivity, but it backfired — critics accused the system of rewriting history and pushing representation too far at the expense of realism. Google paused the feature and promised improvements. This case serves as an important reminder: addressing bias isn't about just adding diversity — it's about doing so in a way that's thoughtful, accurate, and contextually appropriate.

Our Role as Users

As users, we also have a responsibility in how we interact with AI. These systems are incredibly advanced, but they're not perfect — and they don't "know" right from wrong. If an output seems biased, one-sided, or problematic, it's important to question it, flag it, or even test it by rephrasing the prompt or asking follow-up questions.

The great thing about conversational AI like ChatGPT is that you're not stuck with a single answer. Unlike traditional media — where you get whatever the author decided to publish — here, you can ask for multiple viewpoints. For example, you can say:

"What are the arguments on both sides of this issue?"

This makes ChatGPT not just a source of information, but a tool for critical thinking — if you use it that way. But this requires digital literacy. If we blindly accept whatever the AI says — especially when it sounds confident — we risk being misled. The ethical use of AI starts with curiosity, scepticism, and responsibility. It means knowing how to explore a topic from different sides, and being aware of when you might be nudged by subtle biases.

So while developers work to reduce bias from the inside, we can reduce the impact of bias from the outside — by staying engaged, asking better questions, and never switching off our own judgment.

Try This: Asking for Multiple Viewpoints

"Summarize the liberal and conservative views on universal basic income."

"How might a feminist and a traditionalist view this issue differently?"

"What are some criticisms of the point you just made?"

"Can you explain this using a centrist or neutral perspective?"

"If someone strongly disagreed with this view, what might they say?"

#3: Privacy Concerns

One of the most important, and sometimes confusing, concerns around AI tools like ChatGPT is privacy. What happens to the things you type into the chat? Who sees it? And does ChatGPT "learn" from you?

Let's clear this up once and for all.

Training on Public Data

Before ChatGPT could answer your questions, it had to be trained on massive datasets. These came from public sources on the internet: books, Wikipedia, articles, blogs, forums, and more. But here's the catch: some of that public content may have included personal details that people didn't expect to end up in an AI model. For example, someone might post a story or blog that includes their full name or a real-world address, and ChatGPT could technically learn from that.

This sparked concern among privacy advocates and regulators. In 2023, Italy's data protection authority temporarily banned ChatGPT, saying it might have violated European privacy laws (like the GDPR) by collecting and using personal data without proper consent. Italy questioned how the training data was gathered and whether OpenAI had done enough to protect individuals' privacy.

What Happens to *Your* Chat?

By default, ChatGPT uses your conversations to improve and train its models — unless you opt out. That means when you chat with ChatGPT (free or Plus versions), OpenAI may store and later review that content to help fine-tune future updates of the model. So if you type sensitive or personal information into the chat, it could — in theory — become part of the training dataset. While OpenAI has systems in place to remove personally identifiable information, the risk isn’t zero.

Who Sees It?

OpenAI uses both automated systems and human reviewers to monitor and label a small percentage of chats, especially for safety and misuse detection. So while most of your conversations won’t be read by a person, it *can* happen — particularly if they are flagged or sampled for training. This doesn’t mean OpenAI is spying on you — but it does mean your data isn’t entirely private by default.

How to Stay Private: Tools and Options

The good news is that OpenAI gives users multiple ways to control their data and protect their privacy while using ChatGPT. Whether you're a casual user or working in a business environment, you have options to limit or entirely block how your chats are stored or used to train future versions of the model.

1. Turn Off Model Training (but Keep History)

If you're using the regular ChatGPT interface (either the Free or Plus version), you can choose to opt out of your chats being used for training. To do this, go to Settings > Data Controls and turn off the toggle labelled “Improve the model for everyone.” Once this is disabled, your future conversations will not be used to train the model. However, unless you also turn off history, those chats will still be visible in your sidebar and saved to your account. This is a good option if you want to keep a record of your conversations but don’t want them contributing to model improvement.

2. Use “Temporary Chats” (No History and No Training)

If you want maximum privacy, you can go a step further and disable chat history entirely. This enables a mode often referred to as “temporary chat” — similar to incognito browsing in a web browser. When this mode is active, your conversations are not stored, they do not appear in your chat sidebar, and they are not used for model training. These chats are kept on OpenAI’s servers for only 30 days for safety checks (e.g. to detect abuse), and then they are permanently deleted. This is the safest option if you're typing something sensitive, and it gives you peace of mind that nothing will be saved or reused.

3. Use Enterprise or API Services (Business-Grade Privacy)

For organizations and developers, OpenAI offers ChatGPT Enterprise, ChatGPT Team, and access to GPT models via an API. These services come with a higher level of data protection. By default, none of your data is used for training, and all chats are encrypted and stored privately. These versions are designed to meet professional security standards and are ideal for environments that deal with confidential or proprietary information. That’s why many companies choose to restrict their employees to these platforms — they offer the benefits of AI while still protecting sensitive data.

Why Some Companies Ban ChatGPT

Even though OpenAI offers these privacy options, many businesses — especially in banking, law, healthcare, or software development — restrict employee use of ChatGPT. Why?

Because by default, anything you paste into ChatGPT (like client data or proprietary code) might be stored and reviewed. There have been real cases, like Samsung employees accidentally leaking code into ChatGPT, which triggered panic over data loss.

As a result, some companies have internal rules like:

“Never paste client names, confidential info, or internal code into ChatGPT.”

It’s a precaution — and one that makes sense given how data flows by default.

#4: Plagiarism debate

The rise of AI tools like ChatGPT has sparked a lot of confusion — and debate — about what counts as "your own work." If you write an assignment and ask ChatGPT to fix the grammar, is it still yours? What if you feed it your rough notes and let it rewrite the entire thing? Or what if you type a single prompt, copy the answer, and hand it in unchanged? The truth is, we're navigating a grey zone — and the ethical boundaries are not always clear.

Before anything else, it's important to emphasize one thing: always follow the policy of your specific module or institution. Even if you personally believe your use of AI is ethical or justified, your lecturer or university may disagree. And regardless of your stance, their rules matter. Academic integrity policies vary widely — not just because of individual preferences, but often because of the educational goals behind them. In some cases, instructors may ban AI in an assignment because the task is explicitly designed to assess your personal reasoning or your ability to express ideas in your own words. That's not necessarily about distrust — it's about pedagogy.

In other words, using AI isn't automatically wrong — but it can be inappropriate in the wrong context.

A Spectrum of AI Use

To make sense of the issue, it helps to imagine a spectrum of AI use, ranging from light-touch assistance to full-blown ghostwriting. Most students don't fall neatly into one category — instead, their usage often shifts across a range of scenarios. Below are some examples of how students use AI in their academic work:

- Light Editing: A student writes everything themselves but uses AI to fix spelling, grammar, or awkward phrasing. This is comparable to using a spellchecker — widely accepted and unlikely to be considered misconduct.
- Clarity and Restructuring: Some students go a step further and ask AI to improve the flow or coherence of their writing. The content is still theirs, but the structure may be enhanced by AI suggestions.

- Idea Development and Critique: Others use AI as a “sounding board” — asking for feedback, alternative perspectives, or prompts that help sharpen their arguments. The student still makes the final decisions, but the AI plays a developmental role.
- Partial Rewriting: A student might paste their notes or rough draft into ChatGPT and ask for a rewritten version. The core ideas remain theirs, but the phrasing may shift significantly. Here, authorship begins to blur.
- Full Generation: At the far end of the spectrum, a student types a one-line prompt and submits the AI’s response unchanged. This is not only the most ethically questionable case, but also — ironically — the least effective. Because the AI lacks the student’s context, these responses tend to be vague, generic, and superficial. In most university-level assessments, this kind of submission would earn poor marks.

Are We Grading Thoughts, or Phrasing?

This brings us to a deeper question: What are we really assessing in student work? Are we testing the student’s ideas, or their ability to express them clearly?

In practice, the answer is often both. But it’s worth asking whether we sometimes conflate clarity with intelligence. Some students have brilliant insights but struggle with phrasing. Others can write elegantly but lack original ideas. In this context, AI becomes more than a tool — it becomes a bridge for students who have something to say but lack the confidence or language skill to say it well.

This matters. Because in the real world, many people are excluded from academic, professional, or public conversations — not because their ideas aren’t valuable, but because they aren’t expressed in a way that “sounds smart.” AI could play a powerful role in levelling that playing field.

The “One-Prompt Essay” Problem

Let’s return to that last case — the student who types “Explain stakeholder engagement” and submits the response unchanged.

Aside from the obvious ethical problems (it's not your work, you didn’t do the thinking), the product is usually weak. ChatGPT doesn’t know your course, your readings, your lecture themes, or your instructor’s expectations. Without direction, it defaults to a vague, Wikipedia-style overview. These kinds of responses lack originality, depth, or relevance — and lecturers can spot them instantly.

Using AI well means collaborating with it, not delegating to it. You still need to guide the conversation, refine the output, and apply your own judgment.

Practical Advice

Here are a few takeaways for navigating this evolving space:

- Clarify expectations: Don’t assume. Ask your lecturer whether AI tools are allowed — and to what extent.
- Be transparent: If permitted, consider stating how AI was used (e.g., “ChatGPT was used to improve grammar and clarity.”)
- Avoid over-reliance: The goal of university is to learn how to think. Use AI to support your learning, not to skip it.
- Stay aware: This field is changing fast. New rules, tools, and expectations will continue to emerge.

The Real Challenge: Ownership of Thought

Ultimately, the most valuable thing you bring to any assignment isn’t perfect phrasing — it’s your perspective. Your reasoning. Your ability to engage with an idea, challenge it, and articulate it. These are the things that still make your work uniquely yours, regardless of the tools you use to express it.

This is why it feels increasingly inconsistent that AI use often requires formal disclosure, while other tools — like Excel, spellcheck, or even Grammarly — are freely used without question. Take Excel, for example: students and professionals routinely use it to perform calculations, build dynamic dashboards, run financial models, or automate decision-making processes that would be nearly impossible to do by hand. Yet no one is required to disclose that they used Excel — not in a footnote, not in a cover page, not anywhere.

It’s not just that Excel isn’t listed as a co-author — it’s not even mentioned, despite the fact that in many cases, the work is entirely dependent on its computational power.

So why is AI treated differently? Why are we expected to justify or disclose its involvement when we’re not held to the same standard with other digital tools that also save time, offer suggestions, and enhance our capabilities?

Some might argue that the key difference is that AI is “generative” — that it produces content, not just processes it. But even that line is blurry. A formula that generates a financial forecast, or a citation tool that creates a full bibliography, is also generating intellectual output. And no one considers that plagiarism.

These inconsistencies suggest that much of the discomfort around AI stems not from logic, but from novelty — it’s new, so it’s viewed with more suspicion. But over time, we may need to recalibrate our ethical lens, not just to keep up with technology, but to ensure that we are applying principles consistently across all tools and disciplines.

That’s why the most honest answer is: we don’t know yet. And neither do universities. These policies are still being formed, often reactively, and sometimes with contradictions. You might disagree with a particular rule — and in many cases, you’ll have valid reasons to do so. But as long as you are part of a course or institution, your safest bet is to understand and follow the rules as they are written, while also engaging in the larger conversation about how they could be improved.

#5: Impact on Jobs and the Future of Work

One of the biggest societal questions raised by AI’s rapid advancement is: How will it affect jobs and the economy? On the one hand, there’s excitement about improved productivity and efficiency. On the other, there’s growing concern about job displacement. This tension isn’t new — every major technological wave, from the Industrial Revolution to the rise of the internet, has brought a version of this same question: *Will machines replace human workers?*

What makes today's situation different is the kind of work AI can now do. Unlike previous technologies that mostly automated physical labour, modern AI — especially cognitive AI like ChatGPT — is reaching into tasks we once thought were uniquely human: writing, analyzing, planning, even engaging in conversation. It's no longer just factory work that's on the line — it's also office work, creative work, and knowledge work.

The Optimistic View: Transformation, Not Elimination

Let's start with the more hopeful outlook. AI is expected to augment many roles, not outright replace them. For example:

- A customer service agent could handle more clients by letting AI respond to common queries while they focus on the tricky cases.
- A lawyer might use AI to draft contracts quickly and then personalize them for specific clients.
- A marketer could use AI to generate initial copy or visuals, and then refine them with their own creative flair.

This trend is often called job augmentation — where AI becomes a copilot, not an autopilot. Tools like GitHub Copilot (for programmers) or Microsoft 365 Copilot (for Word, Excel, PowerPoint, etc.) reflect this idea: AI helps you work faster and smarter, but the human is still in control.

The 2023 Goldman Sachs report estimated that up to 300 million full-time jobs could be affected by generative AI in some way — but it emphasized that most of these roles will be transformed rather than eliminated. In other words, you might still have the job title, but how you do your work could change dramatically.

And historically, that's been the pattern. New technologies often eliminate some tasks, but they also create new roles we couldn't have imagined before. Think about this: around 60% of today's jobs didn't exist in 1940.

Already, AI is spawning careers like *AI ethicists*, *prompt engineers*, *data labelers*, and *model trainers*.

But... What If This Time *Is* Different?

Not everyone shares the optimism. Some economists, technologists, and philosophers argue that this wave of automation may go deeper than anything we've seen before — because AI isn't just replacing physical effort or repetitive office work. It's also getting better at tasks involving language, judgment, and even creativity.

ChatGPT can now:

- Write news articles and poems
- Analyze legal documents
- Generate business ideas
- Code basic software
- Tutor students in math or science
- Simulate human conversation convincingly

If AI continues to improve at this rate, the fear is that even high-skilled, "thinking" jobs might be at risk. A single person with a powerful AI assistant could do the work of five or ten people. In a competitive market, that might lead businesses to cut staff not because the work is gone — but because AI makes fewer workers necessary.

This leads to some uncomfortable but important questions:

What if AI becomes so efficient that companies reduce their workforce just to save money?

What if "creative" roles become so AI-assisted that fewer human creatives are hired?

What if the productivity gains from AI are captured mostly by big tech firms, widening the wealth gap?

In this more pessimistic scenario, inequality could grow, and job loss could be more severe than in previous tech revolutions. Some experts believe we'll need radical responses — like universal basic income (UBI) or new policies to protect human labour — if we're to navigate these disruptions fairly.

No One Knows for Sure

The truth is: no one can predict exactly how this will play out.

Maybe we'll see a new golden age of productivity, where AI frees people to do more meaningful, creative work. Maybe we'll see major societal upheaval unless we make bold changes. The only thing we know with certainty is that uncertainty itself is now part of the equation.

So what can you do? As future graduates, your best strategy is to be adaptable and proactive. Learn how to use AI tools effectively. Understand what they can and *can't* do. Focus on building the kinds of skills that remain valuable in any future: empathy, critical thinking, communication, leadership, collaboration.

The job landscape will shift — but humans who can work alongside AI will be in the best position to thrive.

At least We'll Always Have Creativity… Right?

We often think of creativity as something uniquely human — a magical spark of inspiration, imagination, or emotion that machines can never replicate. But here's a provocative question: What if creativity isn't magic at all? What if it's just clever remixing?

Let's take music as an example. Western music is built on just 12 notes. That's it. Every symphony, pop song, movie soundtrack, and jazz solo you've ever heard uses some combination of these same notes. What makes them feel new or emotional or groundbreaking isn't the ingredients — it's how those ingredients are arranged.

Composers rearrange the same notes in different orders, rhythms, and harmonies. They manipulate timing, layering, repetition, and dynamics. They draw from past influences, cultural traditions, and personal experiences. And despite the limited raw material, they can still create infinite variety. The same is true for poets using only the 26 letters of the alphabet, or chefs using a fixed set of ingredients. Creativity thrives within constraints.

Now consider what AI does.

Large language models like ChatGPT, or image and music generators like DALL·E, Midjourney, or Suno, are trained on vast libraries of human-created content. They've "seen" more books, paintings, songs, and styles than any single human could in a thousand lifetimes. And when they create something, they do so by drawing on all those patterns — blending, tweaking, recombining — in a way that *can look and feel* astonishingly creative.

If creativity is partly about having lots of reference points and recombining them in surprising ways, then AI might already be functionally creative, even if it isn't conscious. It can generate strange and beautiful images, write poetry, suggest melodies, design logos, or compose music that sounds fresh and new — just like a human artist might.

Of course, there's still something deeply human about the intent behind creativity — the emotional message, the cultural context, the personal story. AI doesn't have feelings or purpose. It doesn't *mean* anything. But then again, not every human creation is meaningful either. Sometimes a song is just catchy, a joke is just funny, and a painting is just pretty.

So maybe the real question isn't whether AI can be creative, but: What kind of creativity matters to us?

Do we value the *process* or the *product*?

Does it matter *who* made something if it still moves us?

These are big questions — and they're not just philosophical anymore. As AI gets better at creating, we'll need to rethink not only what it means to be an artist, writer, or composer… but what it means to be original at all.

#5: The Future of AI: What Lies Ahead

Finally, let's cast our eyes to the horizon and talk about the future. What might AI look like in 5, 10, or 50 years? This is a mix of educated speculation and current directions of research. We'll discuss a few exciting (and sometimes eerie) concepts: the pursuit of Artificial General Intelligence (AGI), the development of brain-machine interfaces, and the evolving relationship between humans and AI – could it become a partnership, a merger, or something else entirely?

Towards Artificial General Intelligence (AGI)

Today's AI, including ChatGPT, is powerful but still narrow in certain ways. ChatGPT can't directly drive a car or perform surgery or truly understand the physical world – it's limited to language tasks. Artificial General Intelligence (AGI) refers to an AI that has a level of intelligence equal to a human's, in a broad sense: able to perform any intellectual task a human can, and to generalize knowledge between different domains. In other words, an AGI would be not just a specialist, but a generalist problem-solver with common sense understanding of the world.

AGI is still a hypothetical concept – we don't know exactly how to build it, and there's debate among experts about how far away it is. Some optimists in the AI community (often at companies like OpenAI or DeepMind) think we might achieve something like AGI within a couple of decades or even sooner, given the rapid progress. They see scaling up models and new algorithms eventually leading to human-level cognition. Others are more sceptical and think AGI could be many decades away, or that we might need fundamentally new ideas (not just bigger neural networks) to get there. It's a hot topic of discussion: every time AI takes a leap (like GPT-4 acing certain exams, or AI beating humans in more games), people speculate we're edging closer to AGI.

It's important to clarify that AGI is not guaranteed – it's an end goal some are aiming for, but we're not sure if current AI approaches will directly get us there. By one definition, AGI means the AI could learn and adapt to handle completely new tasks it hasn't encountered, just like a person can figure things out given time. For example, a human who knows English and Chinese can pick up Spanish, but an AI model trained only on English and Chinese might flounder with Spanish without retraining. AGI would ideally have that kind of adaptability and cross-domain knowledge.

The implications of AGI are huge. If achieved, it could potentially accelerate scientific research (imagine an AI as innovative as a top scientist, working tireless 24/7 – it could help cure diseases or solve climate challenges rapidly). On the other hand, some people like Elon Musk and the late Stephen Hawking have warned that super-intelligent AI (beyond AGI, sometimes called Artificial Superintelligence) could pose existential risks if not aligned with human values – the classic sci-fi trope of an AI that goes out of control. This is why you might have heard of efforts in AI safety or ethics aimed at ensuring advanced AI will be beneficial and under control. In 2023, there were even open letters from some tech leaders calling for a pause on training the most powerful models until safety regulations catch up.

It's a bit like we're on the brink of something that could be amazingly positive or potentially dangerous, depending on how it's handled. For first-year students like you, who may have long careers ahead, it's not inconceivable that you will see an AGI in your lifetime. Or you might even contribute to building it, who knows! But even without true AGI, each step toward it is bringing more capable and autonomous AI systems.

Brain-Machine Interfaces: Merging Mind and Machine

Another futuristic thread gaining momentum is the development of brain-machine interfaces (BMI)—technology that connects our brains directly to computers or artificial intelligence. At first glance, this might sound like pure science fiction, but the truth is, basic forms of BMI already exist today. For example, implants allow paralyzed individuals to control robotic limbs or move computer cursors simply by thinking. These technologies work by reading electrical signals from the brain and translating them into digital commands.

Elon Musk's Neuralink company exemplifies this ambitious field. In 2023, Neuralink received approval in the United States to start human trials of its advanced brain implant. The initial goal is medical: enabling people with paralysis to operate computers, move prosthetics, or potentially restoring vision to the blind. But Musk’s broader vision reaches further. He argues that as artificial intelligence becomes increasingly powerful, humans might need to merge with AI to remain relevant. Put simply: if you can't beat them, join them.

Imagine a future where instead of typing questions into a search engine, you simply think them and receive answers directly into your brain. Or, in a scenario straight from The Matrix, you could "download" a new skill instantly. Even if these radical visions are distant, nearer-term BMIs could revolutionize how we interact with devices, removing the need for keyboards, voice commands, or even gestures. A mere thought could trigger actions in the digital or physical world.

But BMIs raise deeper, more philosophical questions about the very nature of human thought and communication. There's a longstanding debate: Does thought shape language, or does language shape thought? This is known as the Sapir-Whorf hypothesis.

Cognitive psychologist Steven Pinker suggests we think in a raw, non-verbal language he calls "mentalese." According to this theory, our brains first generate abstract thoughts, which we then translate into spoken or written languages. But languages differ significantly—some can express ideas and emotions that others cannot, meaning our thoughts may get distorted or simplified in translation.

Advanced brain-machine interfaces could fundamentally alter this dynamic by tapping directly into these pre-linguistic thoughts. Imagine two people connected via a BMI network, able to understand each other's thoughts, emotions, and experiences with unprecedented clarity, bypassing the constraints and imperfections of language entirely. This could lead to new depths of understanding each other.

However, these technologies come with potentially unsettling implications. With advanced BMIs, it might become possible to directly stimulate the brain with fully immersive digital experiences indistinguishable from real life. Imagine living in carefully crafted simulations, experiencing adventures, relationships, and emotions, all without leaving your chair. At first glance, this might sound appealing—endless happiness at your command—but it raises profound questions about what it means to be "real" and how detachment from physical reality could impact mental health.

Another provocative scenario involves forming deep, meaningful relationships not with humans, but with AI. With BMIs, artificial intelligence could offer constant companionship, tailored perfectly to your desires, never disagreeing or disappointing you. At first, this might seem attractive, especially to individuals feeling lonely or misunderstood. But what would the psychological impact be of engaging primarily with entities designed to agree with us? Could we lose the ability to handle disagreements, complexities, and unpredictabilities of genuine human relationships?

There's a risk that such seamless interactions with agreeable AI could lead to emotional stagnation and a withdrawal from authentic human connection. Over time, this could erode crucial social skills and emotional resilience, leading to unprecedented isolation despite being "connected."

Of course, this all sounds wildly futuristic and borderline absurd—after all, who could imagine humans willingly choosing to detach from reality to live in carefully curated digital bubbles? Anyway, let's quickly check Instagram and TikTok again; it's been at least 10 minutes!