Study Guide Transcript

2025

*This study guide transcript has been provided to support learners in following the **Introduction to AI** course.*

*While the guide serves as a useful resource, we highly recommend that learners watch the course episodes on the **Wayout TV channel** or via the **Video-on-Demand** service to gain a full understanding before completing the answer book.*

*For your convenience, episode times are listed on **page 4 of the answer book**.*

Episode 1 – Video 1 : Introduction to AI and ChatGPT

Introduction to AI and ChatGPT

You've heard of ChatGPT, and it's likely you've used it. But how does it actually work? To understand that, let's look at how it fits within the field of artificial intelligence.

In 1956, the term artificial intelligence (AI) was first coined. A small group of experts met at the Dartmouth Summer Research Project and defined AI as the study of creating machines that can perform tasks that typically require human intelligence.

Today, AI is everywhere—in Gmail spam filters, Siri and Alexa voice assistants, and YouTube's recommendations. These are examples of narrow AI—AI designed for specific tasks. For instance, the AlphaGo AI beat the human Go champion in 2017.

There is also the concept of general AI (artificial general intelligence). Experts don't agree on a precise definition, but generally, it refers to a machine with intelligence equal to, or close to, human intelligence—the ability to solve problems, learn, and plan for the future.

You may have heard scary science-fiction stories about AI taking over the world. ChatGPT is nothing like that. It is designed only to generate text in a conversational style.

ChatGPT is part of deep learning, a subfield of machine learning, which itself is a subfield of AI.

Machine Learning

Machine learning involves developing computer models that automatically learn and improve from data. These models make predictions based on patterns they've learned, solving problems without needing to be explicitly programmed step by step. This is very different from rules-based programming.

Deep Learning

Deep learning uses neural networks with multiple layers of interconnected nodes. This structure allows the model to learn very complex patterns.

Neural Networks

A neural network is inspired by the human brain. It consists of processing units called nodes, organized into layers. Nodes take input, perform mathematical computations, and produce output.

These networks have billions of parameters, making it difficult even for their creators to fully understand how they generate particular outputs.

Key Point about ChatGPT

ChatGPT is not a database of words and sentences. Instead, it's a vast network of mathematical operations that define the probabilities of which words come next.

✅ Summary

- AI: Field of creating machines that perform human-like tasks.
- Narrow AI: Task-specific (e.g., spam filters, AlphaGo).
- General AI: Human-level intelligence (still theoretical).
- Machine Learning: Models that learn patterns from data.
- Deep Learning: Multi-layer neural networks for complex patterns.
- Neural Networks: Brain-inspired systems with nodes and layers.
- ChatGPT: Generates text by predicting word probabilities, not by storing answers.

Episode 1 – Video 2: How ChatGPT was built and trained.

In 2017, an important breakthrough in AI research occurred with the development of **transformer architecture**. This allowed AI models to look at the whole context of words within sentences, rather than just one word at a time.

By 2018, companies like Google and OpenAI began building **large language models (LLMs)** using this transformer architecture. These models were trained on vast amounts of text and could generalize to many new domains, instead of being limited to a single task like playing Go.

ChatGPT was released by OpenAI near the end of 2022. Its developers added an easy-to-use chat interface and made it freely available to the public.

It was trained on enormous datasets, including:

- The entire English-language Wikipedia
- Internet data up to mid-2021
- Reddit conversations (with at least 3 upvotes)
- Two collections of digital books

A key part of training involved **humans rating responses**, helping the model learn which answers were more conversational and human-like. This process is called **Reinforcement Learning from Human Feedback (RLHF)**. It's similar to dog training—the model is "rewarded" for better answers.

Once training is complete, the dataset is set aside. The model doesn't store or copy the data. Instead, it has transformed the data into mathematical patterns and probabilities.

When ChatGPT generates text, it uses this probabilistic math to predict which words are most likely to come next. This process is known as **generative AI**.

Remember: ChatGPT is not a database of sentences—it's a network of math operations that determine word probabilities.

However, there are known problems:

- **Hallucinations** (making up information that isn't true)
- **Bias** in responses, depending on the training data

✅ **Summary**

- **2017 breakthrough**: Transformer architecture → models can understand word context.
- **LLMs (2018)**: Built by Google, OpenAI, trained on massive text datasets.
- **ChatGPT release**: End of 2022, free chat interface for public use.
- **Training data**: Wikipedia, Internet (to mid-2021), Reddit (3+ upvotes), digital books.
- **RLHF**: Human feedback shapes responses to be more natural.
- **Generative AI**: Creates new text using probabilities, not stored sentences.
- **Limitations**: Sometimes makes up facts (hallucinations) and shows bias.

Next Video Preview:
We'll explore why ChatGPT sometimes gives incorrect or biased answers, and what can be done about it.

Episode 2: How does ChatGPT aim to prevent harmful use?

Guardrails & Harmful Use

- ChatGPT avoids harmful use through filters called guardrails, which block content like instructions for making bombs or other dangerous activities.
- Feedback from users helps improve these guardrails over time.
- If a response seems harmful, users can give a thumbs down and submit a comment.

Probabilistic Responses & Bias

- ChatGPT generates text based on patterns learned from vast amounts of data.
- Responses are probabilistic, so the same question may get slightly different answers each time.
- Because training data comes from the internet, answers can be biased.
    - Example: A story about career choices might show a boy in engineering and a girl in teaching.
- Bias can be reduced by writing specific, non-stereotypical prompts.

Hallucination & Limitations

- ChatGPT predicts likely next words and can get facts wrong.
- Early versions couldn't search the web and relied on training data, which could be outdated.
- If it couldn't generate the correct answer, it might produce a plausible-sounding but false answer — this is called hallucination.
    - Example: A lawyer used ChatGPT for cases that didn't exist.
- Without web search, ChatGPT was more of a wordsmith — good for ideas or phrasing, not fact-checking.

Web Search & Hybrid Models

- Current versions can search the web, reducing hallucinations.
- However, outdated or blocked sources can still lead to errors.
- Users must verify information with reliable sources.
- Other AI tools:
    - Perplexity, Gemini, Claude → search web and summarize results with source links.
    - Elicit, Scispace → search scholarly databases and summarize papers.
- ChatGPT has evolved from a wordsmith to a hybrid system, combining probabilistic text generation with factual search results.

✅ **Summary**

- Guardrails: Filters to prevent harmful or biased outputs.
- User Feedback: Helps improve safety over time.
- Probabilistic Model: Answers vary and can be biased.
- Bias Mitigation: Use specific, non-stereotypical prompts.
- Hallucination: False but plausible answers, especially before web search capability.
- Hybrid AI Systems: Modern models combine probabilistic generation with factual web search.

- Verification Needed: Always double-check AI outputs with reliable sources.
- Other AI Tools:
    - Web search: Perplexity, Gemini, Claude.
    - Scholarly search: Elicit, Scispace.

Next Video Preview

Next video will cover generative AI and other models besides ChatGPT.

Episode 3: Generative AI and other Models

Have you heard the term Generative AI?
That's AI that can create new content such as text, images, video, music, or speech.

ChatGPT is a multimodal model, which means it can:

- Generate text
- Create images
- Write computer code
- Act as a voice assistant
- Recognize and describe uploaded images

There are also models designed mainly for generating multimedia (images, video, music, and speech). These will be covered in another video.

The other major type of AI is Discriminative AI.

- Generative AI → Creates new content.
- Discriminative AI → Classifies, predicts, or recognizes patterns in existing data.

Examples of discriminative AI:

- Netflix's recommendations for what to watch next
- Gmail's spam filtering

It's useful to ask: *Is the AI generating new content, or classifying existing data?* These are very different systems, each with unique strengths and weaknesses.

Other Generative AI Models

ChatGPT is not the only text-generating model. Others include:

- Microsoft Copilot
- Google Gemini
- Claude (Anthropic)
- Perplexity AI

There are also open-source language models such as:

- Llama
- Qwen
- Mistral
- Gemma
- DeepSeek

Open-source models allow researchers worldwide to experiment, improve, and adapt them. Some can even run on smaller systems like a gaming PC or smartphone. While they are often less capable than large commercial models, they hold strong future potential.

AI Apps Built on Models

Since many models (including ChatGPT) have APIs, developers worldwide are building countless apps. These apps support:

- Writing and studying
- Creating slides and presentations
- Analyzing data
- Generating images and videos
- Working with music and speech

There are now so many AI-powered apps that it's hard to keep track of them all.

✅ **Summary**

- Generative AI: Creates new content (text, images, audio, etc.).

- Discriminative AI: Recognizes patterns and classifies data.
- ChatGPT: Multimodal — handles text, code, images, and speech.
- Other models: Copilot, Gemini, Claude, Perplexity, and open-source options like Llama, Mistral, Gemma, DeepSeek.
- Open-source models: Enable innovation and can run on smaller devices.
- AI ecosystem: Exploding with apps for writing, studying, data, media, and more.

Episode 4: Using ChatGPT effectively

To use ChatGPT effectively, it's important to understand what it's good at and what it's not good at.

What ChatGPT Is Good For

ChatGPT generates text based on patterns it learned from large datasets. Because its answers are probability-based, it is especially useful for generating ideas and working with text.

Examples of good uses:

- Idea generation: Ask for potential research paper topics.
- Keyword generation: Request keywords for searching in library databases.
- Summarization & outlining: Condense academic papers or other texts.
- Explanation: Simplify complex concepts or specialized jargon.
- Code assistance: Write or debug computer code (with verification by running the code).
- Better summaries: Instead of "Summarize this," try:
  "Summarize this and quote the sentences that best illustrate the overall themes. Identify the most surprising ideas."
- Simplifying difficult texts: Paste in a dense paragraph and ask for it to be explained in simpler terms, or define unfamiliar words.

All these tasks involve creative engagement with text, rather than straightforward fact-finding.

What ChatGPT Is Not Good For

Even though newer versions can search the web, hallucinations (made-up answers) still happen. Other limitations include:

- Outdated or incomplete information: Web access may be blocked or the index may be outdated.
- Facts with high stakes: Not reliable for health, financial, or legal advice.
- Academic research starting point: Because it sometimes invents sources, it's unreliable for scholarly research.

Instead, use:

- Library databases or Google Scholar for comprehensive research.
- Specialized AI tools like Elicit or Scispace, which use semantic search to summarize scholarly papers. These can supplement traditional tools but don't replace them.

Key Point

ChatGPT is best used for creative text-based tasks (ideas, summaries, simplification, explanations), not for authoritative fact-finding or critical decisions.

✅ Summary

Strengths: Generating ideas, keywords, summaries, simplifications, explanations, coding help.

- Weaknesses: Can hallucinate, outdated info, unreliable for high-stakes facts or academic research.
- Best practices: Verify facts with reliable sources; use AI research tools as supplements, not replacements.
- Academic tip: Start research in databases/Google Scholar, then use AI tools for summarization and exploration.

Next Video Preview:
We'll cover prompting strategies—how to write effective inputs for better results with ChatGPT.

Episode 5: AI Media Production: Generative Creativity.

Introduction

Welcome to the world of **AI generative media production**—a set of creative tools that are transforming the media industry. Think of them like new instruments: a paintbrush, a camera, or a musical tool that helps people create in totally new ways.

AI for Images

- AI can generate **detailed images from simple descriptions**.
    - Example: "A futuristic city at sunset."
- Can **edit existing pictures**: add/remove objects (e.g., take someone out of a group photo, add a dog to a landscape).
- Can **apply artistic styles**: turn a photo into the style of a famous painting.
- Built-in **safety rules** prevent harmful or misleading images, especially of real people without consent.

AI for Sound & Music

- Can **compose songs, melodies, or backing tracks** from a few ideas.
- Helps musicians with inspiration and production.
- Can generate **synthetic voices** for audiobooks, games, or accessibility.
- Can **clone voices with consent**, supporting people who've lost the ability to speak.
- Importance of **consent**: essential to avoid misuse.

AI for Video

- Can generate short clips from descriptions (e.g., "a dog chasing a ball in a park").
- Can **animate still photos** (make a person blink, river flow).
- In film, can:
    - Make actors look younger
    - Create digital versions for special effects
- **Deepfakes**: misleading or deceptive AI-generated videos—serious concern when used without consent.
- When used responsibly, these tools can enhance entertainment and education.

Impact on People and Professions

- **Professionals**: Work faster, explore more ideas (e.g., designers creating multiple logo concepts).
- **Society**: Creativity becomes more accessible—no need to be a trained painter to make art.
- **Challenges**:
    - Distinguishing real vs. AI-generated media (especially in news).
    - Ethical and social trust issues.
- **Key perspective**: AI should be seen as a **partner in creativity**, not a replacement.

Legal Considerations

- **Copyright basics**: Normally, creators own the rights to their work.
- With AI:
  - Works generated entirely by AI prompts may **not qualify** for copyright (depends on jurisdiction).
  - If humans **edit, modify, or transform** AI outputs, copyright may apply.
- **Artistic styles**: Styles themselves (e.g., Impressionism) cannot be copyrighted.
- **Risks remain**: Plagiarism and infringement still apply if AI content directly copies or unlawfully imitates someone else's work.

✅ **Summary**

- **AI images**: Create from prompts, edit photos, apply artistic styles.
- **AI music & sound**: Compose new music, generate voices, voice cloning (with consent).
- **AI video**: Create clips, animate photos, de-age actors, enable special effects.
- **Risks**: Deepfakes, misleading content, lack of transparency.
- **Impact**: More accessible creativity; professionals use it as a productivity tool.
- **Law**: Copyright often requires human input; styles aren't protected, but plagiarism laws still apply.
- **Mindset**: Use AI as a creative **partner**, not a replacement.