AI: A Brief Historical Introduction and Guide to Getting it to Work Right

by James Kupetz and Karl Levy

A Very Brief History of AI

Artificial Intelligence (AI) refers to machines or systems that can perform tasks that would typically require human intelligence. These tasks include problem-solving, understanding language, recognizing images, making decisions, and even learning from experience.

The roots of what we now call AI can be traced back to early 20th-century work in mathematics and logic. One key figure was Alan Turing, whose groundbreaking 1936 paper on computable numbers laid the foundation for modern computing. In 1950, Turing published his famous essay, "Computing Machinery and Intelligence," in which he proposed the Turing Test to assess a machine's ability to exhibit intelligent behavior equivalent to or indistinguishable from that of a human, i.e., to think and communicate like one of us.

The term "artificial intelligence" was first coined by John McCarthy at the now famous 1956 Dartmouth Conference, where leading scientists gathered to discuss the possibility of creating machines that could simulate human thinking. Early AI systems were based on symbolic reasoning, essentially using rules and logic to solve problems. For example, the Logic Theorist, developed by Allen Newell and Herbert Simon in the 1950s, was one of the first programs that could prove mathematical theorems by mimicking human problem-solving.

For decades, AI research followed a "rule-based" approach, where experts programmed machines with explicit deterministic knowledge and decision-making rules. However, these systems struggled when faced with more complex, real-world problems. In the 1980s, researchers began to explore neural networks, inspired by the human brain's ability to learn from experience. This approach gained momentum in the 2000s, thanks to advances in computer processing power and access to vast datasets, giving rise to modern techniques like deep learning.

Today, AI is a rapidly advancing field, its methods being used to power technologies as diverse as self-driving cars, speech recognition systems, and recommendation algorithms used by companies like Tesla, Amazon and Netflix. However, as AI's uses expand so to do the ethical and societal questions about AI, questions about biases in algorithms, the future of work, and how to ensure that AI systems act responsibly and fairly.

One example of AI you may encounter daily is voice assistance, like Apple's Siri or Amazon's Alexa. Such systems use AI to understand your spoken language and respond appropriately. When you ask, "What's the weather like today?" AI processes your words, understands the question, and retrieves relevant information, all within a few seconds. While these systems seem intelligent, they are examples of narrow AI, which excels at specific tasks but lacks general reasoning abilities.

Types of AI

AI is broadly classified into three types based on its level of intelligence and capabilities: Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI), and Artificial Superintelligence (ASI).

Artificial Narrow Intelligence (ANI), also known as Weak AI or Narrow AI, refers to AI systems designed to perform a specific task. These systems are very good at what they do but can only operate within the narrow confines of that task. For example, image recognition software like Google's image search can quickly identify and categorize images, but it cannot understand or interpret the broader context of the image. Self-driving cars can navigate streets and obey traffic laws, but they cannot perform unrelated tasks like cooking dinner or composing music. The majority of AI systems in use today fall under the category of Narrow AI. They are highly specialized and, although powerful, are not capable of performing tasks outside their designated function.

Artificial General Intelligence (AGI), also known as Strong AI, is the concept of a machine that possesses the ability to understand, learn, and apply intelligence across a broad range of tasks—much

like a human being. An AGI system would be able to solve problems, understand language, recognize patterns, and exhibit creativity across diverse domains without needing explicit programming for each task. While AGI remains a theoretical goal, they are still far from achieving it. Currently AI systems lack the ability to reason about the world in the flexible, adaptable way humans do. For example, imagine an AI that could not only recognize a picture of a dog but could also understand what it means to pet a dog, train a dog, or identify what might be harmful to a dog—all with no additional training. This kind of broad, adaptable intelligence is the goal of AGI.

Artificial Superintelligence (ASI) refers to a machine that can surpasses human intelligence in all areas—creativity, problem-solving, emotional intelligence, and even social skills. ASI would potentially be able to solve problems that humans are incapable of understanding and could have vast implications for fields like science, economics, and ethics. However, ASI is a speculative concept, and its development raises many concerns about safety, control, and ethics. At present, ASI is a distant goal/fear existing more in the realm of science fiction than in reality. Nonetheless, researchers are actively discussing what safeguards should be in place to ensure that superintelligent machines do not pose a threat to humanity.

A specific example of ANI: Large Language Models (LLMs)

One of the most exciting advancements in modern AI is the development of Large Language Models (LLMs), such as GPT-3 and BERT. These models are capable of processing and generating human-like text by analyzing patterns in vast amounts of text data. Later on in this module we will discuss how best to use them.

Limitations of and Misconceptions about AI

While AI can be incredibly useful, it is important for us to recognize its limitations:

- AI lacks general intelligence. Current AI systems, including LLMs and other advanced models, possess narrow intelligence; they excel at specific tasks but lack the general cognitive abilities of humans. For example, an AI that excels at playing chess cannot transfer its skills to tasks like driving a car or cooking a meal.
- AI is data dependent. AI systems rely heavily on large amounts of data for training. The quality and bias of this data can significantly impact the performance and fairness of AI models. For instance, facial recognition systems trained on unbalanced datasets may perform poorly on underrepresented groups.
- AI's interpretability and transparency can be concerning. Many AI models, especially deep learning models, are often seen as "black boxes" due to their complexity. That is, even the people who have engineered an AI model are not exactly sure how their model work. This lack of interpretability can pose challenges in understanding and trusting AI decisions. For example, it can be difficult to explain why a deep learning model made a specific medical diagnosis.
- AI's use gives rise to ethical concerns. AI systems can inherit and amplify biases present in training data, leading to unfair or discriminatory outcomes. Ensuring ethical AI development and deployment remains a critical concern. An example is hiring algorithms that may favor candidates based on biased historical data, perpetuating historical injustices.
- AI's use gives rise to new security and privacy risk. The use of AI in various applications raises concerns about data security and privacy. AI systems can be vulnerable to attacks and misuse, necessitating robust security measures.

We should also be aware of some of the most common misconceptions about AI:

- People are often afraid that AI will replace all of our jobs. While AI may transform the job market, it is unlikely to replace all jobs. Instead, AI is expected to augment human capabilities

and create new opportunities. For example, AI can handle repetitive tasks, allowing humans to focus on more creative and strategic activities. The industrial revolution in general has done so too, though perhaps at a slower pace, over the last 300 years.

- People often think that AI is sentient, but, despite advancements, AI systems do not possess consciousness or self-awareness. They operate based on algorithms and data without understanding or experiencing emotions. For example, a chatbot can simulate empathy but does not *feel* emotions. It has no limbic system.
- There are those who may think that AI systems are perfect and can't make mistakes, but especially when faced with unfamiliar or ambiguous situations, this is not true. Continuous monitoring and improvement are essential for AI systems. They always need more real-world data! For instance, an autonomous vehicle that has been programmed to drive in sunny southern California may struggle to navigate in snowy Alaska.
- There are those who may think that AI Can Do Anything, but AI has limitations and is not a universal solution. It excels in specific tasks but struggles with tasks requiring general knowledge, creativity, and common sense. For example, AI can generate text but may struggle with tasks requiring deep understanding and context of that text. Wise human supervision, editing and intervention is and, we believe, will always be the key to its effective and safe use!

A Guide to Prompt Engineering.

Now that we've discussed a bit about what AI is and what it can do, let's get into how to go about using it to do something. Specifically, how best to interact with large language models. Here is some useful tips.

A. Define your Objective Clearly

What is it you want the AI to do for you? Start with a clear goal in mind. Are you seeking information or do you want more? Do you want the AI to do creative writing or generate code? Is there something else you need? Before you start, answer these questions for yourself and then be explicit about what you want the AI to do including the output format desired. Should it be a paragraph, a list, bullet points, C++ code or something entirely different?

Example of ineffective query: Tell me about AI.

Example of more effective query: Provide a detailed explanation of AI, focusing on its applications in healthcare and list three examples.

B. Provide Context

Add background details to help the model understand the subject or purpose. If you are depending on a specific perspective or tone, mention it.

Example of ineffective query: Write a story.

Example of a more effective query: Write a story about a detective, from the perspective of his assistant. The tone should be in the style of an American film noir from the late 1940s.

C. Be Specific and Avoid Ambiguity

Avoid vague or overly broad prompts. Provide specific details such as tone, style, length, or complexity.

Example of ineffective query: Summarize this text.

Example of more effective query: Summarize this text in three sentences, focusing on the main argument and excluding examples.

D. Use Examples or Templates

Provide examples of the output you're looking for to set clear expectations.

Example:

Prompt: Create a three sentence poem about nature in the style of Haiku.

Prompt example: Morning dew sparkles,

Sunlight filters through the leaves,

Whispers of the breeze.

E. Require Step-by-Step Instructions

Break down complex requests into smaller tasks. Ask for step by step reasoning if relevant.

Example of ineffective inquiry: Add 5Ω resistance to 5Ω inductance.

Example of more effective inquiry: Show the method of adding 5Ω resistance to 5Ω inductance step by step.

F. Experiment and Refine

Start with a basic prompt and iteratively refine it based on the output. Test multiple versions of the prompt to determine which works best.

Initial Example: Write a marketing slogan for a coffee shop.

Refined Example: Write a catchy and upbeat marketing slogan for a cozy coffee shop that emphasizes organic, locally sourced ingredients.

G. Leverage "Few-Shot" Examples

If the task requires learning from examples, include a few labeled examples in the prompt. For simpler tasks, specify the rules clearly without examples.

Few-Shot Example: Convert the following sentences into passive voice: “The cat chased the mouse” -> “The mouse was chased by the cat.” Now convert: “The chef cooked a meal.”

H. Use “Zero-Shot” Examples

If a task requires recognizing an object, provide means of recognizing them, without providing an example of the actual object.

Example:

If you want an AI which is already able to recognize cars to be able to recognize sports cars, provide it with images of spoilers, air dams, etc, that make the category “sports car” stand out from the category car.

I. Use Constraints and Limits

Impose constraints like word counts, character limits, or stylistic guidelines.

Ineffective Example: Write about climate change.

Effective: Write a 100 word persuasive paragraph on why renewable energy is critical to addressing climate change.

J. Ask for Iterations or Refinements

If you’re not satisfied with the first output, specify changes for improvement.

Example Prompt: Revise this answer to make it more concise while keeping the formal tone.

K. Consider the Audience

Tailor the language and style to suit the intended audience (e.g. technical professionals, children, casual readers).

Example:

For children: Explain photosynthesis in simple terms for a 10-year old.

For experts: Provide a technical overview of photosynthesis, including its biochemical stages.

When it comes down to good prompt engineering, practice helps. Experiment with different wording and build upon what works for your specific needs.

A Brief Bibliography and Recommendations for Further Reading

- Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies.* 2014. (An examination of the potential impacts and ethical considerations of advanced AI.)
- Chollet, François. *Artificial Intelligence: A Guide for Thinking Humans*. Penguin, 2019.
- Crevier, D. *AI: The tumultuous history of the search for artificial intelligence*. BasicBooks, 1993. (A good starting point for exploring the history of AI.)
- Pedro Domingos. *The Master Algorithm.* 2015. (A guide to the various approaches and techniques in machine learning.)
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. (A detailed exploration of deep learning techniques and their applications.)
- O'Neil, Cathy. *Weapons of Math Destruction*. Penguin Books, 2017. (A discussion of many of the ethical concerns that AI has given rise to.
- von Neumann, J. *The computer and the brain*. Yale university press, 1958. (Early exploration of the relationship between computers and the human brain.)

- Russell, S. J., & Norvig, P. *Artificial intelligence: A modern approach*. Pearson Education Limited. 2021. (A comprehensive textbook on AI.)
- Shannon, C. E. A mathematical theory of communication. *Bell system technical journal*, *27*(3), 379-423, 1948. (Foundational work in information theory.)
- Turing, Alan. "Computing Machinery and Intelligence." Mind, 1950. (A foundational work in AI, introducing the concept of the Turing Test.)

This Module was made with the assistance of OpenAI's Chatgpt, Microsoft's Copilot and Google's Gemini.