

Mr. Smith provided the following summary of his talk, which was also a demonstration using live websites, but did not provide the full text.

How did we get here and when will they join us? (Summary)

Presented to the SPHEX Club of Lynchburg on 1/11/24 by Stephen C. Smith

Key Terms: Artificial Intelligence; OpenAI, ChatGPT4, Technology, Natural Language Processing, Large Language Models

Recent developments in Artificial Intelligence, or AI, are in turn promising and concerning. While we are not at the level of Artificial General Intelligence, or AGI, that scientists are referring to when they talk of the risk of extinction of the human race, developments in the field are progressing rapidly. AGI, while not scientifically defined, is understood to be when an autonomous computer model can do any task as well as a typical human. While AI is currently able to do some tasks more effectively than humans, such as complex mathematical computations, there are many more tasks that AI simply can't do, or can't do reliably, at this time.

AI has been around for over 50 years, and it is embedded in much of our daily lives. Its development is marked by numerous significant events, one of the most recent being OpenAI's release of ChatGPT3 in 2022. Since then, AI has been in the news daily, as have people expressing concerns about the risks of AI development. ChatGPT is a Natural Language Processing (NLP) Large Language Model (LLM) Generative Pretrained Transformer (GPT). Current AI models like ChatGPT are not only able to, in some cases, pass the Turing Test proposed by Alan Turing in 1950, a test in which a machine communicates in natural language so effectively that it can't be distinguished from a human, they exceed human performance on numerous tasks.

The question of how did we get to this point with AI is best answered with an overview of the AI development timeline and major areas. In summary, AI started in 1950 with classical classification systems. In the 1980s the focus was on expert systems, narrow and labor intensive database decision tree type programs. Machine learning started in the 1990s, and symbolic machine learning, rule based systems of logic, were preferred over neural networks until AlexNet's development and success in 2012.

There were three primary barriers to the development of systems like ChatGPT4: limited computer resources, limited training data, and required advances in programming algorithms. These barriers were overcome by the development of Graphics Processing Units (GPUs), the Internet and its vast amount of data, and several programming innovations. Current AI models and training datasets are exponentially larger than previous ones, and the scaling of these models has resulting in significant improvements in the usefulness and reliability of their performance.

NLP LLM GPTs are trained on vast datasets, some of which may be larger than all the information on the Internet. They break language into tokens, numbers that represent words, punctuation, or parts of words like prefixes or suffixes. They can do this in any language to which they are exposed. They examine the relationships between and among tokens in context from left to right and right to left and at many levels and in ways humans can't. Their training is unsupervised—they simply mask a word in context, predict it, unmask the word to check the answer, then adjust their program to accommodate for any errors. Some models train this way trillions of times a second, 24 hours a day, 7 days a week, for months on end. The models then undergo Supervised Fine-Tuning and Reinforcement Learning with Human Feedback. Then, when exposed to a prompt, the model will take all it has been trained on, and the prompt, and predict the most likely next token, add that token to its prompt, predict the next, and continue in that way. This is why fine-tuning the model and engineering the prompt are essential to decreasing hallucinations.

There are a great many concerns about the development of AI, and categories of the risks are summarized in a paper entitled “An Overview of Catastrophic AI Risks” written by members of the Center for Artificial Intelligence Safety. Misuse of AI by bad actors is an obvious concern, but misalignment of the goals and methods of an autonomous AI system and humans is even more concerning. There are also many exciting advances in the field of AI, advances that have the potential to benefit humanity. In summary, AI is not as good as you might hope, nor as bad as you might fear...yet.