**AI Code Generation Models: The Big List**

<https://aibusiness.com/nlp/ai-code-generation-models-the-big-list>

ChatGPT wowed the public with its ability to generate text, with marketers and copywriters able to use it to aid their work. Emerging around the same time is another use for generative AI that could revolutionize software development: text-to-code.

Instead of painstakingly writing code line-by-line, developers may soon be able to simply describe what they want the program to do in natural language. AI systems like ChatGPT, Copilot and Ghostwriter will then take care of producing the required code.

*AI Business* explores the workings and abilities of these transformative AI systems that can code entire programs from the ground up based solely on text prompts.

**Text to code**

**What are text-to-code AI models?**

Text-to-code AI models use machine learning to generate snippets of code or entire functions. These models are trained on vast amounts of public code and are designed to aid human developers.

Text-to-code AI models take natural language inputs – plain English – and can turn it into code.

A diagram of a machine learning model

Description automatically generated

**Examples of text-to-code AI models and applications**

**StarCoder**

**Creators:**

* [ServiceNow](https://www.servicenow.com/uk/now-platform/now-intelligence.html) - Santa Clara-based enterprise workflows company
* [Hugging Face](https://huggingface.co/) - Machine learning tools developer and home to one of the internet’s largest libraries of natural language processing AI models

**Related:**[12 Language Models You Need to Know](https://aibusiness.com/nlp/12-language-models-you-need-to-know)

**First published:**May 2023

[StarCoder](https://huggingface.co/bigcode/starcoder) is a 15 billion-parameter AI model designed to generate code for the open-scientific AI research community.

StarCoder was trained on licensed data from GitHub spanning over 80 programming languages, and fine-tuning it on 35 billion Python tokens.

The resulting model outperforms Google’s PaLM 1 and Meta’s LLaMA at popular benchmarks despite its small size.

**Access StarCoder:** <https://huggingface.co/bigcode/starcoder>

**Read more about Starcoder on AI Business:**<https://aibusiness.com/nlp/hugging-face-service-now-launch-coding-llm-starcoder>

**Codex**

**Creator:** [OpenAI](https://openai.com/) - New York-based AI research lab backed by Microsoft

**First published:**August 2021

[Codex](https://openai.com/blog/openai-codex) is a code generation model that powers GitHub Copilot (see below).

Proficient in more than a dozen programming languages, Codex can interpret simple commands in natural language and execute them.

**More on OpenAI’s code generation capabilities:**<https://platform.openai.com/docs/guides/code>

**Read more about Codex on AI Business:**<https://aibusiness.com/ml/openai-upgrades-codex-machine-learning-assistant-says-it-can-turn-natural-language-into-code>

**Copilot**

**Creators:**

* [GitHub](https://github.com/) - Microsoft-owned code hosting platform
* [Microsoft](https://www.microsoft.com/en-us/ai) - Software giant behind Windows, 365
* [OpenAI](https://openai.com/) - New York-based AI research lab backed by Microsoft

**Related:**[The Essential List: AI Text-Generation Models and Apps](https://aibusiness.com/nlp/ai-text-generation-models-and-apps-everything-you-need-to-know)

**First published:**October 2021

**Current version:** [Copilot X](https://github.com/features/preview/copilot-x)

[Copilot](https://github.com/features/copilot) is a generative AI coding tool. It can take text inputs and queries and turn them into coding suggestions across dozens of languages, including Python, JavaScript, TypeScript, Ruby and Go.

The latest iteration, [unveiled in March 2023](https://aibusiness.com/verticals/github-supercharges-copilot-with-gpt-4-new-features), supports voice recognition, GPT-4-powered tags in pull request descriptions and a ChatGPT-style interface where devs can ask questions about business documentation.

**Read more about Copilot X on AI Business:**<https://aibusiness.com/verticals/github-supercharges-copilot-with-gpt-4-new-features>

**Code Interpreter**

**Creator:** [OpenAI](https://openai.com/) - New York-based AI research lab backed by Microsoft

**First published:** July 2023

It is the only plugin on this list – Code Interpreter is an add-on for ChatGPT, enabling users to use ChatGPT to generate and execute code. Without it, ChatGPT can only generate code snippets.

Code Interpreter is only available to subscribers of OpenAI’s premium offering, ChatGPT Plus. Users can write and execute Python code as well as upload a file and ask ChatGPT to analyze data, create charts, edit files and perform math.

Here’s an example of Code Interpreter on ChatGPT Plus creating code for a bot to monitor online prices of Pokémon cards.

**Related:**[AI Image-Generation Models and Tools: The Ultimate List](https://aibusiness.com/nlp/the-essential-list-ai-image-generation-models-and-tools)

A screenshot of a computer program

Description automatically generated

**Read more about Code Interpreter on AI Business:**<https://aibusiness.com/nlp/openai-s-code-interpreter-lets-chatgpt-play-data-scientist>

**CodeT5**

**Creator:** Salesforce - Enterprise cloud giant

**First published:**May 2023

[CodeT5](https://github.com/salesforce/CodeT5) is a [large language model](https://aibusiness.com/nlp/what-is-a-large-language-model-) for code understanding and generation tasks.

CodeT5 is a pre-trained encoder-decoder model built to perform tasks including code defect detection and clone detection, as well as generation tasks.

**Read the CodeT5 research paper:** <https://arxiv.org/pdf/2305.07922.pdf>

**Access the CodeT5 code:**<https://github.com/salesforce/CodeT5>

**Polycoder**

**Creators:**Researchers from Carnegie Mellon University - the full list of authors are in the [paper.](https://arxiv.org/pdf/2202.13169.pdf)

**First published:** May 2022

Built using OpenAI’s [GPT-2](https://github.com/openai/gpt-2) language model, [Polycoder](https://github.com/VHellendoorn/Code-LMs" \t "_blank) was trained on a dataset of 249GB of code spanning 12 programming languages.

It was designed as an open source alternative to OpenAI’s Codex (see above). While not as powerful as some of the code generation models on this list, Polycoder surpasses Codex at writing code in the programming language C.

**Read the Polycoder paper:** <https://arxiv.org/pdf/2202.13169.pdf>

**Access the Polycoder code:** <https://github.com/vhellendoorn/code-lms#models>

**Replit Ghostwriter**

**Creator:** [Replit](https://replit.com/" \t "_blank) - San Francisco-based startup and software development platform

**First published:** October 2022

Replit’s answer to GitHub Copilot (see above), [Ghostwriter](https://replit.com/ai) is an AI-powered programming tool to aid developers building software.

Ghostwriter can complete code, providing users with suggestions, as well as explain to the user code in plain English. It can also re-write and generate code based on natural language prompts.

In February 2023, Replit launched Ghostwriter Chat, [adding conversational AI capabilities](https://blog.replit.com/gw-chat-launch)to the mix.

**Try Replit Ghostwriter:** <https://replit.com/signup>

**Tabine**

[Tabine](https://www.tabnine.com/) uses deep learning to aid code completions. Tabine supports over 50 programming languages and has a free one-user option.

For businesses, Tabine offers an enterprise offering, with the chipmaking giant Nvidia, Elon Musk’s rocket company SpaceX and sportswear maker Nike all using the platform.

**Try Tabine:** <https://www.tabnine.com/install>