

# ChatGPT & Python API

# Content

- + What is ChatGPT?
- + ChatGPT Registration
- + Integrate ChatGPT into Python application

# ChatGPT

## What is ChatGPT?

ChatGPT is basically a long-form question and answering AI which answers your questions conversationally. This ChatGPT tool is developed by OpenAI and launched in November 2022.

ChatGPT is basically a Chatbot and it is based on GPT 3.5. The word '**GPT**' means **Generative Pre-Trained Transformer**. Let's break down the GPT to understand it properly. '**Generative**' means the capability to generate something, '**Pre-Trained**' means who is already trained and not required any other training and '**Transformer**' means such a learning model that understands the text and is transformed as per your need. So we can say that the GPT on which the ChatGPT is built, is such a model of artificial intelligence that is already trained and generates a lot of text based on some hints.



# ChatGPT

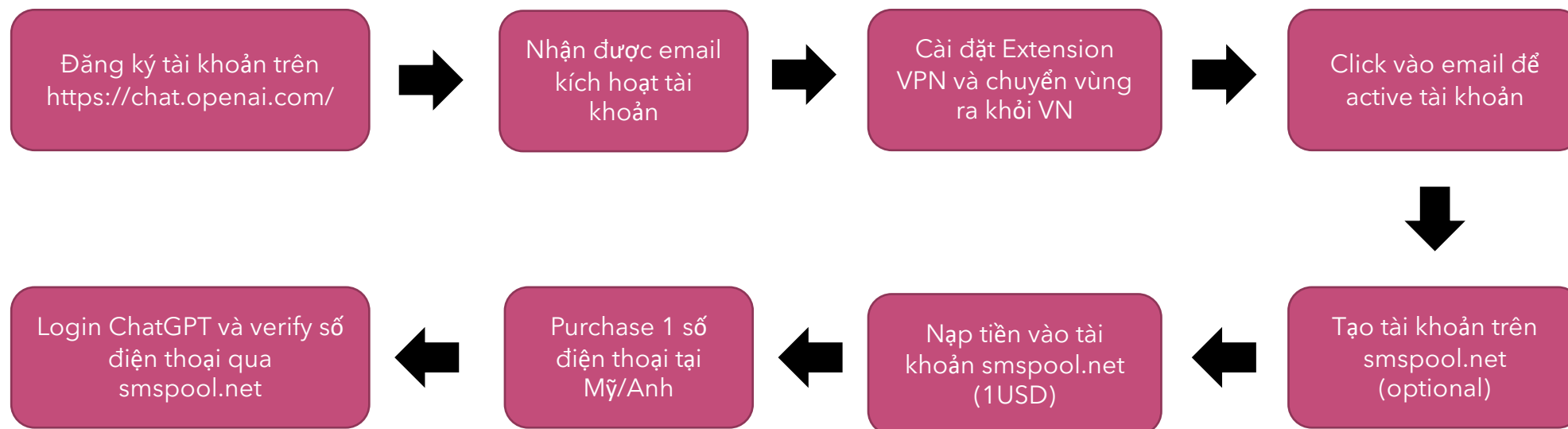
## How Does ChatGPT Work?

ChatGPT takes your questions or requests and quickly answers them. This tool answered your questions using its database. The model is trained using text databases from the Internet. It contains 570 GB of data sourced from books, web texts, Wikipedia, articles, and other pieces of writing on the Internet. To be even more precise, 300 billion words have been fed into the system.

You also get the option to tell whether you are satisfied with the answer given by it or not. Whatever answer you give, it also keeps updating its data continuously.



# ChatGPT Registration



# Models

LATEST MODEL	DESCRIPTION	MAX REQUEST	TRAINING DATA
text-davinci-003	Most capable GPT-3 model. Can do any task the other models can do, often with higher quality, longer output and better instruction-following. Also supports <a href="#">inserting</a> completions within text.	4,000 tokens	Up to Jun 2021
text-curie-001	Very capable, but faster and lower cost than Davinci.	2,048 tokens	Up to Oct 2019
text-babbage-001	Capable of straightforward tasks, very fast, and lower cost.	2,048 tokens	Up to Oct 2019
text-ada-001	Capable of very simple tasks, usually the fastest model in the GPT-3 series, and lowest cost.	2,048 tokens	Up to Oct 2019

# Models

LATEST MODEL	DESCRIPTION	MAX REQUEST	TRAINING DATA
code-davinci-002	Most capable Codex model. Particularly good at translating natural language to code. In addition to completing code, also supports <b>inserting</b> completions within code.	8,000 tokens	Up to Jun 2021
code-cushman-001	Almost as capable as Davinci Codex, but slightly faster. This speed advantage may make it preferable for real-time applications.	Up to 2,048 tokens	



# Token

The GPT family of models process text using **tokens**, which are common sequences of characters found in text. The models understand the statistical relationships between these tokens, and excel at producing the next token in a sequence of tokens.

---

A helpful rule of thumb is that one token generally corresponds to ~4 characters of text for common English text. This translates to roughly  $\frac{3}{4}$  of a word (so 100 tokens  $\approx$  75 words).



# Token

GPT-3 Codex

Hello how are you|

Clear

Show example

**Tokens**

4

**Characters**

17

Hello how are you

# Rate limits

	TEXT & EMBEDDING	CODEX
Free trial users	•20 RPM •150,000 TPM	•20 RPM •40,000 TPM
Pay-as-you-go users (first 48 hours)	•60 RPM •250,000 TPM*	•20 RPM •40,000 TPM
Pay-as-you-go users (after 48 hours)	•3,000 RPM •250,000 TPM*	•20 RPM •40,000 TPM

# Rate limits

The **TPM** (tokens per minute) unit is different depending on the model:

TYPE	1 TPM EQUALS
davinci	1 token per minute
curie	25 tokens per minute
babbage	100 tokens per minute
ada	200 tokens per minute

# Handson

- + Generate OpenAPI Token
- + Check OpenAPI balance & price
- + Integrate ChatGPT into Python application
- + <https://platform.openai.com/docs/introduction/overview>