|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Architecture** | **Training data** | **Release date** | **Training cost** |
| **GPT-1** | 12-level, 12-headed Transformer decoder (no encoder), followed by linear-SoftMax. | Book-Corpus 4.5 GB of text, from 7000 unpublished books of various genres. | June 11, 2018 | 30 days on 8 P600 GPUs, or 1 petaflop/s-day. |
| **GPT-2** | GPT-1, but with modified normalization | Web Text: 40 GB of text, 8 million documents, from 45 million webpages upvoted on Reddit | February 14, 2019 (initial/limited version) and November 5, 2019 (full version) | "Tens of petaflop/s-day”, or 1.5e21 FLOP. |
| **GPT-3** | GPT-2, but with modification to allow larger scaling | 499 billion tokens consisting of Common Crawl (570 GB), Web Text, English Wikipedia, and two books’ corpora (Books1 and Books2). | May 28, 2020 | 3640 petaflop/s-day (Table D.1), or 3.1e23 FLOP. |
| **­­­­­**  **GPT-3.5** | Undisclosed | Undisclosed | March 15, 2022 | Undisclosed |
| **GPT-4** | Also trained with both text prediction and RLHF; accepts both text and images as input. Further details are not public. | Undisclosed | March 14, 2023 | Undisclosed. Estimated 2.1e25 FLOP. |