**Notes:**

**How LLMs work:**

* An LLM is an instance of a foundation model
* Foundation models are trained on unlabelled, unsupervised data, so the model learns from patterns in the data
* LLMs are instances of foundation models applied in text-based circumstances
* Trained on approximately 1.78\*10¹⁴ words worth of data (1 Petabyte)
* The more parameters a model has, the more complex it can be
* GPT-3 was trained on 43TB of data and uses 175B parameters
* LLMs are made up of data, architecture and training
* The data is the large amount of text training data used
* The architecture is the neural network used, for GPT this is a transformer
* The LLM is then trained using the data
* The models can be fine-tuned on specific, smaller datasets

**Environmental impact of AI:**

* AI, crypto and data storage represented almost 2% of global energy demand in 2022 and will double by 2026
* All digital data is stored in physical data centres that require a lot of energy due to constant cooling necessary
* The internet emits more GHG emissions than aeroplanes
* Often, data centres are water-cooled, and use approximately 5 million gallons of water per day
* A lot of e-waste comes from data centres as well, due to constant maintenance of the hardware
* Storing 100GB of cloud data per year would produce 200kg of CO₂
* A prompt on Chat-GPT requires 10 times more energy than a google search
* Training Chat-GPT uses the same amount of energy as the annual consumption of 130 US homes
* AI training and AI usage use massive amounts of energy
* Companies are buying millions of GPUs, so tonnes of natural minerals are being mined, and humans are being exploited to work these mining jobs
* If data centres switched to renewable energy sources, then there would be a drastic improvement on the amount of GHG emissions and the environmental impacts
* Transparent AI could also help, as it can tell users how much energy each model uses

**Table:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Parameters** | **Training CO₂ (kg)** | **Energy Use** | **Performance** |
| GPT-3 | 175B | 500,000 | Very High | Excellent |
| LLaMA-7B | 7B | 15,000 | Medium | Good |
| DistilBERT | 66M | 200 | Low | Moderate |
| Falcon-40B | 40B | 375,000 | High | Excellent |
| BERT-Base | 110M | 635 | Low | Good |
| BLOOM | 176B | 25,000 | Very High | Excellent |
| XLM-RoBERTa-Base | 279M | 635 | Medium | Good |
| DeepSeek-R1-0528 | 685B | 500,000 | High | Excellent |
| Mixtral-8x22B | 141B | 200,000 | High | Excellent |

A graph of a graph with numbers and text

AI-generated content may be incorrect.