Large Language Models (LLMs) – Technical Overview

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Developer | Parameters | Context Window | Capabilities | License | Notable Features |
| GPT-4o | OpenAI | ~1T (est.) | 128K | Multimodal (text, image, audio), advanced reasoning | Proprietary | Real-time voice interaction, integrated into ChatGPT |
| Claude 3.7 Sonnet | Anthropic | Undisclosed | 200K–1M | Multimodal, advanced reasoning | Proprietary | Enhanced coding, chart interpretation, image text extraction |
| Gemini 2.5 Pro | Google DeepMind | Undisclosed | >1M | Multimodal, Mixture-of-Experts (MoE) | Proprietary | Integrated with Google services, multimodal generation |
| Llama 3.3 | Meta AI | 70B | 128K | Multilingual, text-only | Open Source | Cost-efficient, strong performance |
| Mistral Large 2 | Mistral AI | 123B | 128K | Multilingual, code generation | Commercial | High code task performance, cloud availability |
| DeepSeek-V3 | DeepSeek | 671B | Undisclosed | Multilingual, reasoning | MIT | 14.8T tokens, open-sourced |
| Qwen2.5-72B | Alibaba Cloud | 72B | Undisclosed | Multilingual, MoE | Qwen License | MoE variants for efficiency |
| Grok 3 | xAI | 314B (MoE) | 128K | Multimodal, AI agents | Proprietary | Supports vision tasks, advanced reasoning |
| Falcon 180B | Tech Innovation Institute | 180B | Undisclosed | Multilingual, reasoning | Apache 2.0 | Open-source, high benchmark performance |
| Phi-3 | Microsoft | 14B | Undisclosed | Small language model | MIT | 4.8T tokens, optimized for efficiency |
| DBRX | Databricks & Mosaic ML | 136B | Undisclosed | Multilingual | Databricks Open Model License | 12T tokens, cost-effective training |
| Gemma | Google DeepMind | 7B | Undisclosed | Text-only | Gemma Terms of Use | Lightweight model for tasks |
| Ernie 4.0 | Baidu | ~10T (rumored) | Undisclosed | Multilingual, multimodal | Proprietary | Used by 45M+ users |
| Nova Pro | Rubik's AI | Undisclosed | Undisclosed | Multimodal | Proprietary | Includes Nova-Instant & Nova-Air |
| Fugaku-LLM | Fujitsu & Tokyo Tech | 13B | Undisclosed | Multilingual | Undisclosed | Trained on CPUs, energy-efficient |