

DATASET SCALING: GPT TO GENESIS MISSION 2026



GPT-1: Books 994M tokens • 4GB



GPT-2: Web 40B tokens • 160GB



GPT-3: Human data 499B tokens • 2TB

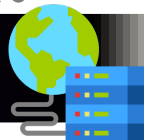


GPT-4: Human data 13T tokens • 52TB

GPT-5: Human + synthetic data 70T tokens • 280TB



Genesis Mission: Physical world data 1Qi tokens • 3.6EB



40×

12×

26×

5×

14,286×

Log scale, estimates only. Some images by Flaticon.com. See also What's in my AI? <https://life architect.ai/whats-in-my-ai/> & What's in GPT-5? <https://life architect.ai/whats-in-gpt-5/> Alan D. Thompson, December 2025, <https://life architect.ai/>



LifeArchitect.ai/genesis