

Code Generation Performance Across Models (HumanEval, 16,400 generations per model)

