

EVIDENCE SUGGESTS THAT COMPUTE PROVIDERS FOR GEN AI OPERATE IN AN INCREASINGLY COMPETITIVE ENVIRONMENT



MANY PROVIDERS OF COMPUTE FOR GEN AI

Evidence suggests that there are **> 100 providers of compute for Gen AI**, including new AI-specialised providers (e.g., CoreWeave, Cerebras, Lambda) and on-premises providers¹



BUYERS MAKE USE OF MULTIPLE OPTIONS

Buyers often use various options for accessing compute for Gen AI workloads, including **multiple cloud providers or a combination of on-prem and cloud services**



SEVERAL EMERGING PLAYERS ARE GROWING FAST

Several emerging AI-specialised players and on-premises providers are experiencing **triple-digit revenue growth**



INDICATIVE EVIDENCE SUGGESTS DECLINING PRICES

Available evidence indicates that the inflation-adjusted **price of GPU processing power has decreased** over the recent years



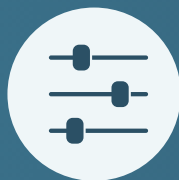
SUBSTANTIAL FURTHER INVESTMENT IS EXPECTED

Total worldwide **investment in data centres** is forecasted to reach **USD 1 trillion annually by 2029**, half of which is expected to support AI workload²



- Competitive conditions could vary between users (e.g. developers, deployers) and types of Gen AI workload (pre-training, fine-tuning, and inference). For example, pre-training requires more intensive workloads than inference.
- Further research would be required to capture any differences in competitive conditions.

VARIOUS EXPECTED DEVELOPMENTS MAY FURTHER STRENGTHEN COMPETITION



SHIFT OF DEMAND TOWARDS **LESS COMPUTATIONALLY INTENSIVE WORKLOADS** (FINE-TUNING AND INFERENCE)



ONGOING INNOVATION IN COMPUTE USAGE TOWARDS **SMALLER, MORE EFFICIENT MODELS, AND ON-DEVICE INFERENCE**

POLICYMAKERS SHOULD CAREFULLY ASSESS COMPETITIVE CONDITIONS BEFORE CONSIDERING ANY POTENTIAL INTERVENTION

The research underlying this report was carried out between April and June 2025. Given the fast-moving nature of this market, subsequent developments may not be fully reflected.