Impact of GenAI on company finances 1/19/24

Associate, GenAI, IT Advisory, PwC Latvia
Gunda Karnīte

Generative artificial intelligence (GenAI) has become an essential business tool that helps companies optimise their processes, improve efficiencies and cut costs. However, to better understand GenAI's impact on finances, it's important to consider the cost of this tool from different aspects.

Every time a GenAI model receives a prompt, it operates large language models (LLMs) using graphic processing units for computing. The processing of data inputs is known as natural language processing, and the data is measured in tokens. Each token equals about four English characters and with 1,000 tokens you can process about 750 words. To generate text using a GPT-4 model, the price of 1,000 tokens for text input is about USD 0.03, while 1,000 tokens for text output cost about USD 0.06. For example, if a company employs 1,000 workers and each of them enters ten prompts a day, 300 words each, generating an output of 100 words, the cost of text generation would be USD 198 a day – about USD 45,000 a year.

Most of the cost is incurred in fine-tuning the GenAI model for a particular task or domain to ensure the model produces the desired result. For example, a GenAI model trained to carry out a financial analysis from financial statements may need fine-tuning due to changes in tax treatment. The process of fine-tuning includes adjusting the model parameters and using a new dataset to make the outputs more accurate in the context of the country's taxation. For example, OpenAI uses a formula to compute the total cost of fine-tuning.

The total fine-tuning price equals the basic cost of 1,000 tokens times the number of tokens in the input file times the number of iterations required to fine-tune the model. The AI terminology defines iterations as epochs. For example, if you require ten iterations with 50,000 tokens each and 1,000 tokens cost USD 0.008 (the training fee for the gpt-3.5-turbo model), the total fine-tuning cost would be:

10 * 50,000 / 1,000 * 0.008 = USD 400

Building the LLM infrastructure requires significant investment, so most companies opt for cloud services to operate their GenAI models. Cloud cost breaks down into several items: time-of-use fee, storage fee, general-purpose instance fee, optimised compute instance fee, annual subscription fee. The largest public cloud service providers are Amazon Web Services, Azure, Google Cloud Platform and Oracle, each charging a different amount for these items.

When it comes to choosing cloud architecture for your GenAI model, the main consideration is whether it will be operated on public cloud, private cloud or multi-cloud solutions. For example, medical institutions having to protect their patient data would favour private cloud storage, which may initially look more advantageous than public cloud but in the long-term may lead to extra costs to maintain and update the models.

The total cost of using and fine-tuning GenAI may vary according to several factors, including services used and model specifications. It's important to note data quality as a key factor in building GenAI models because without high-quality data a GenAI model may need to be fine-tuned more than once, increasing the cost and the consumption of resources. To make calculations more accurate, you should also consider staff training on GenAI usage, data security measures and ethics. Of course, consulting industry experts on

how to use GenAI efficiently will help you maximise your benefits.

To find out more about GenAI technology, we invite you to attend a free webinar ‘Using GenAI in practice’ run by PwC’s Academy on 23 May 2024.