Chapter 11: Generative AI in GCP

In the preceding two chapters we have demonstrated AWS and Azure capabilities within the realm of generative AI; we now turn our attention to Google offering in that space. What differentiates the company from its main competitors in the space – Amazon and Microsoft – is that it is currently the only one offering a comprehensive suite of commercial solutions covering every aspect of the AI-focused suite of tools. From dedicated AI chips, through computing power, to development tools and applications – Google has a complete AI stack that serves as a foundation for its offering. Their proprietary Tensor Processing Units (TPU) are designed specifically for machine learning tasks, optimized for use with TensorFlow, and are boasting faster and more efficient performance for AI workloads; while GPU (and CPU) units have not been dethroned across the industry, over the last few years TPU have emerged as formidable rivals.

The Cambrian explosion of generative AI started in the second half of 2022 and the initial impression was that Google had to play catch-up to Microsoft / OpenAI (especially in the NLP space, due to chatGPT parabolic rise). Upon closer examination, it seems like Google strategy has been more of a comprehensive enterprise-grade approach. We begin by discussing the educational offer on GCP: if you, dear reader, have reached this point in the book, chances are you are already familiar with the basics; nevertheless, the series of mini courses offered on the platform can be useful complement to your knowledge.

Having reviewed the educational offer, we will examine the capabilities of Generative AI Studio. The application is built on the Vertex AI platform and allows developers to create their own generative AI apps in text and vision domains. More advanced users can make us of the Model Garden, which gives access to a collection of pretrained models, for advanced exploration and interaction.

If you prefer to hear the summary of Google Generative AI offerings from the company themselves, feel free to check out the introductory videos: <https://www.youtube.com/watch?v=YCZ6nwGnL4o>. Once you have done that, you can come back here and we will embark on the next stage of our journey.

# Learning path

The educational provision on GCP are designed for a wide range of individuals:

The

The educational provisions on the GCP are designed to cater to a wide range of individuals, from novice enthusiasts eager to expand their understanding to seasoned professionals seeking to refine their expertise. These succinct yet informative courses provide an accessible and engaging learning experience, allowing learners to delve into specific topics within the realm of generative AI. By participating in these educational offerings, you can gain a deeper insight into the principles, techniques, and applications of generative AI, while honing your practical skills in implementing and leveraging these technologies effectively.

By availing yourself of the educational resources available on the GCP, you have the opportunity to enhance your proficiency in generative AI and stay abreast of the latest developments in this dynamic field. The knowledge and competencies gained through these courses can empower you to tackle real-world challenges, create innovative solutions, and unlock the full potential of generative AI within your projects and endeavors

**A picture containing text, screenshot, diagram, design

Description automatically generated**

Figure 11.1: Landing page for the Generative AI track from GCP

# Generative AI Studio

Asd

# Running models on GCP

Asd