Professional Development

What GitHub Copilot is:

GitHub Copilot is an artificial intelligence (AI) powered code completion tool that is designed to help developers write code more efficiently. It is a collaboration between OpenAI and GitHub, and uses machine learning models trained on a large corpus of open source code to generate suggestions for code as developers write.

GitHub Copilot works by analyzing the code you are working on and suggesting code snippets or even entire functions that it believes will be useful. It can be integrated with popular code editors such as Visual Studio Code, and supports a variety of programming languages including Python, JavaScript, TypeScript, Ruby, Go, and more.

While GitHub Copilot is still in beta and not perfect, it has the potential to significantly speed up the development process for developers by reducing the amount of time spent typing out repetitive code or searching for solutions to common problems. However, it's important to note that while Copilot can help generate code, it is ultimately up to the developer to ensure that the code is secure and performs as intended.

What algorithms it uses:

GitHub Copilot uses several machine learning algorithms to generate code suggestions, including deep learning models such as transformer-based language models and neural networks.

The core algorithm used by Copilot is called OpenAI's Codex, which is a large-scale, neural network-based language model that has been trained on a diverse set of code and natural language data. Codex is trained using a combination of supervised and unsupervised learning techniques, allowing it to generate high-quality code suggestions based on the context and intent of the developer's code.

In addition to Codex, Copilot also uses a variety of other machine learning techniques such as contextual embeddings, tree-based models, and graph-based models to enhance its code suggestion capabilities.

Overall, Copilot's algorithms are designed to learn from a vast corpus of code and natural language data, allowing it to suggest relevant and useful code snippets that can help developers be more productive and efficient.

Where the data that trained it came from:

The data used to train GitHub Copilot's machine learning models came from a variety of sources, including public repositories on GitHub, Stack Overflow, and other open-source code and documentation repositories.

GitHub's code search functionality allowed them to index and analyze code from millions of repositories, covering a wide range of programming languages and application domains. They also used natural language text from sources such as software documentation, technical articles, and other online resources.

The dataset used to train Copilot's core algorithm, Codex, was created by OpenAI and is one of the largest and most diverse code datasets ever assembled. The dataset consists of over 750 million lines of code and natural language text, covering a wide range of programming languages, libraries, frameworks, and application domains.

It's worth noting that the data used to train Copilot was carefully selected and filtered to ensure that it did not include any sensitive or proprietary information. Additionally, GitHub and OpenAI have put in place strict privacy and security measures to protect user data and prevent misuse of the platform.

How you used it in this project:

We did not use copilot as our coding skills are perfect and need no changing. That is a massive exaggeration but sadly we did not get to use it as we did not have the time to ask. After doing research it seems like it would have been a very useful tool.