Nimesh Tripathi | Product Management in AI and ML - Elvtr
White Paper on Github Copilot

1. Describe the Gen AI product. Include how it works technically if you are aware/ have access to it and What problems it is solving

The product I am choosing to review here is **Github Copilot**. It acts as a pair programmer where it can suggest generated code, by reading the surrounding code and tabs. Its most popular feature is the autocomplete feature, which displays predictive “gray-text”, which are basically, suggestions about the next lines of code. It gives access to a wide array of libraries which the programmer might not have used before, and it can also automate debugging, vulnerability testing and help with orchestration, with limited human supervision. It can generate entire codebases with the click of a button, and can also offer code translation(C++ to Java etc.)

It also leverages the comment box, where the user can prompt Gen AI to generate a few lines of code based functionality. Furthermore, It aids when a new user is onboarded to a project and wants to understand the code structure, during knowledge transfer. It makes code accessible to different levels of developers. It also allows you to toggle between a few generated solutions, to pick the right one, or to sample available solutions. It also allows the user to hover over a piece of code, and translate it into plain english, about what that code snippet is doing.

Github Copilot is trained over publicly available repositories of code, and its ability to predict accurate generated code, depends on the volume and diversity of data it has been trained on.

2. Who are the customers it is targeting?

Customer Segmentation
Github Copilot offers three variations, to different kinds of users:

1. Github Individual
2. Github Business
3. Github Enterprise

Github Copilot aims at engaging with the entire development community, engaging with individual users with their Github Individual account. It is priced at $10 per month and individuals, especially teachers or students can apply to use it for free.

It also leverages business and enterprise users respectively at $19 and $39 per month where they allow enterprises to have a shared repository, version control, pull request summaries and an integrated Teams platform.

3. What is its GTM strategy? How are they engaging and targeting customers?

Github is the world's leading open-source repository for code, with over 100 million users utilizing Github's code repository for collaboration, self-portfolios and because of this, they leveraged a header on their website mentioning their releases/updates, as it acted as a free marketing opportunity for Github to market the Copilot feature on their own website.

They also held multiple conferences displaying the code generation ability which automatically caught attention, since Github is the mecca for developers of all kinds.

Apart from engagement platforms, Github Copilot offers plugins to a variety of IDEs like JetBrains, VisualStudio etc. which helps compatibility with code development platforms, along with its own repositories. This a dual-pronged strategy at aiming to interact with developers who may or may not commit their projects on repositories.

4. Do you have recommendations or suggestions on how to improve the product/ experience/ GTM?

Although Github Pilot is touted as the next product to "revolutionize programming",it certainly fails to do so. Its suggestions can be random, out of context, and at a lot of times entirely problematic, as it's trained on public repositories and it can lead to copyright infringement, as it might display code which is proprietary to another user/organization. Copilot does not display the source where it suggested the code from, which has led to it not being adopted on enterprise level. My primary suggestion would be the ability to quote its sources, which might be a general LLM/Gen AI problem as well.

Another approach to this would be a plagiarism check which is inbuilt which verifies the copyright issues, which might save a lot of coders from "hot-water" in their organizations. However this might be computationally expensive to match against every piece of code, available, even if it's done through vector databases.