CHATGPT PROMPTS FOR VALIDATION OF COMMENTS:

**📄 Annotation Guidelines for Manual Validation**

* **Title:** *Validation of Security-Relevant Comments in Developer Discussions*
* **Used in:** Manual validation phase of the thesis: *Security Concerns in Generative AI Coding Assistants*
* **Purpose:** To assess whether each comment from Reddit, Stack Overflow, and Hacker News truly reflects developer concerns about security or privacy in GitHub Copilot or similar tools.

#### **Reviewer Instructions (Prompt)**

* You are acting as a **security expert** validating whether a given developer-authored comment reflects a **true security or privacy concern** regarding GitHub Copilot or other AI-assisted code generation tools.
* For each comment, assign one of the following labels:
  + ✅ **True Positive (TP):** The comment clearly raises a **security, privacy, trust, or risk-related** concern regarding the behavior, outputs, or design of Copilot.
  + ❌ **False Positive (FP):** The comment does **not reflect a security or privacy concern**, even if it contains a keyword. This includes off-topic remarks, usability complaints, productivity issues, or general opinions without a security framing.
* Please **ignore** comments that only mention bugs, performance, feature requests, or general praise/criticism unless they are framed explicitly as security risks.
* After the labelling, we need to agree on cohen’s kappa, the author and the second reviewer is you (ChatGPT).