**SURVEY STUDY – CODEBOOK**

**Security Concerns in Generative AI Coding Assistants**

| **THEME** | **CODE** | **DESCRIPTION** | **EXAMPLES** |
| --- | --- | --- | --- |
| **Exposure of Secrets and**  **Training Data Leakage** | **Human-in-the-Loop Code Verification** | Concerns about Copilot suggesting malicious or compromised code from training data. | *P178: "I wonder if there's any potential for Copilot to suggest malicious code because it's been trained on an open source projects containing intentionally malicious code."* |
| **Developer Dependency & Trust Issues** | Persistent prompts to use Copilot leading to unintentional use of proprietary code. | *P68: "How do you disable it again? You can't!Once you pressed "Start using Copilot" the copilot menu shows up on all repositories within GitHub.com for careless or uninformed users to leak proprietary code to their training set."* |
| **Security Risks and Vulnerability Propagation** | Copilot could emit sensitive credentials like private keys into generated code. | *P307: "Of course not. Copilot spits out private keys if they happen to be uploaded to the repo (yeah, I know, bad practice). They can't risk that."* |
| **Insecure Code Suggestions and Vulnerability**  **Patterns** | **AI Skepticism, Surveillance Anxiety, Distrust of Tooling** | Preference for offline or privacy-safe alternatives due to closed source restrictions. | *P135: "There is a real need for offline co-pilot alternative.Privacy concerns due to copilot makes it unusable for companies with closed source code.I'm hoping that eventually there will be a stable diffusion level solution in this field."* |
| **Debate Over AI “Stealing” Code** | Arguments around AI reusing structured or copyrighted works without attribution. | *P50: "... but I wonder what the perspective would look like for other structured works, like books? If an author is using a "copilot for writers" and the AI emits text verbatim to another work, then I would think it would be plagiarism. If the text emitted is similar, but not the same, then I would think it would be considered paraphrasing which still requires attribution."* |
| **Developer Trust Erosion**  **and Overdependence on**  **AI** | **Legal Loopholes & License Ineffectiveness** | Generated code may bypass original licensing obligations. | *P115: "This protects GitHub from infringing copyright by distributing code with whatever license, not consumers from infringing copyright if they as a result of Copilot get nonpermissivelylicensed code in their codebase."* |
| **Developer Attitudes Toward Ethical Use** | Lack of clear method for preventing one’s code from being used in training data. | *P280: "Not a fan of copilot. I have several small open source projects that are unlikely to be copied directly, but I do not want my work to be used to train a machine like copilot. How do I opt out? Or is this going to be like when Google mined free public translations to improve their paid autotranslating service without compensating the people who did the work they exploited?"* |
| **User Control and Privacy Safeguards** | Concerns over telemetry, data association, and privacy in Copilot usage. | *P33: "Login to GitHub Copilot using the device auth flow and authorized GitHub Copilot IntelliJ plugin with your GitHub Account in an external browser.> Read and agree to the GitHub Copilot additional telemetry terms.Can anybody comment on the privacy aspects of this? Is the telemetry reasonable? Why on earth do I need to login: presumably so that they can associate my coding with my account to structure the data they are gathering?"* |
| **Legal, Licensing, and At-**  **tribution Ambiguity** | **License Violation and Legal Exploitation Risks** | Copilot potentially reproducing GPL-licensed code without attribution. | *P3: "So they got complaints that copilot generated public code and then as aftermath they hashed all public snippets to bandaid their GPL offense?"* |
| **Legal Ambiguity and Copyright Compliance** | Copilot output conflicting with intended license restrictions. | *P32: "No. When you put your code out, you define the terms of use in your license, and you expect others to follow your license. If your license disallows it to be used in ML algorithm, it shouldn't be. Having your own license doesn't violate TOS.The ethics of copilot is clearly questionable."* |
| **Prompt Injection and**  **Training Data Poisoning** | **Trust, Uncertainty & Risk Awareness** | Risk of attackers using poisoned training data to cause harmful outputs. | *P285: "I wonder if CoPilot uses Github's private repositories to train itself, which would allow malicious users to somehow obtain code or designs that they otherwise would not be able to view."* |