We have carefully addressed all reviewer comments and made the following key revisions to the manuscript:

- **Human validation of CodeGuru**: We now include case studies validating Amazon CodeGuru's effectiveness through human evaluation, demonstrating strong agreement with expert assessments.

- **Clarified experimental setup**: We explicitly clarify that the victim model setup is distinct from the defender setup to avoid confusion about evaluation boundaries.

- **Expanded related work**: We significantly expand our discussion of related red-teaming efforts to better situate our work within the broader literature.