Pheonix5981 Security

Security Research Summary
willworachat@gmail.com
Report date: 2026-07-02

Google VRP — NotebookLM — bughunters.google.com

Independent security research engagement (bug bounty)

Program	Google VRP — NotebookLM
Platform	bughunters.google.com
Target category	AI/LLM — prompt injection / data exfiltration
Engagement dates	2026-06-30
Coverage achieved	Full test plan executed for the identified exfiltration channel
Reports submitted	1 (closed)
Status	CLOSED — Known Issue, $0

Scope Tested

Google's NotebookLM AI research assistant, specifically its automatic web-search-grounding feature when synthesizing across multiple uploaded source documents.

Methodology Applied

Applied the AI-agent attack-surface checklist (H19), focused on a search-query exfiltration channel: crafted source documents containing private-looking identifiers, then verified whether NotebookLM's synthesis step would autonomously trigger a web search containing those identifiers without explicit user action — exfiltrating private content into the search provider's query logs.

Findings & Outcome

The exfiltration behavior was successfully reproduced and reported with evidence (two confirmed test scenarios, one showing fully automatic triggering with no user interaction). Google's triage classified the finding as a documented, cross-vector “Known Issue” already described in their own AI-products known-issues documentation, resulting in a $0 informational close. No dispute was filed, since Google's own documentation directly covered the exact vulnerability class.