Snyk AI FAQ

General AI Product Information:

1. Q: What is the intended use of Snyk's AI?
   A: The intended use is fundamentally the same as Snyk's other core products; namely, identifying security issues in code and proposing fixes to those issues.
2. Q: What inputs will be provided from customers to the AI tool?
   A: Your software code.
3. Q: Does this include any Personal Information?
   A: Our product is designed to process software code, not Personal Information.
4. Q: Where does this data come from? Is it internal data, customer tenant data, etc.?
   A: Your software code. We provide an internal use code security tool and do not process your customer data.
5. Q: How are inputs provided to the tool? Do customers have complete control over the inputs, or does the tool automatically extract data from customers' system(s)?
   A: Customers have complete control over what data is made available to Snyk.
6. Q: What are the outputs that will be provided by the AI tool?
   A: The "outputs" of the tool are the proposed fixes to identified issues in your code.
7. Q: Will the tool be used to provide any customer-facing capabilities (e.g., a support chatbot, search assistance, or content generation)?
   A: Our product is an internal use tool used by your own personnel (your developers and product security team); we are not providing customer-facing capabilities.

AI Security:

8. Q: How often are security reviews and vulnerability scans conducted on the AI tool?
   A: At least annually.
9. Q: How is data stored and transmitted by the tool? What encryption mechanisms are used?
   A: Customer data is encrypted in transit using our approved TLS library. This is currently TLSv1.2 or greater (TLSv1.3 is preferred when supported by the client). This aspect of encryption is under continual review. Sensitive customer data is encrypted at rest using AES-256-CBC
10. Q: What data sources are used to train the model?
    A: Our model is training using permissively licensed open source repositories.

11. Q: Are AI models trained off customers' data? If so, are these models used by other customers?
    A: No, we do not use customer code to train our models.

12. Q: How do you ensure data separation from other customer's data?
    A: Data is logically separated.

13. Q: What methods do you have in place for monitoring the model's behavior?
    A: Our model is generally trained twice with new, labelled training data every week. Results of a newly trained model are compared against a validation set for coverage and accuracy. Accordingly, at present we track how the model performs and behaves on a weekly basis.

14. Q: How do you identify and prevent hallucinations?
    A: Any model can hallucinate, but we reduce the likelihood of hallucination through a combination of training data set curation and human validation. We run validation tests designed to validate that the proposed fix for an issue does not introduce new issues, that the fix is parsable, that it does not break the underlying code, and that it doesn't change the original syntax.

15. Q: Do you have an AI Governance Program for the use and development of AI? If yes, please describe.
    A: Yes, we maintain an AI Governance Program through our cross-functional AI Advisory Board and via the issuance of policies and procedures with respect to the responsible development of AI.