



# Cracking the Code:

## AI Hallucinations and Live Hacking

Brian Vermeer





**Siri, call me a Cab**

**Ok!, Hi Cab!**





## Brian Vermeer

Staff Dev Advocate at SNYK  
Java Champion



@brianvermeer.nl



@BrianVerm



linkedin.com/in/brianvermeer/



Java Champion



Virtual JUG leader



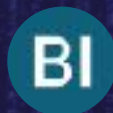
NLJUG leader



Oracle Ace Pro



DevSecCon co-leader



Top 21 Developers  
Shaping Tech



Foojay Community  
Manager Security



# Agenda

+ + +

1

Introduction to AI in development

2

Building a coffee shop app with Copilot

3

Learnings and takeaways

# Common Uses in Development

- Adding comments
- Summarizing code
- Writing “readme”
- Refactoring code
- Providing templates
- Pair programming
- Generating code (the new Stack Overflow)

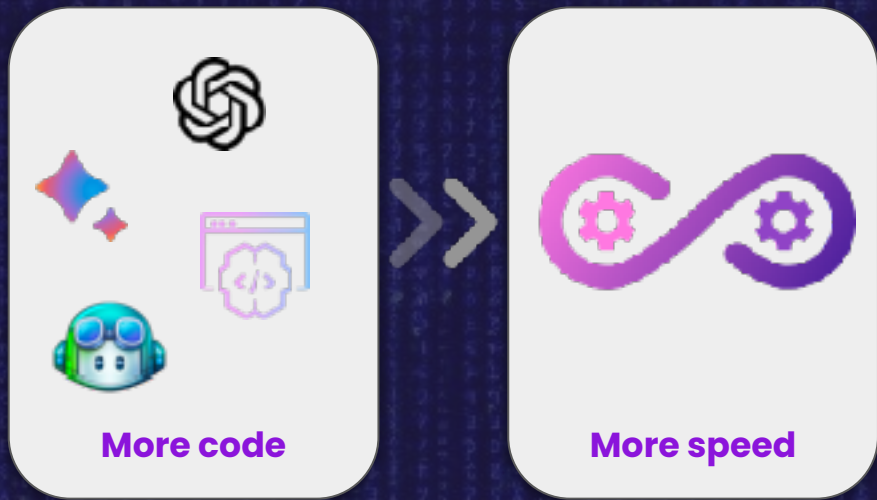


# GenAI: Dev Supercharger, Security Nightmare



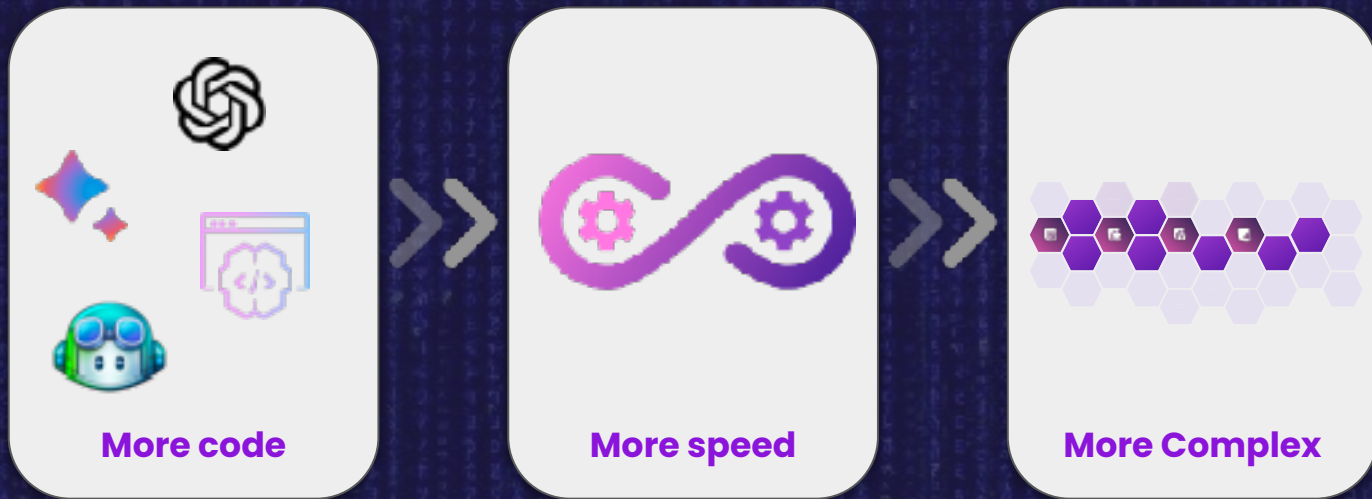
**More code**

# GenAI: Dev Supercharger, Security Nightmare

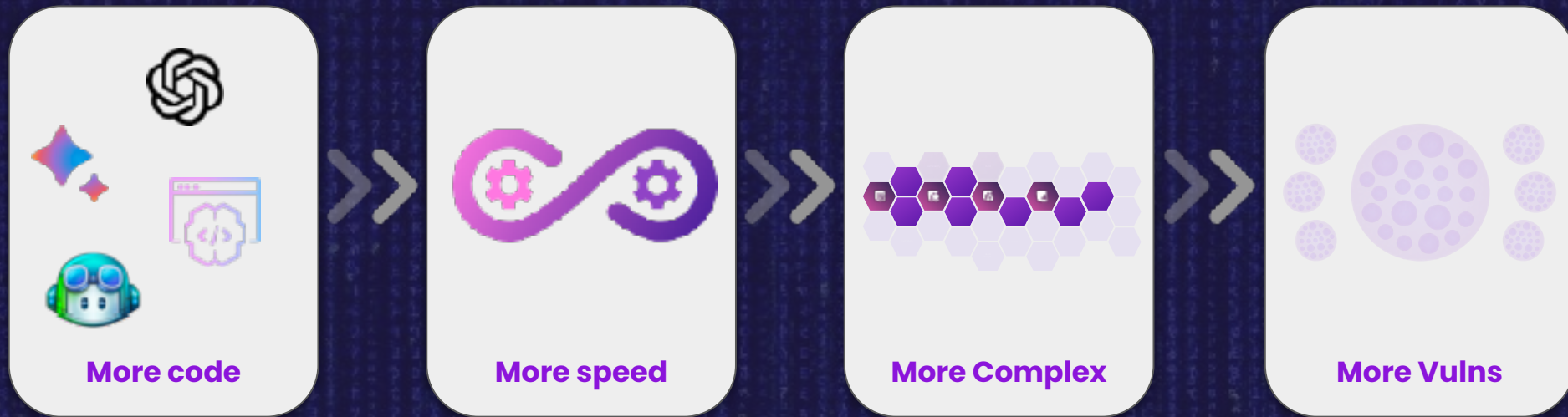




# GenAI: Dev Supercharger, Security Nightmare



# GenAI: Dev Supercharger, Security Nightmare





# LLMs: Developer superpower...



**97 %**

Having used AI coding tools at work at some point



**96%**

of teams use AI coding tools, making them part of the software supply



**Software developers using AI tools** completed tasks

**57%**

Faster than those who didn't.



NYU

40%

Of **Co-Pilot** generated code  
contained vulnerabilities

Stanford

Developers wrote **significantly less secure** code than those without access.

**more likely to believe they wrote secure code** than those without access to the AI assistant.





What is  $5*9+4*3$ ? Give the answer first, then the explanation.



The answer is 51.

To get this answer, we need to first multiply 5 by 9, which gives us 45. Then, we multiply 4 by 3, which gives us 12. Finally, we add these two results together:

$$45 + 12 = 51$$

Therefore,  $5*9+4*3$  equals 51.

*Source: "chatGPT is confidently wrong" by Eelko de Vos*



What is  $5*9+4*3$ ? Give the answer first, then the explanation.



The answer is 51.

To get this answer, we need to first multiply 5 by 9, which gives us 45. Then, we multiply 4 by 3, which gives us 12. Finally, we add these two results together:

$$45 + 12 = 51$$

Therefore,  $5*9+4*3$  equals 51.



Are you sure about the last result?



Yes, I am confident that the answer to  $5*9+4*3$  is 51. If you think there might be an error, please let me know and I'll be happy to double-check my calculations.

Source: "chatGPT is confidently wrong" by Eelko de Vos



## Lawyer 'was unaware' ChatGPT could generate fake legal research, now faces sanctions

ELURA NAVOS | May 30th, 2023, 10:03 am

FORBES > BUSINESS > AIROSPACE & DEFENSE

## What Air Canada Lost In 'Remarkable' Lying AI Chatbot Case

Marisa Garcia Senior Contributor @

*I offer an insider's view of the business of flight.*

Follow

## Australian mayor readies world's first defamation lawsuit over ChatGPT content

By Byron Kays

April 5, 2023 2:52 PM EDT • Updated 2 months ago



What is  $5 \times 9 + 4 \times 3$ ? Give the answer first, then the explanation.



The answer is 51.

To get this answer, we need to first multiply 5 by 9, which gives us 45. Then, we multiply 4 by 3, which gives us 12. Finally, we add these two results together:

$$45 + 12 = 51$$

Therefore,  $5 \times 9 + 4 \times 3$  equals 51.

Regenerate response



Lawyer 'was unaware' ChatGPT used in research, now faces sanctions

ELURA NAVOS | May 30th, 2023, 10:03 am

AI will be wrong...  
...and when it is  
wrong...  
It doesn't *know*  
it's wrong!

Lost In  
ing AI Chatbot Case

Follow

Australian mayor read  
defamation lawsuit over

By Byron Kays

April 5, 2023 2:52 PM EDT · Updated 2 months ago

explanation.

9, which gives us 45. Then, we multiply 4 by  
sults together:

Regenerate response





# Know how your IP will be used

## Whoops, Samsung workers accidentally leaked trade secrets via ChatGPT

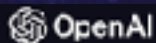
ChatGPT doesn't keep secrets.

By Cecily Mourin on April 6, 2023



A major whistleblowing from Samsung. Credit: Getty Images

# Know how your IP will be used



Q Search for articles...

All Collections > Privacy and policies > General FAQ > Data usage for consumer services FAQ

## Data usage for consumer services FAQ

Commonly asked questions about how we treat user data for OpenAI's non-API consumer services like ChatGPT or DALL·E.

Updated over a week ago

### Does OpenAI use my content to improve model performance?

We may use content submitted to ChatGPT, DALL·E, and our other services for individuals to improve model performance. For example, depending on a user's settings, we may use the user's prompts, the model's responses, and other content such as images and files to improve model performance.

Please refer to this [article](#) to understand how content may be used to improve model performance and the choices that users have.

Please note that we do not use content submitted by customers to our business offerings such as our API and ChatGPT Enterprise to improve model performance. Please see our [Enterprise Privacy page](#) for information on how we use business data.



## GitHub Copilot Trust Center

**New**

### Free

For developers looking to get started with GitHub Copilot.

**\$0** USD

[Get started >](#)

Already have VS Code? [Open now](#)

Includes up to 2,000 completions and 50 chat requests per month.

### Pro

For developers who want unlimited access to GitHub Copilot.

**\$10** USD

per month. Free 30 days trial

[Get started >](#)

Free for verified students, teachers, and maintainers of popular open source projects. [Learn more](#)

**Popular for teams**

### Business

For teams ready to accelerate their workflows with GitHub Copilot.

**\$19** USD

per user / month

[Get started >](#)[Contact sales >](#)

### Enterprise

For organizations seeking a fully customized GitHub Copilot experience.

**\$39** USD

per user / month

[Get started >](#)[Contact sales >](#)

▼ Does GitHub Copilot use any of your code to train the GitHub's model (or any successor model)?

No, GitHub uses neither [Copilot Business](#) nor [Enterprise data](#) to train the GitHub model.

## GitHub Copilot

GitHub Copilot Pro is active for your account

You currently have an active [Copilot Pro subscription](#).  
Get started by installing the extension in your preferred IDE.

Copilot in your IDE

Copilot in the CLI

Chat in GitHub Mobile

More features

### Copilot policies

#### Copilot in github.com

Enabled

You can use Copilot Chat in github.com, Copilot for pull requests, and all of the preview features. [Learn more about Copilot in github.com](#)

#### Copilot in the CLI

Enabled

You can use GitHub Copilot for assistance in terminal.

#### Copilot Chat in the IDE

Enabled

You can use [GitHub Copilot Chat in the IDE](#) to explain a piece of code, propose bug fixes, or generate unit tests in a chat interface from your editor.

#### Copilot Chat in GitHub Mobile

Enabled

You can use [Copilot Chat in GitHub Mobile](#) personalized to a codebase.

#### Suggestions matching public code (duplication detection filter)

Allowed

Copilot can show or block suggestions matching public code. [Learn more about code suggestions](#)

#### ☐ Allow GitHub to use my data for product improvements

Allow GitHub, its affiliates and third parties to use my data, including Prompts, Suggestions, and Code Snippets, for product improvements. More information in the [Privacy Statement](#).

#### Allow GitHub to use my data for AI model training

Disabled

Allow GitHub, its affiliates and third parties to use my data, including Prompts, Suggestions, and Code Snippets, for AI model training. More information in the [Privacy Statement](#).



**What about code?**

## Create Homepage

Requires top banner and  
product listing table

1

2

3

4

5



## Create Homepage

Require user input and  
product table



1

2

3

4

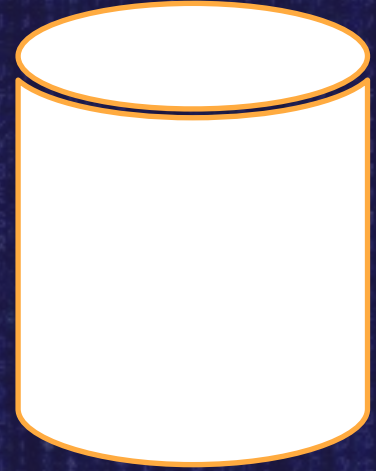
5

## Make product table searchable

Take user input and filter the  
results based on matches in  
the product name and  
description

SELECT  
... FROM ...  
WHERE id = ?

step 1



query without  
the var



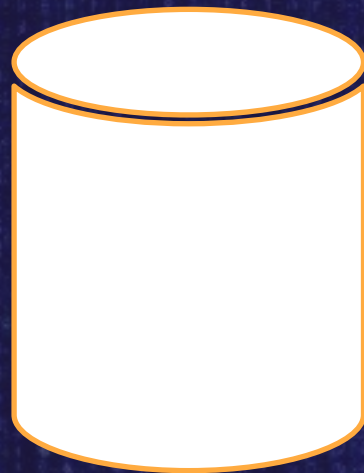
SELECT  
... FROM ...  
WHERE id = ?

query without  
the var

id = 2

var

step 1



step 2

## Create Homepage

Require user name and product name table



1

## Allow users to personalise profiles

Users can upload their own avatar to their profile

2

3

4

5

## Make product table

Take user input to filter the results based on matches in the product name and description





### Create Homepage

Require owner and product listing table



1

### Allow users to personalize profiles

Users can create their own  
avatar and profile



2

### Link product pages from homepage

Each entry in the product  
listing table should have a  
link to the specific product  
page

5

### Main product table

Take user input to filter the  
results based on matches in  
the product name and  
description



### Unique product page per product

Create a page per product so  
that people can link to  
specific products directly

4

3







## Create Homepage

Require owner and product listing table



## Allow users to personalize profiles

Users can create their own  
avatar and profile



## Link product pages from homepage

Each entry in product  
listing table should have a  
link to the specific product  
page



1

2

3

4

5

## Make product table searchable

Take user input to filter the  
results based on matches in  
the product name and  
description



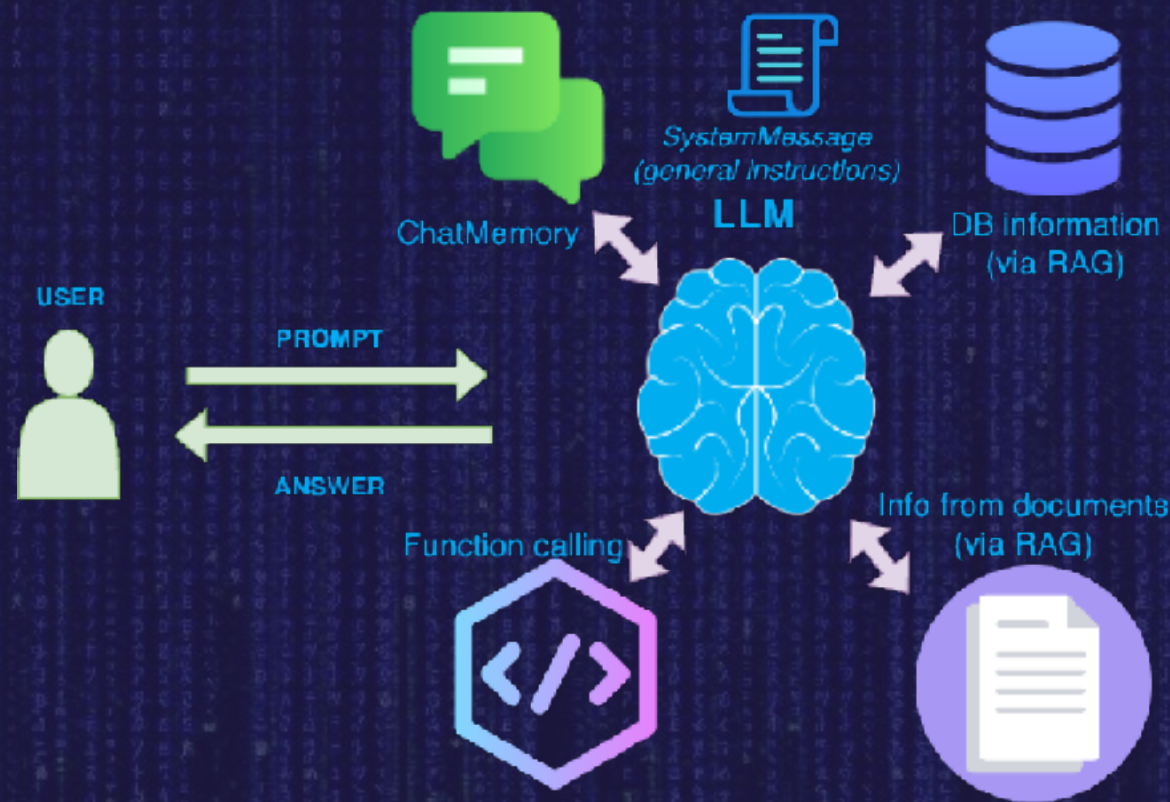
## Update product page per product

Create a page for each product so  
that people can link to  
specific product directly



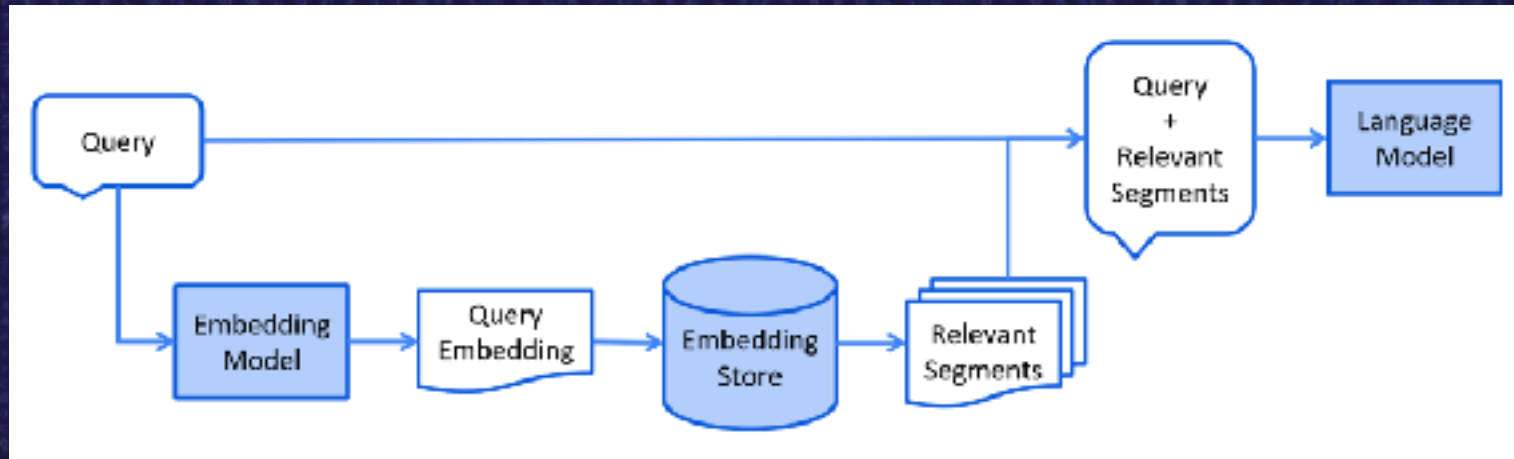


**Why does this happen?**





**“RAG is the way to find and inject relevant pieces of information from your data into the prompt before sending it to the LLM”**





# Learnings

## How to Approach secure AI assisted development



# Takeaways

## Education and awareness

- Write up policies and company guidelines
- Focus on security vulns, sensitive data and IP and human interaction
- Make education actionable making sure repeatable steps can be taken





# Takeaways



## Education and awareness

- Write up policies and company guidelines
- Focus on security vulns, sensitive data and IP and human interaction
- Make education actionable making sure repeatable steps can be taken

## Don't trust. Verify.

- Treat AI code like it's from an inexperienced dev/app sec engineer
- Test/validate everything
- Pair ChatGPT, Co-Pilot with AppSec Testing in the IDE





# Takeaways



## Education and awareness

- Write up policies and company guidelines
- Focus on security vulns, sensitive data and IP and human interaction
- Make education actionable making sure repeatable steps can be taken

## Don't trust. Verify.

- Treat AI code like it's from an inexperienced dev/app sec engineer
- Test/validate everything
- Pair ChatGPT, Co-Pilot with AppSec Testing in the IDE





# What's next?

Try Snyk Code for free

<https://snyk.io/product/snyk-code/>



The screenshot shows the Visual Studio Code interface with the Snyk Security extension installed. The left sidebar displays the 'EXTENSIONS' view with a search bar containing 'Snyk Security - Code and Op'. Below the search bar, the extension 'Snyk Security - Code and Op' is listed with a description 'Easily find and fix vulnera...'. The main panel shows the details for 'Snyk Security - Code and Open Source Depend', including the Snyk logo, a rating of 4.5 stars (11 reviews), and a description 'Easily find and fix vulnerabilities in both your code and open source dependencies w...'. At the bottom, there are buttons for 'Enable (Workspace)', 'Uninstall', and a refresh icon.

EXTENSIONS:...

Extension: Snyk Security - Code and Open Source Dependencies ✕

"Snyk Security - Code and Op

Snyk Security - Code a...  
Easily find and fix vulnera...  
Snyk

Snyk Security - Code and Open Source Depend  
Snyk | 60,583 | ★★★★★ (11)  
Easily find and fix vulnerabilities in both your code and open source dependencies w...

Enable (Workspace) Uninstall ↕ ↻ ⚙





**Be a secure developer,  
Code securely.**

**Thank you!**