



International Council on Systems Engineering
A better world through a systems approach

Value of Using Large Language Models in Building Software for Systems

Mark Sherman

Software Engineering Institute
Carnegie Mellon University



[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Copyright 2025 Carnegie Mellon University.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

References herein to any specific entity, product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by Carnegie Mellon University or its Software Engineering Institute nor of Carnegie Mellon University - Software Engineering Institute by any such named or represented entity.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

This work product was created in part using generative AI.

CERT® and Carnegie Mellon® are registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM25-0966

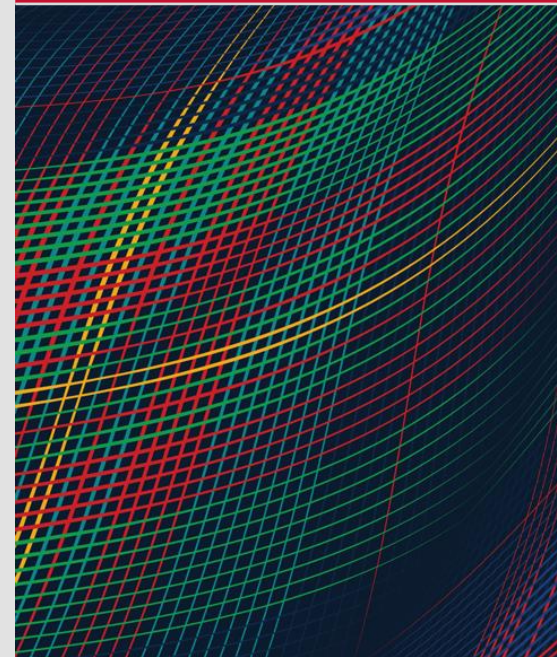


**Carnegie
Mellon
University**
Software
Engineering
Institute

Value of Using Large Language Models in Building Software for Systems

INCOSE JULY 30, 2025

David Schulker
Mark Sherman



Acknowledgments



Robert Schiela
Deputy Technical Director



David Svoboda
Software Security Engineer



David Schulker
Senior Data Scientist



Jamie Glenn
Operations Coordinator



Genavieve Chick
Intern

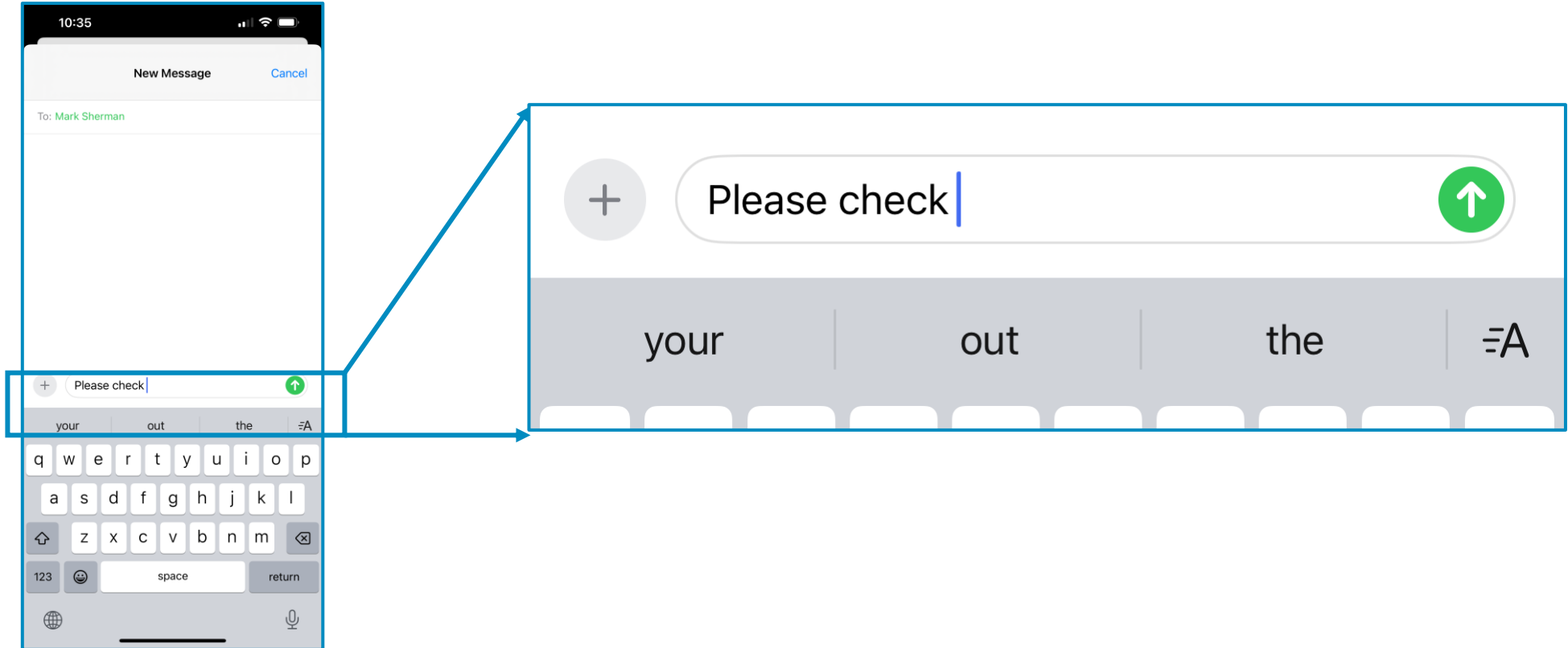


Vedha Avali
Intern

I want to thank the following members of the Cybersecurity Foundations group at CERT for their expert assistance with evaluating and preparing these materials:

- **JRobert Schiela SME** Technical Review
- **David Svoboda SME** (Representative, ISO/IEC JTC1/SC22/WG14 (C Programming Language) and lead of Undefined Behaviors Study Group (UBSG))
- **David Schulker** Data Scientist and co-Lead
- **Genavieve Chick** Data Science Intern
- **Vedha Avali** Data Science Intern
- **Jamie Glenn** Experiment management

Autocomplete is familiar to everyone



Autocomplete scaled on steroids

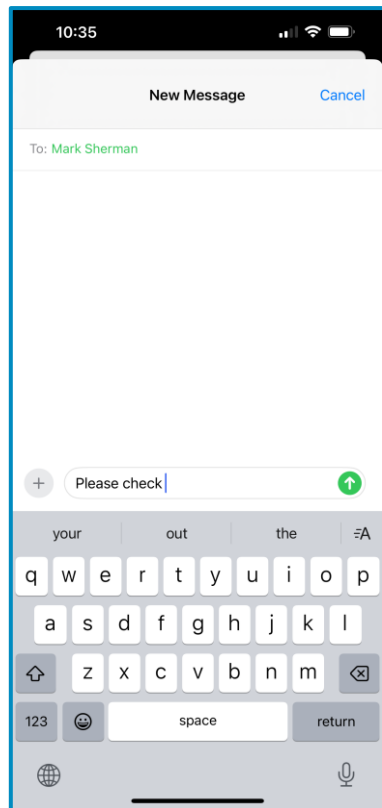


Photo credit: AI-generated photo by @ greta. <https://stock.adobe.com/images/aerial-view-of-a-data-center/962228532>

Autocomplete now!

ChatGPT 4o: "Please write a sonnet in the style of Shakespeare about my love of spicy food."

Autocomplete now!

ChatGPT 4o: "Please write a sonnet in the style of Shakespeare about my love of spicy food."

A Sonnet to Spice

O fiery muse that dost my tongue enflame,
With heat that dances, fierce yet sweetly bold,
No dull repast could ever taste the same,
Once kissed by flames of crimson, bright and gold.
The chili's burn doth wake my slumb'ring soul,
Like passion's touch that lingers on my lips,
Its embered fire I willingly extol,
Though scorching pain may bring me to eclipse.
Yet from this blaze I turn not, nor retreat,
For in the heat, my heart doth find delight.
Each molten bite, a battle won in heat,
A war of taste where pleasure meets the fight.
So let the peppers blaze and bring me bliss,
For love is fire, and spice, a burning kiss.

Autocompletion is Established Technology for IDEs Years Ago

Using AI For Autocompletion

```
1 import os
2 import sys
3
4 # Count lines of code in the given directory, separated by file extension
5 def main(directory):
6     line_count = {}
7     for filename in os.listdir(directory):
8         ext = os.path.splitext(filename)[1]
9         if ext not in line_count:
10             line_count[ext] = 0
11         for line in open(os.path.join(directory, filename)):
12             line_count[ext] += 1
13         line_count[ext] += 1
14         line_count[ext] += 1
15         line_count[ext] += 1
16         line_count[ext].append(
17             line
18
19
```

© 2019 TabNine, See <https://tabnine.com/eula>

Safe, correct code could be written incrementally

- Using n-grams
- Using deep learning (Generative Pretrained Transformer 2)

Sources:

E. Schutte, Autocomplete from StackOverflow, 2016, <https://emilsschutte.com/stackoverflow-autocomplete/>

(Jacob Jackson) TabNine, "Autocompletion with deep learning," July 18, 2019, <https://tabnine.com/blog/deep>

L. Tung, "New tool promises to turbo-charge coding in major programming languages," July 25, 2019, <https://www.zdnet.com/article/new-tool-promises-to-turbo-charge-coding-in-major-programming-languages/>

Mark Sherman, Using AI to Build More Secure Software, Abstractions II Conference, August 23, 2019, https://insights.sei.cmu.edu/documents/4246/2019_017_001_553383.pdf

Autocompletion is Established Technology for IDEs Years Ago

Using AI For Autocompletion

```
1 import os
2 import sys
3
4 # Count lines of code in the given directory, separated by file extension
5 def main(directory):
6     line_count = {}
7     for filename in os.listdir(directory):
8         ext = os.path.splitext(filename)
9         if ext not in line_count:
10             line_count[ext] = 0
11         for line in open(os.path.join(directory, filename)):
12             line_count[ext] += 1
13             line_count[ext] += 1      13%
14             line_count[ext] += 1      Tab 20%
15             line_count[ext] += 1      3 14%
16             line_count[ext].append( 4 3%
17             line                    5 23%
18
19
```

© 2019 TabNine, See <https://tabnine.com/eula>

Safe, correct code could be written incrementally

- Using n-grams
- Using deep learning (Generative Pretrained Transformer 2)

Note: “GPT”

Sources:

E. Schutte, Autocomplete from [StackOverflow](https://stackoverflow.com/stackoverflow-autocomplete/), 2016, <https://emilischutte.com/stackoverflow-autocomplete/>

(Jacob Jackson) TabNine, “[Autocompletion with deep learning](https://tabnine.com/blog/deep),” July 18, 2019, <https://tabnine.com/blog/deep>

L. Tung, “New tool promises to turbo-charge coding in major programming languages,” July 25, 2019, <https://www.zdnet.com/article/new-tool-promises-to-turbo-charge-coding-in-major-programming-languages/>

Mark Sherman, Using AI to Build More Secure Software, Abstractions II Conference, August 23, 2019, https://insights.sei.cmu.edu/documents/4246/2019_017_001_553383.pdf

Autocomplete now!

ChatGPT 4: "Write a Pascal program that will convert a Julian calendar date to a Hebrew (lunar) calendar date."

```
program JulianToHebrew;
uses SysUtils, Math;

type
  TDate = record
    Year, Month, Day: Integer;
  end;

var
  JulianDate, HebrewDate: TDate;

function JulianToHebrewDate(Julian: TDate): TDate;
var
  HebrewYear, HebrewMonth, HebrewDay: Integer;
begin
  // This is where the actual conversion algorithm should
  // be implemented.
  // Placeholder values are assigned for demonstration.
  HebrewYear := (Julian.Year - 3760); // Simplified
  // example, adjust accordingly
  HebrewMonth := 1; // Placeholder
  HebrewDay := 1; // Placeholder

  Result.Year := HebrewYear;
  Result.Month := HebrewMonth;
  Result.Day := HebrewDay;
end;
```

```
procedure ReadDate(var Date: TDate);
begin
  Write('Enter year: ');
  ReadLn(Date.Year);
  Write('Enter month: ');
  ReadLn(Date.Month);
  Write('Enter day: ');
  ReadLn(Date.Day);
end;

procedure PrintDate(Date: TDate);
begin
  WriteLn('Year: ', Date.Year, ' Month: ', Date.Month, '
Day: ', Date.Day);
end;

begin
  WriteLn('Enter the Julian date:');
  ReadDate(JulianDate);

  HebrewDate := JulianToHebrewDate(JulianDate);


  WriteLn('Converted Hebrew date:');
  PrintDate(HebrewDate);
end.
```

"Silicon Valley" is Enthusiastic

TECH

Mark Zuckerberg says AI could soon do the work of Meta's midlevel engineers

Lakshmi Varanasi Jan 11, 2025, 1:28 PM ET [Share](#) [Save](#)



Meta CEO Mark Zuckerberg says the work of midlevel software engineers can soon be outsourced to AI. BRENDAN SMIALOWSKI/ Getty Images

- Mark Zuckerberg said Meta will start automating the work of midlevel software engineers this year.
- Meta may eventually outsource all coding on its apps to AI.

<https://www.businessinsider.com/mark-zuckerberg-meta-ai-replace-engineers-coders-joe-rogan-podcast-2025-1>



Marc Benioff, Salesforce Founder: Why Salesforce Isn't Hiring Software Engineers | E1236

"We have increased the productivity this year ... AI ... by more than 30%. ... we will have less support engineers next year."

<https://www.youtube.com/watch?v=Xgsxi7IGMEU>

<https://www.salesforceben.com/salesforce-will-hire-no-more-software-engineers-in-2025-says-marc-benioff/>

Telling the Tale to Wall Street



We need 70% less coders from third parties to code as the AI handles most of the coding, the human only needs to look at the final 30% to validate it.

Murray Auchincloss, CEO & Director
BP Oil
Earnings Call, May 7, 2024

<https://seekingalpha.com/article/4690194-bp-p-l-c-bp-q1-2024-earnings-call-transcript>

Today, more than a quarter of all new code at Google is generated by AI, then reviewed and accepted by engineers.

Sundar Pichai, CEO Alphabet and Google
Earnings call, Oct 29, 2024

<https://abc.xyz/assets/1f/d2/b4b2a1c4437395dce09645f71173/2024-q3-earnings-transcript.pdf>

Academic Research More Nuanced

Participants who had access to an AI assistant wrote significantly less secure code.

Do Users Write More Insecure Code with AI Assistants
Perry, Srivastava, Kmar, Boneh, CCS '23, Nov 26-20, 2023,
<https://arxiv.org/pdf/2211.03622>

Copilot [produced] 1,689 programs. Of these, we found approximately 40% to be vulnerable.

Pearce, Ahmad, Tan, Dolan-Gavin, Larri, Asleep at the Keyboard?
Assessing the Security of GitHub CoPilot's Code Contributions, IEEE Symposium on Security and Privacy, 2022,
<https://arxiv.org/pdf/2108.09293>

AI coding tools continue to consistently generate insecure code.

"AI Code, Security and Trust: Organization Must Change Their Approach," 2023 Snyk AI-Generated Code Security Report,
<https://go.snyk.io/2023-ai-code-security-report-dwn-ty.html>

Researchers from Johannes Gutenberg University and University College London ... gave 40 pieces of buggy code to [ChatGPT]. ... ChatGPT solved 19 problems.

<https://www.pcmag.com/news/watch-out-software-engineers-chatgpt-is-now-finding-fixing-bugs-in-code>
<https://dl.acm.org/doi/10.1145/3135932.3135941>
<https://arxiv.org/abs/2301.08653>

We introduce an evaluation framework consisting of 2,294 software engineering problems ... The best-performing model ... is able to solve a mere 1.96% of the issues.

Jimenez, Yang, Wettig, Yao, Pei, Press, and Narasimhan, "SWE-BENCH: CAN LANGUAGE MODELS RESOLVE REAL-WORLD GITHUB ISSUES?", Nov 11, 2024,
<https://arxiv.org/pdf/2310.06770>

Can Large Language Models be Used in Building Software for Systems? – Our Experiment

Ask ChatGPT to identify the presence or absence of errors in 1223 examples of compliant and noncompliant code from SEI CERT C, C++ and Java Secure Coding Standards

- ChatGPT 3.5 as of March 27, 2023, and Jun-Aug 2024
- ChatGPT 4 and GitHub CoPilot as of February 16, 2024
- ChatGPT 4o as of Jun-Aug 2024
- 2684 runs, each trial run as a new conversation
- No trial systematically repeated
- All examples (with solutions) available on the web during ChatGPT's training
- Some examples have comments suggesting error
- ChatGPT's performance evaluated by SMEs

Mark Sherman, "Using ChatGPT to Analyze Your Code? Not So Fast," February 12, 2024, <https://insights.sei.cmu.edu/blog/using-chatgpt-to-analyze-your-code-not-so-fast/>
David Schulker, "GenAI for Code Review of C++ and Java," November 18, 2024 <https://insights.sei.cmu.edu/blog/genai-for-code-review-of-c-and-java/>

CERT Secure Coding Standards



Collected wisdom from thousands of contributors on community wiki since Spring 2006

<http://securecoding.cert.org>

- SEI CERT C Coding Standard
- SEI CERT C++ Coding Standard
- CERT Oracle Secure Coding Standard for Java

Secure Coding Training and Professional Certificates

- CERT Secure Coding in C and C++
- CERT Secure Coding in Java

International Standards Participation

- ISO/IEC C Programming Language
- ISO/IEC C++ Programming Language



Example Standard Page (FIO47-C:Use valid format strings)

Noncompliant Code Example

Mismatches between arguments and conversion specifications may result in [undefined behavior](#). Compilers may diagnose type mismatches in formatted output function invocations. In this noncompliant code example, the `error_type` argument to `printf()` is incorrectly matched with the `s` specifier rather than with the `d` specifier. Likewise, the `error_msg` argument is incorrectly matched with the `d` specifier instead of the `s` specifier. These usages result in [undefined behavior](#). One possible result of this invocation is that `printf()` will interpret the `error_type` argument as a pointer and try to read a string from the address that `error_type` contains, possibly resulting in an access violation.

```
#include <stdio.h>

void func(void) {
    const char *error_msg = "Resource not available to user.";
    int error_type = 3;
    /* ... */
    printf("Error (type %s): %d\n", error_type, error_msg);
    /* ... */
}
```

Compliant Solution

This compliant solution ensures that the arguments to the `printf()` function match their respective conversion specifications:

```
#include <stdio.h>

void func(void) {
    const char *error_msg = "Resource not available to user.";
    int error_type = 3;
    /* ... */
    printf("Error (type %d): %s\n", error_type, error_msg);
    /* ... */
}
```

Example Standard Page (FIO47-C:Use valid format strings)

Noncompliant Code Example

Mismatches between arguments and conversion specifications may result in [undefined behavior](#). Compilers may diagnose type mismatches in formatted output function invocations. In this noncompliant code example, the `error_type` argument to `printf()` is incorrectly matched with the `s` specifier rather than with the `d` specifier. Likewise, the `error_msg` argument is incorrectly matched with the `d` specifier instead of the `s` specifier. These usages result in [undefined behavior](#). One possible result of this invocation is that `printf()` will interpret the `error_type` argument as a pointer and try to read a string from the address that `error_type` contains, possibly resulting in an access violation.

```
#include <stdio.h>
```

```
void func(void) {  
    const char *error_msg = "Resource not available to user.";  
    int error_type = 3;  
    /* ... */  
    printf("Error type %s): %d\n", error_type, error_msg);  
    /* ... */  
}
```

(type %s): %d\n

Compliant Solution

This compliant solution ensures that the arguments to the `printf()` function match their respective conversion specifications:

```
#include <stdio.h>
```

```
void func(void) {  
    const char *error_msg = "Resource not available to user.";  
    int error_type = 3;  
    /* ... */  
    printf("Error type %d): %s\n", error_type, error_msg);  
    /* ... */  
}
```

(type %d): %s\n

Finding and fixing an error (FIO47-C:Use valid format strings)

ChatGPT 3.5

Example of noncompliant code:

```
#include <stdio.h>

void func(void) {
    const char *error_msg = "Resource not available to
user.";
    int error_type = 3;
    /* ... */
    printf("Error (type %s): %d\n", error_type, error_msg);
    /* ... */
}
```

ChatGPT 3.5:

The format specifier in the printf statement is incorrect.

ChatGPT 4.0:

The correct `printf` statement should be as follows:
**printf("Error (type %d): %s\n",
error_type, error_msg);**

CoPilot:

Here is the corrected code:
**printf("Error (type %d): %s\n",
error_type, error_msg);**

All Versions Missing Errors

DCL38-C Use the correct syntax when declaring a flexible array member

ChatGPT3.5: There is **nothing technically wrong with this code**

ChatGPT4: the provided **code snippet is an appropriate use** of a flexible array member within a struct for dynamic array allocation in C.

CoPilot: The **code seems to be correct** in terms of syntax and logic.

Improved Technology Sometimes Gives Better Results

EXP33-C Do not read uninitialized memory

ChatGPT3.5: The code appears to be incomplete as it **doesn't have any obvious errors**.

ChatGPT4: There are a couple of issues with the provided code ... ****Uninitialized Variable:****

CoPilot: The issue with this code is that the **sign variable** in the `is_negative` function is **not initialized**.

Evaluating Correction Can be Difficult

EXP42-C Standard

Compliant recommendation – correct code

0 == memcmp(left->buffer, right->buffer, 13)

```
char c;  
int i;  
char buffer[13];  
};  
  
void compare(const struct s *left, const struct s *right) {  
    if ((left && right) &&  
        (left->c == right->c) &&  
        (left->i == right->i) &&  
        (0 == memcmp(left->buffer, right->buffer, 13))) {  
        /* ... */  
    }  
}
```

ChatGPT 3.5's Subtly Mistaken Fix

The code appears to be comparing two instances of a struct **s** using **memcmp()**.

However, this may not always produce the expected results, especially if the struct contains padding bytes between its members. The padding bytes may have different values in the two instances of the struct, which would cause **memcmp()** to return a non-zero value even if the actual data members are the same.

To avoid this issue, it is recommended that, for example, the code could be modified as follows.

```
void compare(const struct s *left, const struct s *right) {  
    if ((left && right) &&  
        (left->c == right->c) &&  
        (left->i == right->i) &&  
        (0 == strcmp(left->buffer, right->buffer))) {  
        /* ... */  
    }  
}
```

Here, the struct members are compared individually using **==** for **char** and **int** members and **strcmp()** for the **buffer** member.

0 == **strcmp**(left->buffer, right->buffer)

- Assumed “buffer” is a string
- (If string, should be using **strncmp**, not **strcmp**)

Trying a serious security violation – Enabling Command Injection

(ENV33-C. Do not call system())

In this noncompliant code example, the `system()` function is used to execute `any_cmd` in the host environment.

```
#include <string.h>
#include <stdlib.h>
#include <stdio.h>

enum { BUFFERSIZE = 512 };

void func(const char *input) {
    char cmdbuf[BUFFERSIZE];
    int len_wanted = snprintf(cmdbuf, BUFFERSIZE,
                             "any_cmd '%s'", input);

    if (len_wanted >= BUFFERSIZE) {
        /* Handle error */
    } else if (len_wanted < 0) {
        /* Handle error */
    } else if (system(cmdbuf) == -1) {
        /* Handle error */
    }
}
```



system(cmdbuf)

Identifies a non-existent problem and misses real problem

(ENV33-C. Do not call system())

In this noncompliant code example, the `system()` function is used to execute `any_cmd` in the host environment.

```
#include <string.h>
#include <stdlib.h>
#include <stdio.h>

enum { BUFFERSIZE = 512 };

void func(const char *input) {
    char cmdbuf[BUFFERSIZE];
    int len_wanted = snprintf(cmdbuf, BUFFERSIZE,
                             "any_cmd '%s'", input);

    if (len_wanted >= BUFFERSIZE) {
        /* Handle error */
    } else if (len_wanted < 0) {
        /* Handle error */
    } else if (system(cmdbuf) == -1) {
        /* Handle error */
    }
}
```

snprintf(...)

system(cmdbuf)

any_cmd

ChatPGT3.5:

it's recommended to **use** a function that is specifically designed to handle string concatenation and avoid buffer overflows, such as **snprintf()**.

Revisiting ENV33 – Better

ChatGPT4:

****Security Vulnerability
(Command Injection):**** The most critical issue with this code is that it directly incorporates user input (``input``) into a command string that is executed by the system.

Revisiting ENV33 – Better – and surprising guardrails

ChatGPT4:

****Security Vulnerability (Command Injection):**** The most critical issue with this code is that it directly incorporates user input (``input``) into a command string that is executed by the system.

CoPilot:

Sorry, the response was filtered by the Responsible AI Service. Please rephrase your prompt and try again. [Learn more.](#)

(Tried 3 times.)

Consolidated Results

		Insecure Programs*			Secure Programs		
		Correct**	Incorrect	%	Correct***	Incorrect	%
C	ChatGPT 3.5	114 (110)	124	48%	97	112	46%
	ChatGPT 4	187 (175)	51	79%			
	ChatGPT 4o				175	34	84%
	CoPilot	172(160)	66	72%			
C++	ChatGPT 3.5	104	61	63%			
	ChatGPT 4o	138	27	84%			
Java	ChatGPT 3.5	108	166	39%	204	135	60%
	ChatGPT 4o	183	91	67%	268	71	79%

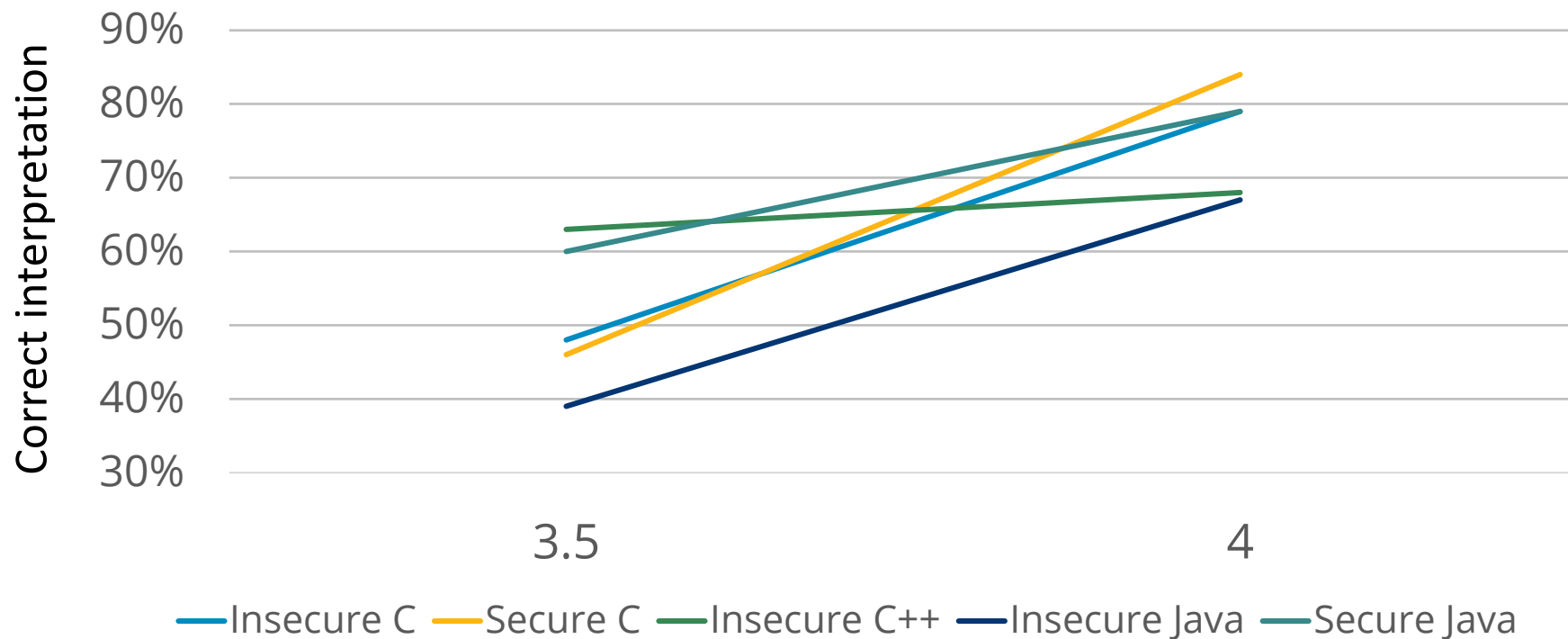
*Combining reports as “noticed but not error” as a “error found”

** Parenthetical values exclude “noticed but not error” reports

*** Not combining “noticed” and “found”

Improved performance

Improvement from 3.5 to 4 (4, CoPilot, 4o)



Field Experience – Consultant and Developer



Photo credit: AI-generated photo by [Tomislav](https://stock.adobe.com/images/old-fictional-wizard-using-a-laptop-computer/583224250).
<https://stock.adobe.com/images/old-fictional-wizard-using-a-laptop-computer/583224250>

“If you’re not a great coder, you’re not going to be able to have the judgement to know if the code that comes out is going to do what you want in the way you want.

Bret Greenstein, Data and AI Leader, PwC,

Can AI coding Tools meet enterprise expectations, (quoted by Lindsey Wilkinson, Feb 26, 2024, <https://www.ciodive.com/news/generative-ai-coding-tools-reality-versus-hype/708426/>)

“When LLMs make mistakes — and they do, regularly — it takes an experienced developer to spot these issues, understand their implications, and know how to fix them.”

Rahul Agarwal, Machine Learning Engineer, Facebook, LLMs Won’t “Why Replace Programmers: A Reality Check,” Nov 25, 2024, <https://mlwhiz.medium.com/why-llms-wont-replace-programmers-a-reality-check-73133bd3b9cc>

Users Defer to Perceived AI Authority



Photo generated by Firefly in response to “Mystical wizard with the appearance of a futuristic computer.” April 2024 release. Adobe. June 2024.

<https://www.adobe.com/products/firefly/features/text-to-image.html>

“Participants with access to an AI assistant were also more likely to believe they wrote secure code, suggesting that such tools may lead users to be overconfident about security flaws in their code.”

Do Users Write More Insecure Code with AI Assistants

Perry, Srivastava, Kmar, Boneh, CCS '23, Nov 26-20, 2023, <https://arxiv.org/pdf/2211.03622>

“Respondents commonly found security issues with AI suggestions but they also **strongly believed that AI suggestions were secure.**”

“AI Code, Security and Trust: Organization Must Change Their Approach,” 2023 Snyk AI-Generated Code Security Report, <https://go.snyk.io/2023-ai-code-security-report-dwn-typ.html>

Amazon's Experience

Amazon Experience using LLMs for Software Development

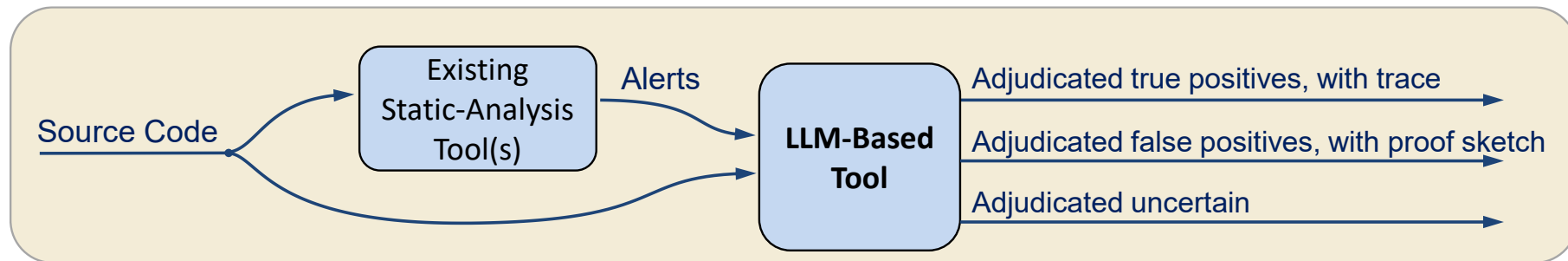
Task: Upgrade existing Java 8 and Java 11 programs to Java 17

Claim: saved the equivalent of 4,500 developer-years

Approach

- Train an LLM specifically on Java 8/11 to Java 17 upgrade
- Process programs written according to Amazon standards
- Leverage OpenRewrite recipes for describing the changes
- Reuse existing unit tests for existing programs to validate successful conversion

Drive LLM to Validate Itself



- Use other tools to generate candidate issues
- Use prompt engineering to focus on candidate issues
- Use LLM to generate patches
- Use LLM to generate evidence of patch correctness
- Validate patch correctness externally, e.g., SMT solvers (aka theorem prover)

Figure 1. Using LLMs for SA Alert Adjudications. Klieber, Flynn, “Using LLMs for SA alert adjudication and rationales”, CrossTalk, 2024

Klieber and Flynn, “Using LLMs to Adjudicate Static-Analysis Alerts,” 58th Hawaii International Conference on System Sciences (HICSS), January 10, 2025, <https://scholarspace.manoa.hawaii.edu/items/42b220fa-17f8-4cfe-9f2b-ba4553e21bc3>

Take Aways

- Don't abandon existing methods for improving software security
 - Secure coding practices and source code analysis tools remain valuable
- Educate management (and developers) on practical applications of LLMs
 - Separate hype from capability
- Adopt low hanging fruit, e.g., autocompletion, documentation surfing, contained code “snippets”
 - Easy productivity gains from basic function
 - Correctness and security manifest
- If situation allows, consider investing in specialty applications
 - Extremely well-defined context and target
 - Have or acquire technical talent to construct LLM applications
 - Practical method for evaluation of LLM output

Thank You

Mark Sherman

Carnegie Mellon University
Software Engineering Institute
4500 Fifth Avenue
Pittsburgh, PA 15213-2612
412-268-5800, 888-201-4479
info@sei.cmu.edu