MANCHESTER
1824
The University of Manchester

ESBMC-AI
Bounded Model Checking enhanced by Large Language Models. Frontend for ESBMC.

Yiannis Charalambous
yiannis-charalambous.com

Motivation

Ease

ESBMC’s counterexample is difficult to read and understand, it can be explained using LLMs. Questions can also be asked about the code and counterexample and answered by the LLM.

Unique Use Cases

LLMs enable fixing and optimizing the code. ESBMC can then verify the efforts.

Progress of ESBMC-AI

April

- User Chat mode.
- AI uses in-context learning to learn the source code and ESBMC output.
- LLM is asked questions regarding topic and answers.

May

Fix Code Mode

Very basic implementation, but showed very promising results in small samples. Lot’s of room for improvement.

Progress of ESBMC-AI

June

Research Paper Released

- A New Era in Software Security: Towards Self-Healing Software via Large Language Models and Formal Verification

Added Additional LLMs

- Experiments showed limitations in the open source domain.
- Good performance with OpenAI GPT3 and GPT4.

September

Optimize Code Mode Experimental Prototype Release

- Very poor performance. Lot’s of room for improvement!

Progress of ESBMC-AI

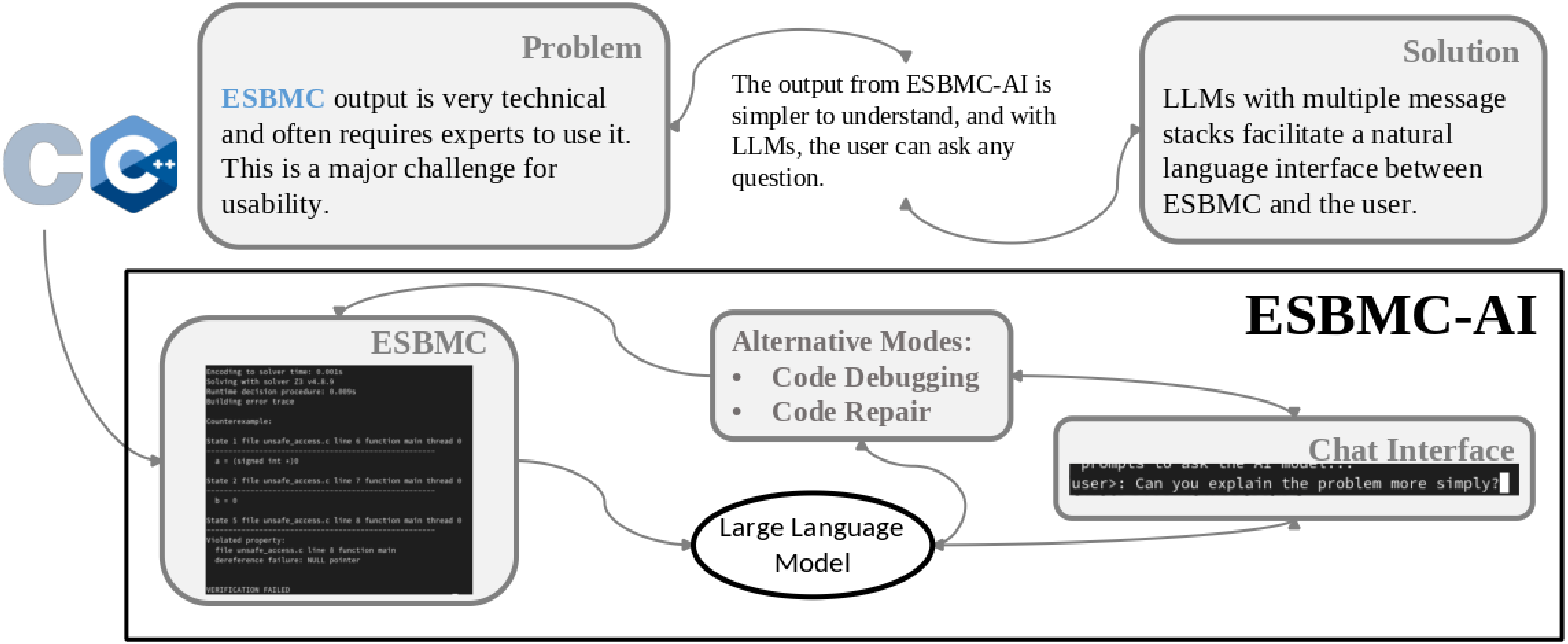
Planned Features

- Optimize Code Mode improvements in optimization and partial equivalence checking accuracy.
- Fix Code Mode 2.0: Better accuracy in fixing larger samples!
- Add and test other LLMs (Meta/Google)

Possibly Also Add

- Build a Static Verification LLM for ESBMC-AI (efficient in fixing vulnerable code and optimizing it).

ESBMC-AI: AI-driven, software development and debugging for C/C++ applications. An AI augmentation layer for ESBMC.



Problem

ESBMC output is very technical and often requires experts to use it. This is a major challenge for usability.

The output from ESBMC-AI is simpler to understand, and with LLMs, the user can ask any question.

Solution

LLMs with multiple message stacks facilitate a natural language interface between ESBMC and the user.

ESBMC-AI

ESBMC

```
Encoding to solver time: 0.001s
Solving with solver Z3 v4.8.9
Runtime decision procedure: 0.009s
Building error trace

Counterexample:

State 1 file unsafe_access.c line 6 function main thread 0
----------------------------------------------------------
  a = (signed int *)0

State 2 file unsafe_access.c line 7 function main thread 0
----------------------------------------------------------
  b = 0

State 5 file unsafe_access.c line 8 function main thread 0
----------------------------------------------------------
Violated property:
  file unsafe_access.c line 8 function main
  dereference failure: NULL pointer


VERIFICATION FAILED
```

Alternative Modes:

- Code Debugging
- Code Repair

Large Language Model

Chat Interface

```
 prompts to ask the AI model...
user>: Can you explain the problem more simply?
```

ESBMC-AI is an AI-powered augmentation layer for ESBMC that uses LLM to simplify feedback in the form of natural language. It incorporates extra functionality that utilizes AI, such as automatic code repair that uses ESBMC in the back-end to ensure the generated code is correct concerning a specification

ESBMC-AI Modes of Operation

User Chat Mode

Chat assistant that explains the output of ESBMC to the user. Allows for natural conversations to occur regarding the code.

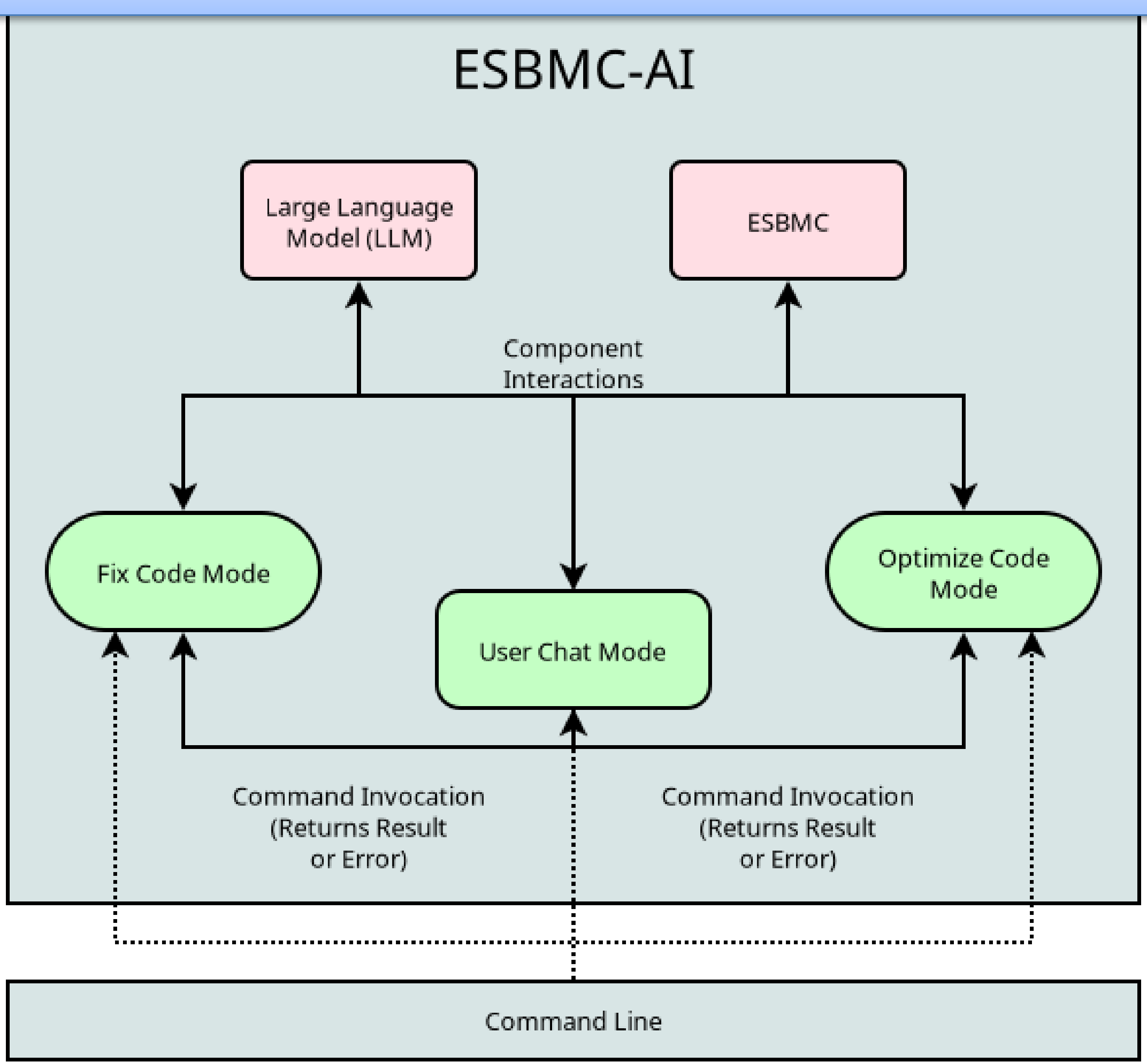
Fix Code Mode

Fix code using the LLM, verify it using ESBMC. Wrong generations are returned to the LLM for revaluation.

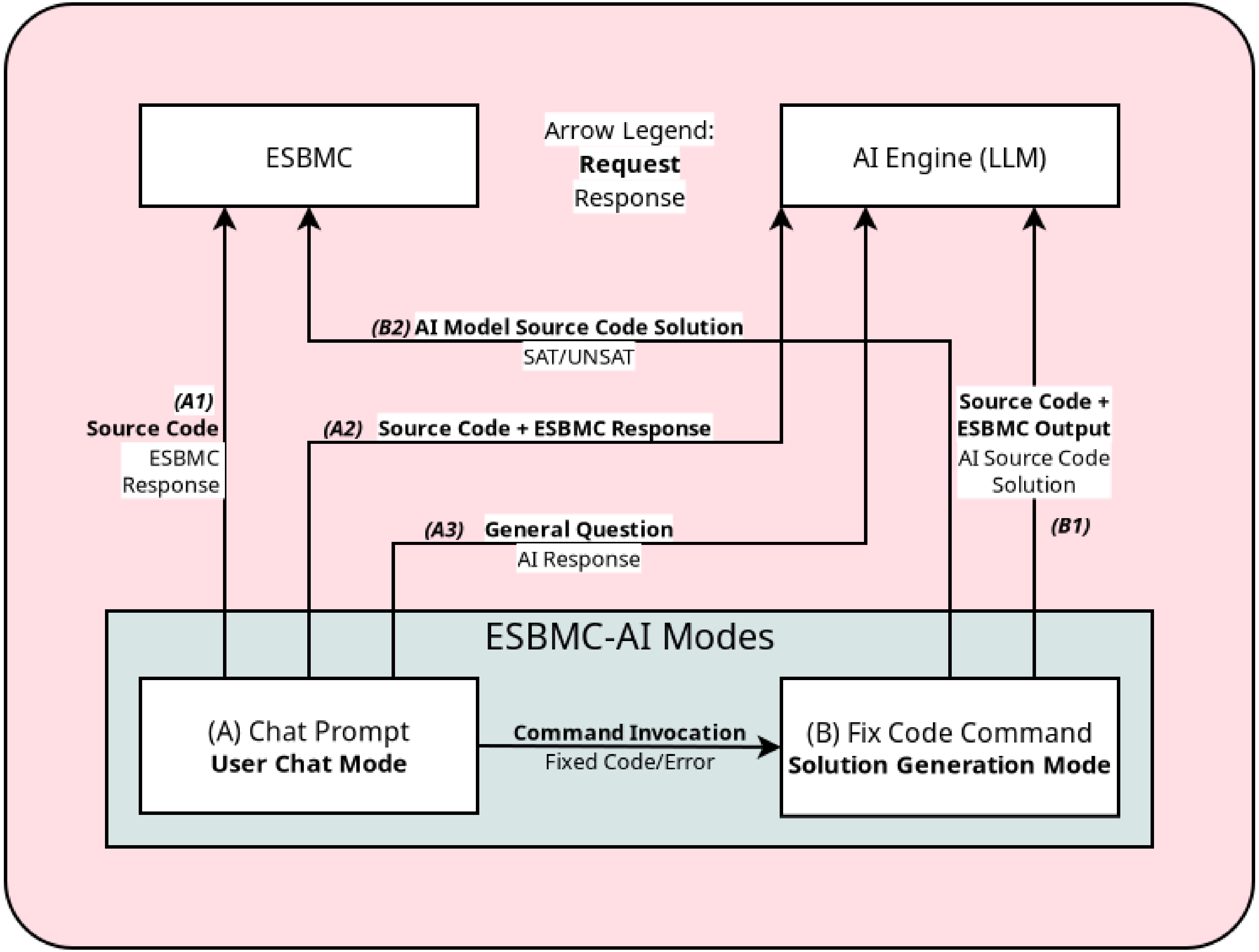
Optimize Code Mode

Optimize code segments the source code into function blocks and optimizes the functions individually.

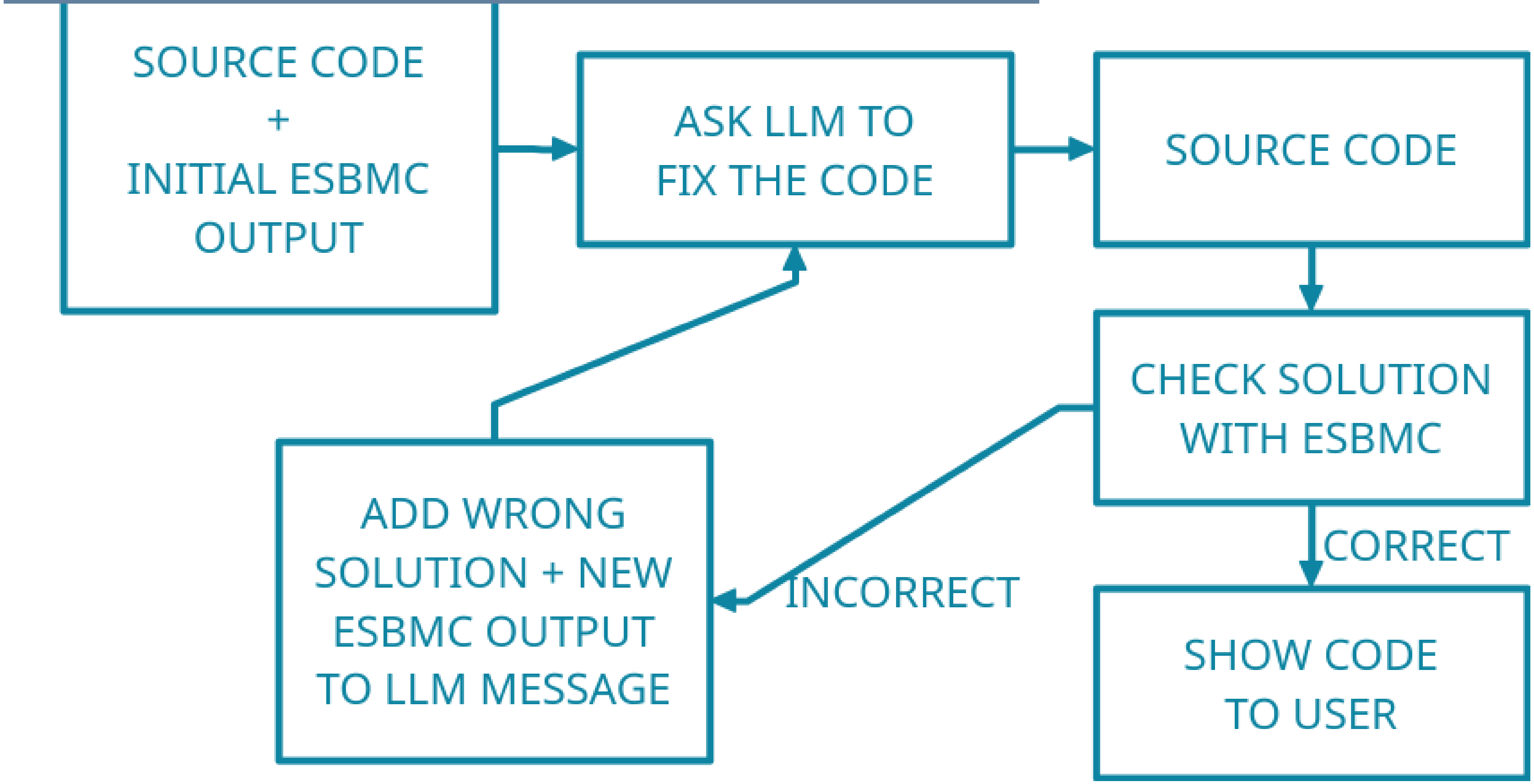
General Info


ESBMC-AI
Large Language Model (LLM)
ESBMC
Component Interactions
Fix Code Mode
User Chat Mode
Optimize Code Mode
Command Invocation (Returns Result or Error)
Command Invocation (Returns Result or Error)
Command Line

Fix Code Mode (FCM)



Fix Code Mode


SOURCE CODE + INITIAL ESBMC OUTPUT

ASK LLM TO FIX THE CODE

SOURCE CODE

CHECK SOLUTION WITH ESBMC

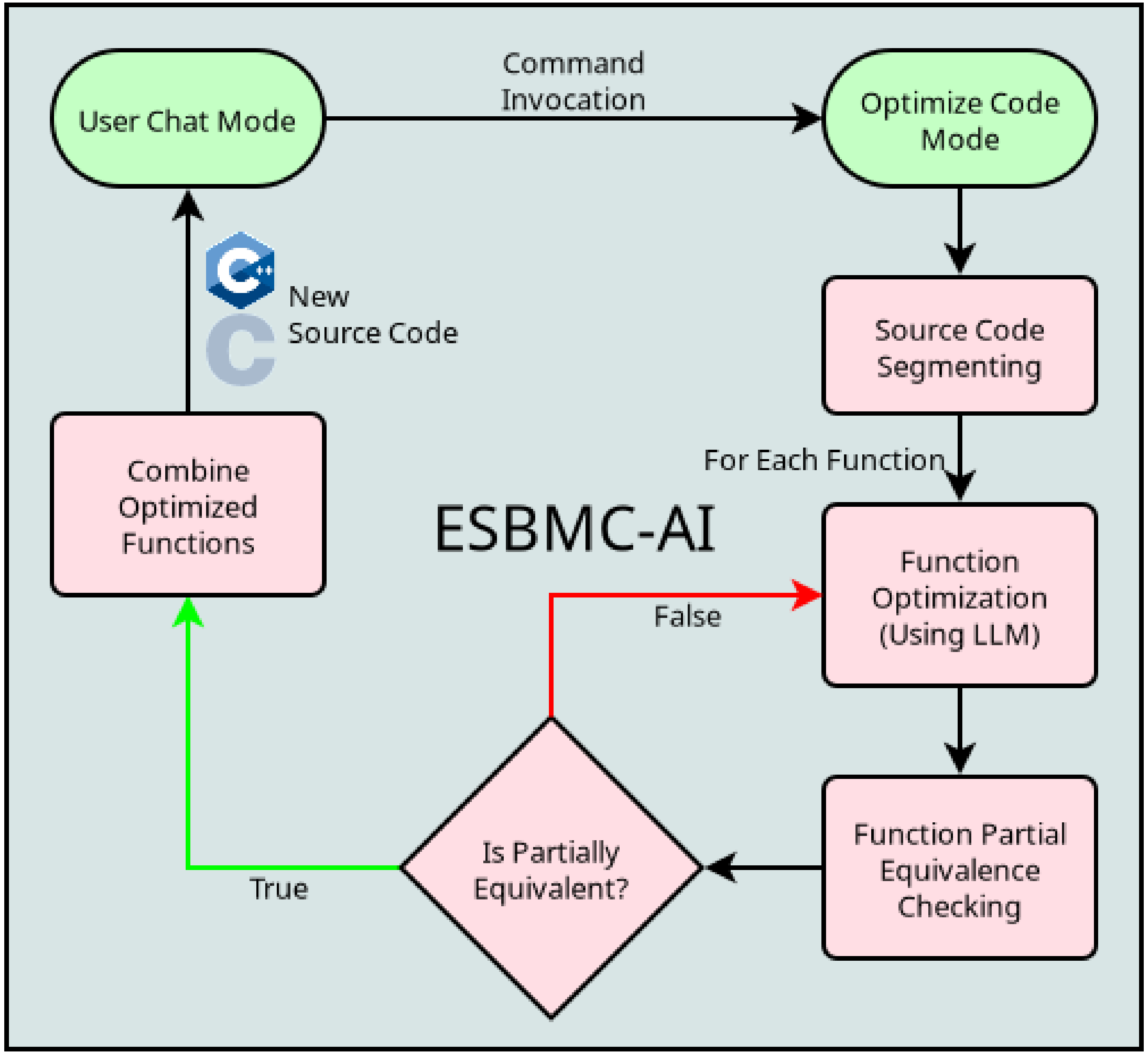
CORRECT

SHOW CODE TO USER

INCORRECT

ADD WRONG SOLUTION + NEW ESBMC OUTPUT TO LLM MESSAGE

Optimize Code Mode (OCM)



Demo

- YouTube Channel: https://www.youtube.com/@esbmc-ai
- Multi-threaded Example
- Optimize Code Mode Prototype Showcase

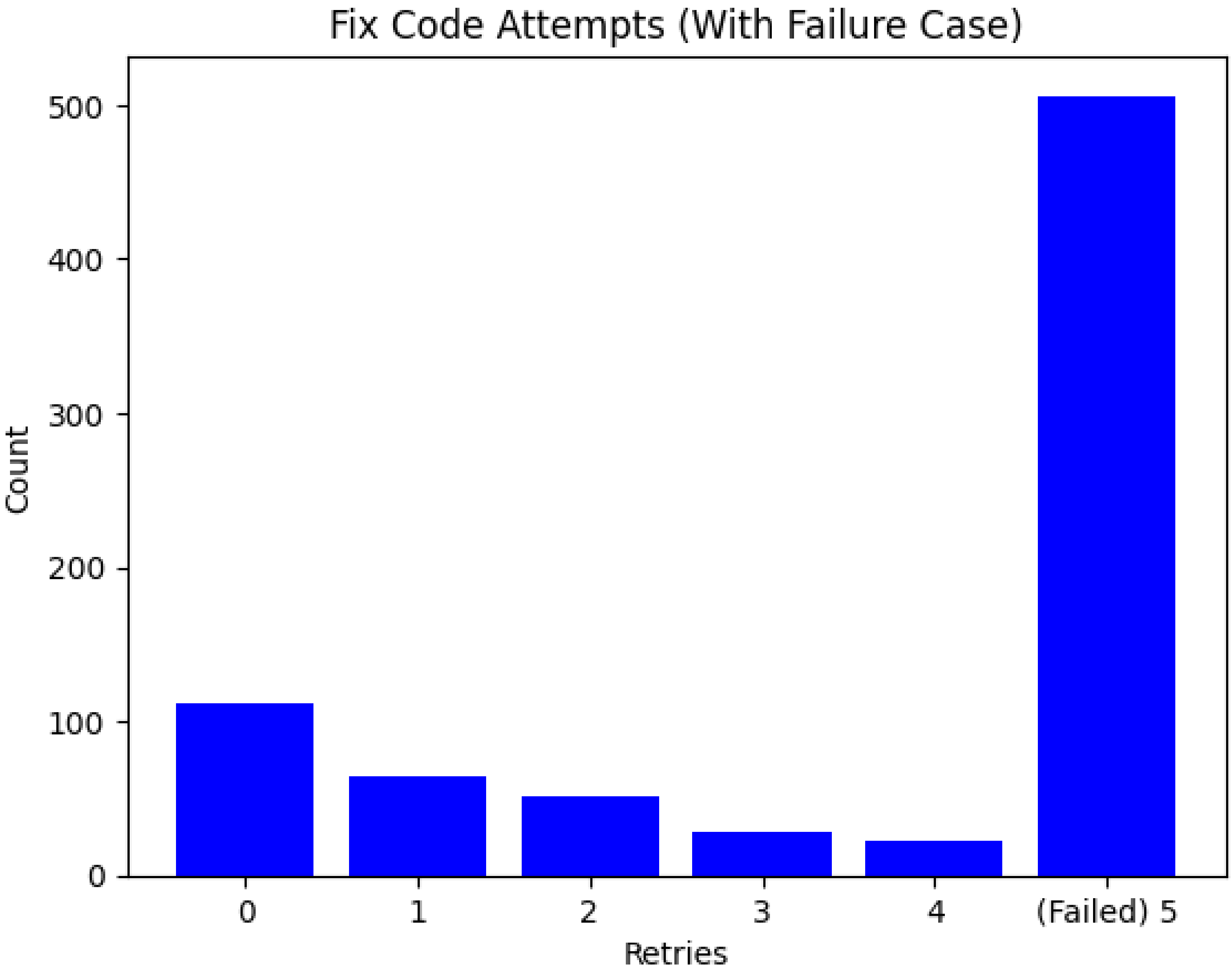
Performance

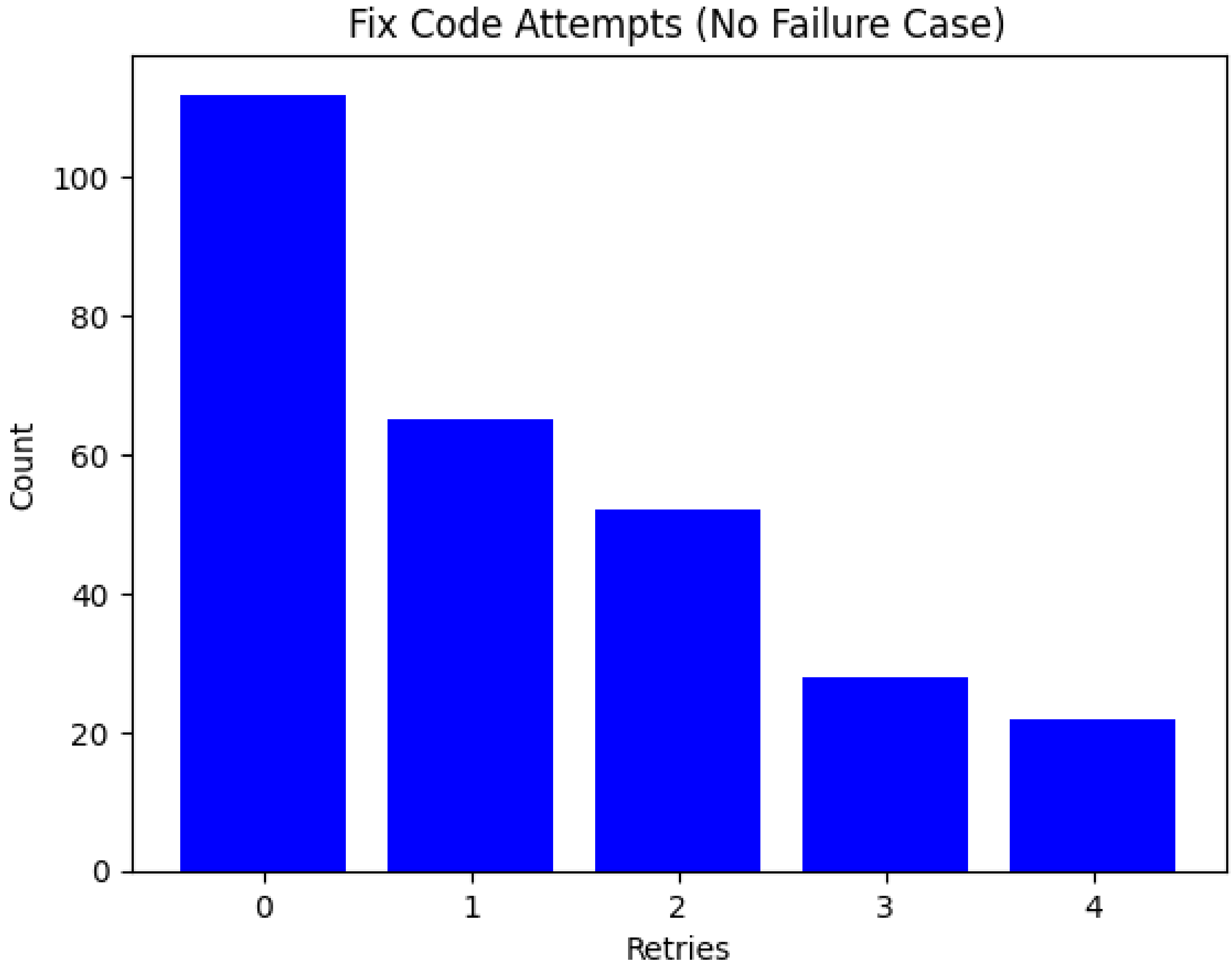
FormAI Dataset

Mode	Accuracy
Fix Code Mode	35.5%

Perceived Performance

Mode	Small Samples	Medium Samples	Large Samples
User Chat Mode	5	4	4
Fix Code Mode	3	2	2
Optimize Code Mode	2	0	0


Fix Code Attempts (With Failure Case)
Count
500
400
300
200
100
0
0
1
2
3
4
(Failed) 5
Retries



Fix Code Attempts (No Failure Case)
Count
0
20
40
60
80
100
Retries
0
1
2
3
4

Thank You

Yiannis Charalambous
yiannis-charalambous.com