

Experiment ‘P_3_1_classify_int’ Results

December 18, 2025

Experiment outcome: SUCCESS

Bad responses: 0

Responses containing assume : 0

Resolution attempts: 1

Hard fails (resolution): 0

Soft fails (resolution): 0

Verification attempts: 1

Problem Specification

Problem name: P_3_1_classify_int

Natural language statement: Write a method that takes an integer and returns whether it is negative, zero, or positive.

Method signature: P_3_1_classify_int (x:int) returns (s: string)

Ensures

- $x == 0 \implies s == \text{"zero"}$
- $x > 0 \implies s == \text{"positive"}$
- $x < 0 \implies s == \text{"negative"}$

GenAI interactions

Below you will find all interactions between the ‘user’ (program) and the ‘assistant’ (OpenAI).

Program → GenAI

You are given the following task to perform in Dafny:

Write a method that takes an integer and returns whether it is negative, zero, or
→ positive.

The signature should be:

```
method P_3_1_classify_int (x:int) returns (s: string)
```

The method should respect the following contract:

```
ensures x == 0 ==> s == "zero", ensures x > 0 ==> s == "positive", ensures x < 0 ==> s ==  
→ "negative"
```

Produce and show only the Dafny body of this method, including the curly braces that
→ surround it. Do not show the signature nor contract. You must not use 'assume'
→ anywhere in your code.

GenAI → Program

System fingerprint:

ID: chatcmpl-Co4S0XKxA1nr3ffR69VSb1ljVrpHS

Model: gpt-5.1-chat-latest

Created at: 1766049544

Finish reason: stop

Usage: 312 tokens in, and 70 tokens out

```
{  
    if x == 0 {  
        s := "zero";  
    } else if x > 0 {  
        s := "positive";  
    } else {  
        s := "negative";  
    }  
}
```

Final Program

```
method P_3_1_classify_int (x:int) returns (s: string)  
    ensures x == 0 ==> s == "zero"  
    ensures x > 0 ==> s == "positive"  
    ensures x < 0 ==> s == "negative"  
{
```

```
if x == 0 {  
    s := "zero";  
} else if x > 0 {  
    s := "positive";  
} else {  
    s := "negative";  
}  
}
```

Total Token Usage

Input tokens: 312

Output tokens: 70

Reasoning tokens: 0

Sum of ‘total tokens’: 382

Experiment Timings

Overall Experiment started at 1766049545918, ended at 1766049549055, lasting 3137ms (3.14 seconds)

Iteration #1 started at 1766049545921, ended at 1766049549055, lasting 3134ms (3.13 seconds)