Experiment 'P_2_4_4_avarage' 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_2_4_4_avarage
Natural language statement: Write a method that takes two integers and then returns the average.
Method signature: average(x: int, y: int) returns (result:real)

Ensures

- `result == (x + y) as real / 2.0`

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 two integers and then returns the average.

The signature should be:

method average(x: int, y: int) returns (result:real)

The method should respect the following contract:

ensures result == (x + y) as real / 2.0

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-Co4PCLULvG5PnH4NMhSIFUT9vXfVq
Model: gpt-5.1-chat-latest
Created at: 1766049370
Finish reason: stop
Usage: 260 tokens in, and 37 tokens out

```
{
  result := (x + y) as real / 2.0;
}
```

Final Program

```
method average(x: int, y: int) returns (result:real)
       ensures result == (x + y) as real / 2.0
{
  result := (x + y) as real / 2.0;
}
```

Total Token Usage

Input tokens: 260
Output tokens: 37
Reasoning tokens: 0
Sum of 'total tokens': 297

Experiment Timings

Overall Experiment started at 1766049372553, ended at 1766049375075, lasting 2522ms (2.52 seconds)
Iteration #1 started at 1766049372554, ended at 1766049375075, lasting 2521ms (2.52 seconds)