

Experiment 'P_3_7_Order' 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_7_Order

Natural language statement: Write a method that takes three integers and returns "in order" if they are sorted in ascending or descending order, or "not in order" otherwise.

Method signature: P_3_7_Order (i:int, j: int, k: int) returns (s: string)

Ensures

- `if (i < j < k) || (i > j > k) then s == "in order" else s == "not in order"`

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 three integers and returns "in order" if they are sorted in
→ ascending or descending order, or "not in order" otherwise.

The signature should be:

```
method P_3_7_Order (i:int, j: int, k: int) returns (s: string)
```

The method should respect the following contract:

```
ensures if (i < j < k) || (i > j > k) then s == "in order" else s == "not in order"
```

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-Co4SovRRcuHFYhAaNEakyzMt0GkJq

Model: gpt-5.1-chat-latest

Created at: 1766049594

Finish reason: stop

Usage: 346 tokens in, and 134 tokens out

```
{
  if ((i < j && j < k) || (i > j && j > k)) {
    s := "in order";
  } else {
    s := "not in order";
  }
}
```

Final Program

```
method P_3_7_Order (i:int, j: int, k: int) returns (s: string)
  ensures if (i < j < k) || (i > j > k) then s == "in order" else s == "not in order"
  → "
{
  if ((i < j && j < k) || (i > j && j > k)) {
    s := "in order";
  } else {
    s := "not in order";
  }
}
```

```
}  
}
```

Total Token Usage

Input tokens: 346

Output tokens: 134

Reasoning tokens: 64

Sum of ‘total tokens’: 480

Experiment Timings

Overall Experiment started at 1766049596599, ended at 1766049600527, lasting 3928ms (3.93 seconds)

Iteration #1 started at 1766049596600, ended at 1766049600527, lasting 3927ms (3.93 seconds)