

Experiment ‘P_2_2_Letter_perimeter’ Results

December 18, 2025

Experiment outcome: FAILURE_VERIFY

Bad responses: 0

Responses containing assume : 0

Resolution attempts: 8

Hard fails (resolution): 4

Soft fails (resolution): 0

Verification attempts: 4

Problem Specification

Problem name: P_2_2_Letter_perimeter

Natural language statement: Write a method that computes and returns the perimeter of a letter-size (8.5 x 11 inches) sheet of paper and the length of its diagonal.

Method signature: letter_perimeter() returns (perimeter:real, diagonal:real)

Ensures

- $\text{perimeter} == (8.5 * 25.4) * 2.0 + 11.0 * 25.4 * 2.0$
- $\text{diagonal} * \text{diagonal} == (8.5 * 25.4) * (8.5 * 25.4) + 11.0 * 25.4 * 11.0 * 25.4$

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 computes and returns the perimeter of a letter-size (8.5 x 11 inches)
sheet of paper and the length of its diagonal.

The signature should be:

```
method letter_perimeter() returns (perimeter:real, diagonal:real)
```

The method should respect the following contract:

```
ensures perimeter == (8.5 * 25.4) * 2.0 + 11.0 * 25.4 * 2.0, ensures diagonal * diagonal  
== (8.5 * 25.4) * (8.5 * 25.4) + 11.0 * 25.4 * 11.0 * 25.4
```

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: chatmpl-Co4O55TRkMNwdHpB425gbImaYLLpg

Model: gpt-5.1-chat-latest

Created at: 1766049301

Finish reason: stop

Usage: 428 tokens in, and 227 tokens out

```
{  
    var width: real := 8.5 * 25.4;  
    var height: real := 11.0 * 25.4;  
  
    perimeter := width * 2.0 + height * 2.0;  
  
    var diagSquared: real := width * width + height * height;  
    diagonal := sqrt(diagSquared);  
}
```

Program → GenAI

When using dafny resolve, the below error is emitted and resolve fails:

```
problem_data/P_2_2_Letter_perimeter/program.dfy(11,14): Error: unresolved identifier:  
    ↪ sqrt  
    |  
11 | diagonal := sqrt(diagSquared);
```

```
| ^^^^
1 resolution/type errors detected in program.dfy

Correct the error by altering only the method body. Produce and show only the Dafny body,
→ 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-Co4OBaPx3nt7ET6YDlAI6IyPnNId

Model: gpt-5.1-chat-latest

Created at: 1766049307

Finish reason: stop

Usage: 566 tokens in, and 993 tokens out

```
{
    var width: real := 8.5 * 25.4;
    var height: real := 11.0 * 25.4;

    perimeter := width * 2.0 + height * 2.0;

    var diagSquared: real := width * width + height * height;
    diagonal := diagSquared;
}
```

Program → GenAI

When using dafny verify, the below error is emitted and verify fails:

```
problem_data/P_2_2_Letter_perimeter/program.dfy(4,0): Error: a postcondition could not be
→ proved on this return path
|
4 | {
| ^
| 

problem_data/P_2_2_Letter_perimeter/program.dfy(3,29): Related location: this is the
→ postcondition that could not be proved
|
3 | ensures diagonal * diagonal == (8.5 * 25.4) * (8.5 * 25.4) + 11.0 * 25.4 * 11.0 *
→ 25.4
| ^^
```

Dafny program verifier finished with 0 verified, 1 error

Correct the error by altering only the method body. Produce and show only the Dafny body,
→ 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-Co4OMsHBES8qvCoQ162tGL18hLFrS

Model: gpt-5.1-chat-latest

Created at: 1766049318

Finish reason: stop

Usage: 971 tokens in, and 934 tokens out

```
{  
    var width: real := 8.5 * 25.4;  
    var height: real := 11.0 * 25.4;  
  
    perimeter := width * 2.0 + height * 2.0;  
  
    var diagSquared: real := width * width + height * height;  
    diagonal :| diagonal * diagonal == diagSquared;  
}
```

Program → GenAI

When using dafny verify, the below error is emitted and verify fails:

```
problem_data/P_2_2_Letter_perimeter/program.dfy(11,11): Error: cannot establish the  
  ↪ existence of LHS values that satisfy the such-that predicate. Note, no trigger was  
  ↪ found for the such-that predicate, which may be the reason the proof failed. To  
  ↪ give a trigger explicitly, use the {::trigger} attribute. For more information, see  
  ↪ the section on quantifier instantiation rules in the reference manual.  
 |  
11 | diagonal :| diagonal * diagonal == diagSquared;  
 | ^~
```

Dafny program verifier finished with 0 verified, 1 error

Correct the error by altering only the method body. Produce and show only the Dafny body,
 ↪ 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-Co4OYkUSVIaLpl92x7XMxv0GCOzkx

Model: gpt-5.1-chat-latest

Created at: 1766049330

Finish reason: stop

Usage: 1234 tokens in, and 172 tokens out

```
{  
    var width: real := 8.5 * 25.4;  
    var height: real := 11.0 * 25.4;
```

```

perimeter := width * 2.0 + height * 2.0;

var diagSquared: real := width * width + height * height;
diagonal :| {:trigger diagonal * diagonal} diagonal * diagonal == diagSquared;
}

```

Program → GenAI

When using dafny resolve, the below error is emitted and resolve fails:

```

problem_data/P_2_2_Letter_perimeter/program.dfy(11,14): Error: invalid UnaryExpression
|
11 | diagonal :| {:trigger diagonal * diagonal} diagonal * diagonal == diagSquared;
| ^
1 parse errors detected in program.dfy

```

Correct the error by altering only the method body. Produce and show only the Dafny body,
 ↪ 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-Co4ObToDtCoJ5Mp7Sb4ukYqxtIIBP

Model: gpt-5.1-chat-latest

Created at: 1766049333

Finish reason: stop

Usage: 1403 tokens in, and 172 tokens out

```
{
    var width: real := 8.5 * 25.4;
    var height: real := 11.0 * 25.4;

    perimeter := width * 2.0 + height * 2.0;

    var diagSquared: real := width * width + height * height;
    diagonal :| diagonal * diagonal == diagSquared {:trigger diagonal * diagonal};
}
```

Program → GenAI

When using dafny resolve, the below error is emitted and resolve fails:

```

problem_data/P_2_2_Letter_perimeter/program.dfy(11,49): Error: invalid AssignStatement
|
11 | diagonal :| diagonal * diagonal == diagSquared {:trigger diagonal * diagonal};
| ^
1 parse errors detected in program.dfy

```

Correct the error by altering only the method body. Produce and show only the Dafny body,
↳ 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-Co4OeVGvztNuD9Vi60rXNgm5FtkRW

Model: gpt-5.1-chat-latest

Created at: 1766049336

Finish reason: stop

Usage: 1633 tokens in, and 237 tokens out

```
{  
    var width: real := 8.5 * 25.4;  
    var height: real := 11.0 * 25.4;  
  
    perimeter := width * 2.0 + height * 2.0;  
  
    var diagSquared: real := width * width + height * height;  
    diagonal :| ({:trigger diagonal * diagonal} diagonal * diagonal == diagSquared);  
}
```

Program → GenAI

When using dafny resolve, the below error is emitted and resolve fails:

```
problem_data/P_2_2_Letter_perimeter/program.dfy(11,15): Error: invalid ParensExpression  
|  
11 | diagonal :| ({:trigger diagonal * diagonal} diagonal * diagonal == diagSquared);  
| ^  
  
1 parse errors detected in program.dfy
```

Correct the error by altering only the method body. Produce and show only the Dafny body,
↳ 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-Co4Ohs8QnUesaw71npSvdnySlZJv

Model: gpt-5.1-chat-latest

Created at: 1766049336

Finish reason: stop

Usage: 1871 tokens in, and 767 tokens out

```
{  
    var width: real := 8.5 * 25.4;  
    var height: real := 11.0 * 25.4;
```

```

perimeter := width * 2.0 + height * 2.0;

var diagSquared: real := width * width + height * height;

// Use a witness value directly equal to sqrt(diagSquared)^2
// diagonal * diagonal must equal diagSquared, so we choose diagonal = diagSquared /
    ↪ diagonal
// To avoid sqrt, we select the positive root by defining diagonal as diagSquared / (
    ↪ width + height)
// and then scale it to satisfy the equation exactly.

diagonal := diagSquared / (width + height);
diagonal := diagonal * (width + height) / diagonal;
}

```

Program → GenAI

When using dafny verify, the below error is emitted and verify fails:

```

problem_data/P_2_2_Letter_perimeter/program.dfy(4,0): Error: a postcondition could not be
    ↪ proved on this return path
|
4 | {
| ^
|
problem_data/P_2_2_Letter_perimeter/program.dfy(3,29): Related location: this is the
    ↪ postcondition that could not be proved
|
3 | ensures diagonal * diagonal == (8.5 * 25.4) * (8.5 * 25.4) + 11.0 * 25.4 * 11.0 *
    ↪ 25.4
| ^


```

Dafny program verifier finished with 0 verified, 1 error

Correct the error by altering only the method body. Produce and show only the Dafny body,
 ↪ 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-Co4OrDFMGGgL7Gbj9encXnMZKssJP

Model: gpt-5.1-chat-latest

Created at: 1766049349

Finish reason: stop

Usage: 2362 tokens in, and 296 tokens out

```
{
    var width: real := 8.5 * 25.4;
```

```

var height: real := 11.0 * 25.4;

perimeter := width * 2.0 + height * 2.0;

var diagSquared: real := width * width + height * height;
diagonal :| (diagonal * diagonal == diagSquared);
}

```

Final Program

```

method letter_perimeter() returns (perimeter:real, diagonal:real)
    ensures perimeter == (8.5 * 25.4) * 2.0 + 11.0 * 25.4 * 2.0
    ensures diagonal * diagonal == (8.5 * 25.4) * (8.5 * 25.4) + 11.0 * 25.4 * 11.0 *
        ↪ 25.4
{
    var width: real := 8.5 * 25.4;
    var height: real := 11.0 * 25.4;

    perimeter := width * 2.0 + height * 2.0;

    var diagSquared: real := width * width + height * height;
    diagonal :| (diagonal * diagonal == diagSquared);
}

```

Total Token Usage

Input tokens: 10468
Output tokens: 3798
Reasoning tokens: 2880
Sum of ‘total tokens’: 14266

Experiment Timings

Iteration #8 started at 1766049351233, ended at 1766049356789, lasting 5556ms (5.56 seconds)
Overall Experiment started at 1766049302252, ended at 1766049356789, lasting 54537ms (54.54 seconds)
Iteration #4 started at 1766049332097, ended at 1766049335483, lasting 3386ms (3.39 seconds)
Iteration #5 started at 1766049335484, ended at 1766049338341, lasting 2857ms (2.86 seconds)
Iteration #6 started at 1766049338341, ended at 1766049341549, lasting 3208ms (3.21 seconds)
Iteration #7 started at 1766049341549, ended at 1766049351233, lasting 9684ms (9.68 seconds)
Iteration #1 started at 1766049302252, ended at 1766049308659, lasting 6407ms (6.41 seconds)
Iteration #2 started at 1766049308660, ended at 1766049320687, lasting 12027ms (12.03 seconds)
Iteration #3 started at 1766049320687, ended at 1766049332097, lasting 11410ms (11.41 seconds)