Waterproof River: An AI Tutor for Waterproof

Dick Arends Pim Otte Mikołaj Pujanek Jim Portegies

Radboud University Nijmegen

Utrecht University

TU Eindhoven

ThEdu 2026

- **Waterproof**: Educational software aimed at helping students learn to write mathematical proofs.
- **Rocq Proof Assistant**: Provides immediate feedback.
- **Controlled Natural Language**: Flattens the learning curve, better transfer to pen-and-paper style proofs.[1]

```
Lemma exercise_1 : 4 is the supremum of [0, 4).
Proof.
  It suffices to show that
    4 is an _upper bound_ for [0, 4) ∧
    ∀ ε > 0, ∃ a ∈ [0, 4), 4 - ε < a.
  We show both statements.
  - We need to show that ∀ c ∈ [0, 4), c ≤ 4.
    Take c ∈ [0, 4).
    We conclude that c ≤ 4.
  - We need to show that
      ∀ ε > 0, ∃ a ∈ [0, 4), 4 - ε < a.
    Take ε > 0.
    Either ε < 2 or ε ≥ 2.
    + Case ε < 2.
      Choose a := 4 - ε / 2.
      { Indeed, a ∈ [0, 4). }
      We conclude that & 4 - ε < 4 - ε / 2 = a.
    + Case ε ≥ 2.
      Choose a := 3.
      { Indeed, a ∈ [0, 4). }
      We conclude that & 4 - ε < 3 = a.
Qed.
```

[1] Wemmenhove et al., *"Comparative analysis of student proof structure following the use of Waterproof"* (2026).

Waterproof

Standard Rocq

```
Goal forall x : nat, x + 1 = 1 + x.
Proof.
  intro x.
  apply Nat.add_comm.
Qed.
```

Waterproof

```
Goal ∀ x ∈ ℕ, x + 1 = 1 + x.
Proof.
  Take x ∈ ℕ.
  We conclude that x + 1 = 1 + x.
Qed.
```

Standard Rocq

```
Goal forall x : nat, x + 1 = 1 + x.
Proof.
  intro x.
  apply Nat.add_comm.
Qed.
```

Waterproof

```
Goal ∀ x ∈ ℕ, x + 1 = 1 + x.
Proof.
  Take x ∈ ℕ.
  We conclude that x + 1 = 1 + x.
Qed.
```

The Dilemma

Controlled Natural Language flattens the learning curve but is unforgiving.

Rigidity stalls student momentum: Students spend time fixing syntax, not writing proofs.[2]

$\longrightarrow$ **Our Vision: An AI tutor to bridge the gap between rigid syntax and student intuition.**

[2] Wemmenhove et al., *"Service design for the improvement of Intelligent Tutoring Systems: a case study"* (2025).

```
Goal ∀ x ∈ ℕ, x + 1 = 1 + x.
Proof.
take x ∈ ℕ.
We conclude that x + 1 = 1 + x.
```

Syntax error: [ltac2_use_default] expected after [ltac2_expr] (in [ltac2_command]).

The variable x was not found in the current environment.

```
Qed.
```

```
Goal ∀ x ∈ ℕ, x + 1 = 1 + x.
Proof.
take x ∈ ℕ.
We conclude that x + 1 = 1 + x.
```

Syntax error: Unfortunately, this sentence cannot be understood. Check for instance whether all parentheses match and whether all words are spelled correctly. To reduce syntax errors, it can be helpful to input sentences using the autocomplete functionality.

The variable x was not found in the current environment.

```
Qed.
```

Waterproof River

- **Proposed Solution**: Waterproof River, an LLM-based tutor for Waterproof.
- **Virtual Tutor**: Available 24/7, lower barrier to ask.
- **Goal**: Make Waterproof easier to use $\longrightarrow$ improve the proof writing skills of students.

Context & Reception

- **Setting**: Analysis 1 at Eindhoven University of Technology (~ 300 students). Waterproof mandatory, Waterproof River was not.
- **Adoption**: Used by 50-100 students, who rated the development versions *moderately useful*.
- **Experimental**: So far, we leveraged student input for active development, but have not conducted any systematic educational study yet.

Key Features

- **Tight Integration**: Interaction with the Rocq proof assistant using the 'rocq-lsp' language server[3] and its Pétangue[4] plugin.
- **Embedding**: River is available in the same interface as the normal Waterproof interface.
- **Opt-in**: Students can opt-in to use River.

[3] Gallego Arias et al., *Rocq-lsp: Visual Studio Code Extension and Language Server Protocol for Rocq* (2025).

[4] Teodorescu et al., *"NLIR: Natural Language Intermediate Representation for Mechanized Theorem Proving"* (2024).

Related Technologies

- **HaLLMos**[5]: Gives AI-based feedback on typed natural language proofs.
  River is backed by Rocq for verification.
- **LeanTutor**[6]: Checks and helps students with natural language mathematical proofs.
  Interaction with the student is through natural language. Uses Lean to verify student proof through auto-formalization and provide hints on incomplete proofs.

[5] `hallmos.com`

[6] Patel et al., *LeanTutor: Towards a Verified AI Mathematical Proof Tutor* (2026).

Modes of Interaction

1. Chat Conversations
   - **Socratic**: River is instructed to guide by asking socratic questions instead of answering directly.
   - **Frustration Detection**: The model is allowed to become more helpful the longer the student is stuck.

Modes of Interaction

1. Chat Conversations

- **Socratic**: River is instructed to guide by asking socratic questions instead of answering directly.
- **Frustration Detection**: The model is allowed to become more helpful the longer the student is stuck.

2. Commands

- `/syntaxHelp`: Helps with the syntax error on the current line.
- `/hint`: Generate a *verified* hint for the next step in a proof.
- `/translateProof`: Translate a 'handwritten proof' into valid Waterproof.
- Highly reproducible, trigger specific actions; valuable for debugging.

Chat Conversation

Editor

```
Lemma exercise : 4 is the supremum of [0, 4).
Proof.
   It suffices to show that
     4 is an upper bound for [0, 4) ∧
     (∀ ε > 0, ∃ a ∈ [0, 4), 4 - ε < a).
   We show both statements.
   - We conclude that 4 is an upper bound for [0, 4).
   - We need to show that ∀ ε > 0, ∃ a ∈ [0, 4), 4 - ε < a.
     Take ε > 0.
     Choose a := 4 - ε / 2.

Qed.
```

Chat

@River Could you help me with my proof?

River

Proof context

It looks like you are making great progress! You have successfully split the supremum definition into two parts, proved the first one, and just made a choice for `a` in the second part.

When you use the `Choose` tactic to pick an element from a specific set (in this case, `[0, 4)`), what is the immediate next step you need to take according to the proof framework? (Hint: The Waterproof system actually gives you a very strong hint about this directly in your goal state!)

Tools

- To really work as a tutor, River needs to be *context aware*.
- LLM receives prompt including tutorial file, lecture notes paragraph, few-shot examples, and question.
  However, the answer might also depend on state of the proof.

Tools

- To really work as a tutor, River needs to be *context aware*.
- LLM receives prompt including tutorial file, lecture notes paragraph, few-shot examples, and question.
  However, the answer might also depend on state of the proof.
- **Proof Context**: "Returns the goals and proof script at the place of the user cursor. This tool should be used when the user mentions the current proof in a follow up question and the answer is dependent on the current proof state."

Basic Prompt

```
You are River, an encouraging, empathetic, socratic instructor for an introductory
mathematics course that uses the Waterproof language in Rocq to formally state and
prove mathematics.

Your aim is to help students with writing statements by asking them questions.
To get information on the Waterproof language, please see the tutorial:
...

An overview of the Waterproof tactics is supplied here:
...

For inspiration for the tone, how we communicate about certain proof steps, and on
what we value in this course, please have a careful look at
...
```

Features

With suitable tools and prompts, what can we achieve?

- Help with syntax.
- Provide hints.
- Translate a proof.
- Help with mathematical theory.

Feature: Help with Syntax

```
Goal ∀ x ∈ ℕ, x + 1 = 1 + x.
Proof.
take x ∈ ℕ.
We conclude that x + 1 = 1 + x.
```

Syntax error: Unfortunately, this sentence cannot be understood. Check for instance whether all parentheses match and whether all words are spelled correctly. To reduce syntax errors, it can be helpful to input sentences using the autocomplete functionality.

River: "Syntax error: tactic names are case-sensitive; use `Take` instead of `take`."

Replace ↩ Take x ∈ ℕ.

```
Qed.
```

Feature: Provide Hints

- **Idea**: Already have Waterproof `Help.` command that gives a hint on the structure of the proof $\longrightarrow$ extend this to hints like an instructor.

```
Lemma exercise_3_11_4 :
  (∀ M1 ∈ ℝ, ∃ y > 0, ∀ x > y, f (x) > M1) ⇒
    (∀ M2 < 4, ∃ z > 0, ∀ x > z, g (x) > M2) ⇒
      ∃ c ∈ ℝ, f(c) + g(c) > 10.
Proof.
  Assume that ∀ M1 ∈ ℝ, ∃ y > 0,
    ∀ x > y, f(x) > M1 as (A).
  Assume that ∀ M2 < 4, ∃ z > 0,
    ∀ x > z, g(x) > M2 as (B).
  Use M1 := 10 in (A).
  { Indeed, 10 ∈ ℝ. }
  It holds that ∃ y > 0, ∀ x > y, f(x) > 10.
  Help.
```

📋 The goal is to show a 'there exists'-statement (∃).
Choose a specific variable in ℝ.

```
Lemma exercise_3_11_4 :
  (∀ M1 ∈ ℝ, ∃ y > 0, ∀ x > y, f (x) > M1) ⇒
    (∀ M2 < 4, ∃ z > 0, ∀ x > z, g (x) > M2) ⇒
      ∃ c ∈ ℝ, f(c) + g(c) > 10.
Proof.
  Assume that ∀ M1 ∈ ℝ, ∃ y > 0,
    ∀ x > y, f(x) > M1 as (A).
  Assume that ∀ M2 < 4, ∃ z > 0,
    ∀ x > z, g(x) > M2 as (B).
  Use M1 := 10 in (A).
  { Indeed, 10 ∈ ℝ. }
  It holds that ∃ y > 0, ∀ x > y, f(x) > 10.
  Help.
```

📋 The goal is to show a 'there exists'-statement (∃). Choose a specific variable in ℝ.

@River Can you give me a hint for this proof?

River

Proof context

Result of verified hint generation

Nice step — you already derived

- `∃ y > 0, ∀ x > y, f(x) > 10`.

A useful next question: how do you **use** an existential hypothesis in Waterproof?

Try extracting its witness first:

```
Obtain y according to (_H1).
```

Hint Generation: Pétanque

Proof

```
Lemma exercise :
  ∀ x ∈ ℝ, 0 * x = 0 ∧ x + 1 > x.
Proof.
........................
Qed.
```

- Thanks to Emilio J. Gallego Arias for all the help!
- Completely hidden from the student.

Pétanque

Goal

`∀ x ∈ ℝ, 0 * x = 0 ∧ x + 1 > x`

Proof so far

`........................`

Step to try

`Let x ∈ ℝ`

Output

Syntax Error

Hint Generation: Pétanque

Proof

```
Lemma exercise :
  ∀ x ∈ ℝ, 0 * x = 0 ∧ x + 1 > x.
Proof.
........................
Qed.
```

- Thanks to Emilio J. Gallego Arias for all the help!
- Completely hidden from the student.

Pétanque

Goal

`∀ x ∈ ℝ, 0 * x = 0 ∧ x + 1 > x`

Proof so far

`........................`

Step to try

`Take x ∈ ℝ.`

Output

Success

Remaining Goal: `0 * x = 0 ∧ x + 1 > x`

Hint Generation: Pétanque

Proof

```
Lemma exercise :
  ∀ x ∈ ℝ, 0 * x = 0 ∧ x + 1 > x.
Proof.
Take x ∈ ℝ.
........................
Qed.
```

Pétanque

Goal

`0 * x = 0 ∧ x + 1 > x`

Proof so far

`Take x ∈ ℝ.`

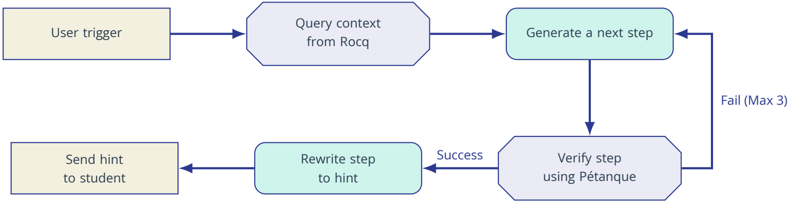
Step to try

`We conclude that 0 * x = 0 ∧ x + 1 > x.`

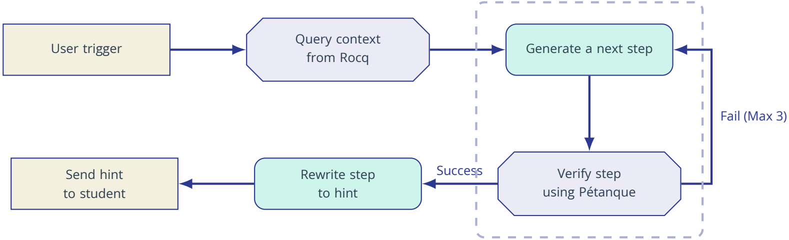
Output

Proof finished (no goals left)

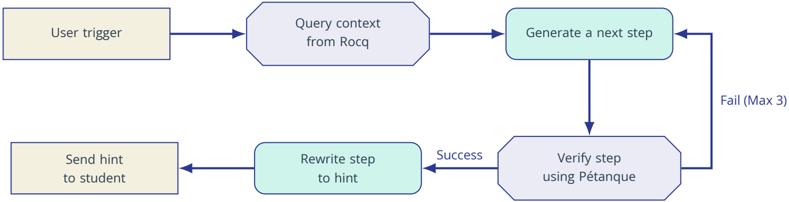
Hint Generation: Complete Pipeline



Hint Generation: Complete Pipeline



Hint Generation: Complete Pipeline



Double-Prompt Strategy

Separate the **Step Generation** from **Hint Formulation**. River guarantees correctness without delivering a copy-paste answer to student.

Feature: Translate a Proof

- **Idea**: Students might have the correct idea for a proof, but don't know how to enter it in Waterproof.
- **Verified**: The translated proof is verified in the background using Rocq.

```
Variable f : ℝ → ℝ.
Variable g : ℝ → ℝ.

Lemma exercise :
  (∃ M1 ∈ ℝ, ∀ x ∈ ℝ, f (x) > M1) ⇒
    ((∃ M2 < 4, ∀ x ∈ ℝ, g (x) > M2) ⇒
      ∃ M3 ∈ ℝ, ∀ x ∈ ℝ, f(x) + g(x) > M3).
Proof.
Assume there exists M1 in ℝ
  such that for all x in ℝ, f(x) > M1.
Assume as well that there is an M2 < 4,
  such that for all real x, g(x) > M2.
Use M1 + M2 for M3.
Qed.
```

@River /translateProof

River

To translate your ideas into Waterproof, you can use the `Assume that` tactic for both of your premises since they are implications in the goal. Be sure to give them a label, such as `(h1)` and `(h2)`. Because these assumptions start with an existence (∃) quantifier, you need to extract the variables using the `Obtain` tactic before you can do arithmetic with them. Finally, to supply the value for M_3, use the `Choose` tactic.

```
Assume that ∃ M1 ∈ ℝ, ∀ x ∈ ℝ, f(x) > M1 as (h1).
Obtain M1 according to (h1).
Assume that ∃ M2 < 4, ∀ x ∈ ℝ, g(x) > M2 as (h2).
Obtain M2 according to (h2).
Choose M3 := M1 + M2.
(* Verify that M3 is in ℝ and complete the proof *)
```

Feature: Help with Mathematical Theory

- **Idea**: Modern LLMs can provide extensive mathematical knowledge, but don't always present it in line with a curriculum or a course.

River

@River What is the definition of continuity?

River

Snippet from the lecture notes

Great question.

In this course, continuity at a point is defined as:

$$f \text{ is continuous in } a \iff \lim_{x \to a} f(x) = f(a).$$

And the practical ϵ–δ form is:

$$\forall \epsilon > 0,\ \exists \delta > 0,\ \forall x \in D,\ \left(0 < \mathrm{dist}_X(x,a) < \delta \Rightarrow \mathrm{dist}_Y(f(x), f(a)) < \epsilon\right).$$

For real-valued functions on $\mathbb{R}$ (usual distance), this is usually written as:

$$\forall \epsilon > 0,\ \exists \delta > 0,\ \forall x \in D,\ \left(0 < |x-a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon\right).$$

A function is **continuous (on its domain)** if it is continuous at every point of that domain.

ChatGPT

What is the definition of continuity?

The definition of **continuity** depends on the context:

- **General meaning:** Continuity is the state of something continuing over time without interruption or significant change.
  - Example: "The company ensured continuity of service during the power outage."
- **In mathematics (functions):** A function is **continuous** at a point if small changes in the input produce small changes in the output, with no jumps, holes, or breaks in the graph. More formally, a function $f(x)$ is continuous at $x = a$ if:

Challenges & Limitations

- **Non-Reproducibility**: The output from the LLM is inherently random.
  Difficult to debug and impossible to guarantee output quality.
- **Model Differences**: Larger, and hence more expensive, models catch more errors.
  Students could choose to pay for the more expensive model, which could lead to inequality.
- **Wrong Explanations**: The model provides the correct fix, but fails to properly explain the underlying problem.
  Models frequently claim a tactic needs an upper-case letter when the real error is elsewhere.
- **(Rapidly) Changing LLM Field**: Models and technology are replaced rapidly.
  Prompt that worked well for the former model, may not work as well for the newer one.
  No free models in VS Code anymore.

Outlook & Next Steps

- **Method of Interaction**: Investigate how students should interact with River.
  Chat interaction useful for programming contexts, but does shift the work from writing in Waterproof to chatting with River.
- **Proactive Assistance**: Provide student with assistance independently of whether they ask for it.
- **Fine-Tuning**: Fine-tune a lightweight, local model specifically with Waterproof River in mind.

Conclusion

- **Waterproof River**: Successfully bridges the strict notation of Waterproof with the forgiving, free-flow communication of natural language.
- **Lessons Learned**: Building an AI tutor is not just about creating a wrapper around an LLM. It requires tight integration with the system and careful design.
- **Classroom**: Already proven useful in Eindhoven for helping students with syntax errors and allowing them to focus on math.
- **Future**: Move towards smaller, local models to guarantee privacy, sustainability, and complete pedagogical control.

References

[1] J. Wemmenhove, N. Hoofd, A. Schüler-Meyer, and J. Portegies, "Comparative analysis of student proof structure following the use of waterproof," in *ThEdu '25 proceedings, To Appear*, 2026.

[2] J. Wemmenhove, D. Bór, R. Conijn, and J. Portegies, "Service design for the improvement of intelligent tutoring systems: A case study," *Journal of Computer Assisted Learning*, vol. 41, no. 4, e70088, 2025.

[3] E. J. Gallego Arias and rocq-lsp contributors, *Rocq-lsp: Visual studio code extension and language server protocol for rocq*, version v0.2.5, GitHub repository, 2025. Accessed: Nov. 30, 2025. [Online]. Available: `https://github.com/ejgallego/rocq-lsp`.

[4] L. Teodorescu, G. Baudart, E. J. G. Arias, and M. Lelarge, "Nlir: Natural language intermediate representation for mechanized theorem proving," in *The 4th Workshop on Mathematical Reasoning and AI at NeurIPS'24*, 2024.

[5] M. Patel et al., *Leantutor: Towards a verified ai mathematical proof tutor*, 2026. arXiv: `2506.08321 [cs.AI]`. [Online]. Available: `https://arxiv.org/abs/2506.08321`.