Teaching Mathematical Proof in the Age of Generative AI: Toward AI-Supported Proof Inquiry

Weng Kin Ho
Nanyang Technological University
wengkin.ho@nie.edu.sg

Recent advances in generative artificial intelligence (Gen AI) have enabled students to obtain mathematical solutions and proofs almost instantly, raising important questions about the teaching of proof in school mathematics. Drawing on research on proof in mathematics education, this paper argues that proof learning should extend beyond proof production toward students' abilities to interpret, critique, compare, and explain mathematical reasoning, including AI-generated proofs. The paper proposes AI-Supported Proof Inquiry (ASPI) as a pedagogical framework for supporting inquiry-oriented engagement with proof in AI-rich learning environments.

Introduction

The rapid development of generative artificial intelligence (AI) is beginning to transform many aspects of education, including the learning and teaching of mathematics. Contemporary AI systems are increasingly capable of producing symbolic manipulations, solving mathematical problems, explaining concepts, and generating mathematical proofs in response to natural language prompts. As generative AI tools such as ChatGPT, Gemini, Claude, Deep Seek, etc., become widely accessible, students can obtain step-by-step mathematical explanations and arguments with minimal effort, raising important questions about the nature of mathematical learning in the classroom. Recent research in mathematics education has begun to examine how students interact with generative AI when solving and proving mathematical statements, highlighting both new learning possibilities and potential challenges for instruction (Yoon et al., 2024).

One area where the impact of generative AI is particularly significant is the teaching of mathematical proof. Proof has long been regarded as a central component of mathematical practice and a key element of mathematical thinking (Hanna, 2000; Hanna & de Villiers, 2008; Stylianides, 2007). Within mathematics education, reasoning and proof are widely recognised as fundamental aspects of students' mathematical learning and understanding (Hanna & de Villiers, 2008). These multiple roles of proof have been widely discussed in mathematics education research (de Villiers, 1990; Hanna & de Villiers, 2008; Stylianides, 2007).

However, the emergence of generative AI capable of producing mathematical arguments challenges some of the assumptions underlying traditional approaches to teaching proof. If students can obtain a complete proof almost instantly from an AI system, the educational value of simply producing a proof may be diminished. Studies examining the integration of generative AI in mathematics learning environments have noted that while such systems can assist students in constructing mathematical arguments, they also raise concerns about overreliance on automated solutions and the need for instructional designs that promote critical engagement with AI-generated outputs (Park & Manley, 2024).

In this paper, we argue that the increasing availability of generative AI calls for a reconsideration of proof instruction in secondary mathematics education. Rather than focusing primarily on students' independent production of proofs, instruction may place greater emphasis on interpreting, critiquing, comparing, and explaining mathematical reasoning, including AI-generated arguments. At the same time, generative AI creates opportunities for students to explore alternative proof strategies and engage more deeply with proof as reasoning,

(2026). In N. Berger, M. Thompson, J. Ley & C. Wilsmore (Eds.), *Beyond boundaries: Strengthening futures in mathematics education. Proceedings of the 48th annual conference of the Mathematics Education Research Group of Australasia, Sydney* (pp. 202–209). MERGA.

exploration, and communication. Drawing on research on proof in mathematics education, this paper proposes AI-Supported Proof Inquiry (ASPI) as a pedagogical framework for inquiry-oriented engagement with proof in AI-rich learning environments.

The Role of Proof in School Mathematics

Mathematical proof has long been regarded as a defining feature of mathematical activity and an essential component of mathematical reasoning. Within mathematics education, proof plays a central role in helping students understand why mathematical statements are true and in developing their ability to reason logically (Stylianides, 2007; Hanna, 2000). Consequently, the teaching and learning of proof has been the focus of extensive research in mathematics education. Research has shown that proof serves several important functions in mathematical practice. In a widely cited framework, de Villiers (1990) identified multiple roles of proof in mathematics, including verification, explanation, systematisation, discovery, and communication. Verification refers to establishing the truth of a mathematical statement, while explanation concerns providing insight into why the statement holds. Proof can also contribute to the organisation of mathematical relationships, support the discovery of new mathematical insights, encourage intellectual challenge, and serve as a means of communicating mathematical reasoning.

Although these functions are all important in mathematical practice, research suggests that proof in school mathematics is often presented primarily as a means of verification (Hanna, 2000; Stylianides, 2007). Students were frequently asked to produce products that confirm the correctness of a result, while other roles of proof – such as explanation, exploration, and comparison of different arguments – may receive less emphasis in classroom instruction. Verification remains important in school mathematics, but engaging students with multiple proofs may deepen understanding and reveal different mathematical structures (Hanna & de Villiers, 2008).

In this context, the emergence of generative AI presents a potentially useful development for proof instruction by making multiple mathematical arguments and alternative proof strategies more readily accessible for classroom exploration. Because AI systems can readily produce proofs of many standard results in school mathematics, classroom activities involving proof may increasingly focus not only on constructing proofs independently, but also on analysing the validity of arguments, comparing alternative strategies, and explaining the reasoning underlying mathematical claims. In this way, generative AI may create new opportunities for students to engage with proofs in ways that more fully reflect its multiple functions in mathematical practice.

Generative AI as a Mathematical Partner in Proof Learning

Recent advances in generative artificial intelligence have introduced new possibilities for engaging with mathematical reasoning in education settings. Large language models (LLMs) are increasingly capable of producing symbolic manipulations, solving mathematical problems, and generating proofs in response to natural language prompts. As a result, students now have unprecedented access to step-by-step proofs of many results encountered in school mathematics.

From the perspective of mathematics education, the availability of AI-generated proofs raises important pedagogical questions. If students can obtain complete proofs with minimal effort through generative AI, classroom activities that focus primarily on reproducing standard proofs may no longer provide insight into students' understanding of mathematical reasoning. However, rather than viewing generative AI solely as a challenge to proof instruction, it may also be understood as a mathematical partner that can support students' engagement with mathematical reasoning. By generating alternative arguments and proof strategies, generative AI may support students in analysing, comparing, and critiquing mathematical reasoning.

To illustrate the type of reasoning that students may encounter when interacting with generative AI, consider the following response produced by an AI system to a prompt requesting a proof.

Figure 1

A Sample Proof of a Mathematical Statement Generated by ChatGPT

Example Transcript

Prompt

Prove that $n^2 - n$ is always even for all integers n.

AI-generated response

We want to show that the expression

$$n^2 - n$$

is always even.

First factor the expression:

$$n^2 - n = n(n-1).$$

This is the product of two consecutive integers. Since consecutive integers differ by 1, one of them must be even. Therefore the product $n(n-1)$ must be even. Hence $n^2 - n$ is always even.

Although the argument appears convincing, it contains a step requiring further justification regarding why one of two consecutive integers must be even. Examining such examples can help students reflect on the structure and completeness of mathematical reasoning.

Interactions of this kind illustrate how generative AI may support richer engagement with proof. Rather than simply copying an AI-generated argument, students can analyse its reasoning, identify missing steps, and revise the argument to make it more complete. In addition, students may prompt the AI system to generate alternative approaches to the same problem, allowing them to compare different proof strategies and reflect on which argument offers clearer explanations of the underlying mathematics. Such interactions engage students with several functions of proof identified by de Villiers (1990), including verification, explanation, discovery, and communication.

A useful analogy may be drawn with the impact of dynamic geometry software on geometry education. Just as tools such as GeoGebra supported students' exploration of geometric relationships and conjectures (Laborde, 2001), generative AI may support proof learning by enabling students to generate, compare, and analyse alternative mathematical arguments.

At the same time, the availability of AI-generated proofs also raises important considerations for assessment. Traditional proof tasks often ask students to construct complete proofs independently. When generative AI systems can readily produce such arguments, assessment practices may need to evolve to capture students' understanding of mathematical reasoning more effectively. For example, tasks that ask students to interpret, critique, or compare proofs may provide richer insights into students' reasoning than tasks that focus solely on proof production.

At the same time, the integration of generative AI into proof learning also presents important challenges. AI-generated proofs may contain incomplete reasoning, hidden assumptions, or mathematically incorrect steps that appear superficially convincing. Students may also become over-reliant on AI-generated arguments without critically evaluating their validity. These concerns highlight the importance of designing classroom activities that support students in analysing and critiquing mathematical reasoning rather than accepting AI-generated outputs uncritically.

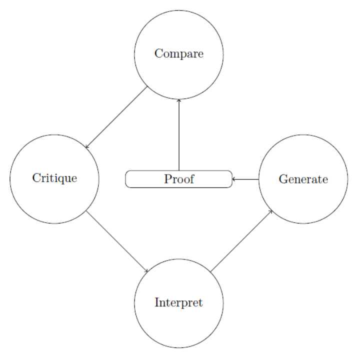
These developments suggest that generative AI may provide additional opportunities for students to engage with mathematical arguments in inquiry-oriented ways that are already valued in mathematics classrooms, such as analysing, comparing, and critiquing proofs. While such pedagogical practices are not unique to AI-supported environments, generative AI may increase students' access to multiple candidate arguments and alternative proof strategies. Further empirical research is needed to examine how these interactions influence students' proof learning in classroom settings.

AI-Supported Proof Inquiry (ASPI)

The discussion in the previous section suggests that generative AI has the potential to reshape how students engage with mathematical proof. Rather than focusing exclusively on the production of proofs, classroom activities may increasingly involve analysing, comparing, and refining mathematical arguments. To conceptualise how such interactions might be structured pedagogically, this paper proposes the framework of *AI-Supported Proof Inquiry (ASPI)*.

Figure 2

ASPI Conceptual Framework



The ASPI framework conceptualises proof learning as an inquiry process in which students engage with mathematical arguments through four complementary practices: *generate*, *interpret*, *compare*, and *critique* (Figure 2). Although the four practices are represented visually with directional arrows, the ASPI framework is not intended to describe a fixed instructional sequence. Rather, the practices of generating, interpreting, comparing, and critiquing proofs are conceptualised as interconnected and iterative forms of engagement with mathematical reasoning. Depending on the task and classroom context, students may move flexibly between these practices rather than progressing through them linearly. Generative AI serves as a tool that supports these practices by providing candidate proofs, suggesting alternative reasoning strategies, and enabling students to examine different forms of mathematical argumentation. The framework serves as both a pedagogical proposal and a conceptual model for future research on proof learning in AI-supported environments.

At the centre of the framework is *proof* itself as an object of inquiry. Students interact with proofs in multiple ways, moving between the practices of generating, interpreting, comparing, and critiquing mathematical arguments. Rather than viewing proof solely as a product to be produced, the ASPI framework emphasises proof as a dynamic process of reasoning and explanation. The four practices involve progressively deeper engagement with mathematical reasoning, from generating arguments to evaluating their validity and explanatory power.

The four components of ASPI are described below.

Generate

The first component of the framework involves generating candidate proofs or proof ideas though interaction with generative AI tools. In traditional classroom settings, students typically attempt to construct proofs independently before receiving feedback from teachers or textbooks. With generative AI, students can prompt an AI system to produce a proof or suggest a possible proof strategy for approaching a problem.

Within the ASPI framework, the purpose of generation is not merely to obtain an answer but to provide a starting point for inquiry. AI-generated proofs serve as artefacts that students can analyse and discuss. In some cases, the generated proof may be correct; in others, it may contain gaps or incomplete reasoning. Both situations provide valuable opportunities for learning. The ability to quickly generate candidate arguments also allows students to explore multiple approaches to a problem, supporting the discovery function of proof identified in mathematics education research (de Villiers, 1990).

Interpret

The second component involves interpreting the reasoning within a proof. Students examine the structure of an argument, identify the key logical steps, and explain why each step is valid. This practice emphasises understanding the meaning of a proof rather than simply verifying its correctness.

Interpretation is closely related to the explanatory role of proof discussed in the literature (Hanna, 2000). A proof that is technically correct may still fail to provide insight into why a result holds. By asking students to explain the reasoning behind each step of a proof, teachers can encourage deeper engagement with the mathematical ideas involved. Generative AI can support this process by providing explanations or alternative formulations that students can examine and discuss.

Compare

The third component of ASPI involves comparing different proofs of the same mathematical statement. In mathematics, it is common for a single result to admit multiple proofs, each highlighting different aspects of the underlying structure. For example, an algebraic identity might be proved using symbolic manipulation, geometric reasoning, or mathematical induction.

Generative AI makes it relatively easy to obtain multiple proofs of the same result. Students can prompt the system to "give another proof" or "explain the result in a different way," producing alternative arguments that can be analysed and compared. Through such comparisons, students may gain insight into which proofs provide clearer explanations, reveal deeper structure, or generalise more readily to other contexts. This practice reflects the explanatory and systematising functions of proof described by de Villiers (1990).

Critique

This final component of the framework involves critiquing mathematical arguments. Students evaluate whether a proof is logically sound, identify gaps in reasoning, and suggest improvements to the argument. This practice emphasises proof as a process of verification and validation.

AI-generated proofs can provide particularly rich opportunities for critique. Although such systems often produce plausible reasoning, their arguments may occasionally contain subtle errors, incomplete justifications, or ambiguous explanations. When students examine these proofs critically, they develop the ability to assess mathematical arguments independently rather than relying solely on external authority. This practice aligns with the verification function of proof while also promoting mathematical judgement.

Taken together, the four practices of generating, interpreting, comparing, and critiquing proofs represent a shift in emphasis from proof production toward *proof inquiry*. Generative AI supports this shift by enabling students to interact with mathematical arguments in multiple ways and by providing a flexible environment for exploring alternative reasoning strategies. In the next section, several classroom task designs are presented to illustrate how the ASPI framework may be implemented in secondary mathematics classrooms.

Classroom Task Designs Using the ASPI Framework

To illustrate how the ASPI framework can be implemented in classroom practice, this section presents several example task designs suitable for secondary mathematics classrooms. Each task engages students with mathematical proofs through different stages of AI-Supported Proof Inquiry (ASPI). These tasks are intended to demonstrate how generative AI can support students in generating, interpreting, comparing, and critiquing mathematical arguments.

Task 1: Exploring Alternative Proofs

Consider the identity $1 + 3 + 5 + \cdots + (2\mathrm{n} - 1) = n^2$. Students are asked to prompt a generative AI system to prove this result.

Generate. Students request a proof from the AI system. The AI may produce a proof by mathematical induction showing that if the identity holds for $n = k$, it also holds for $n = k + 1$.

Interpret. Students examine the proof and explain the reasoning behind the inductive step, discussing how the expression $k^2 + (2k + 1)$ leads to $(k + 1)^2$.

Compare. Students then prompt the AI system for another proof. The AI may produce a geometric argument showing that successive odd numbers correspond to adding layers of dots that build a square of side length n. Students compare the two arguments and discuss which proof provides a clearer explanation of why the identity holds.

Through this activity, students encounter multiple proofs of the same statement and reflect on how different arguments reveal different mathematical insights.

Task 2: Comparing Proofs

Students consider the Pythagorean theorem and examine two proofs: one obtained from a textbook using similar triangles and another generated by an AI system based on geometric dissection.

Interpret. Students analyse the reasoning in each proof, identifying the key mathematical ideas involved.

Compare. Students discuss how the two proofs differ in structure and explanatory power. For example, one proof may emphasise geometric similarity while another highlights area relationships. Students then reflect on which proof helps them understand the theorem more clearly and why.

Such activities encourage students to view proofs not merely as formal procedures but as explanations that illuminate mathematical relationships.

Task 3: Critiquing AI-Generated Proofs

Students are presented with the following AI-generated proof of the statement:

"If n^2 is even, then n is even."

Generate. A sample proof generated by ChatGPT is obtained from a prompt for a proof.

Figure 3

A Sample Proof of a Mathematical Statement

AI-generated proof

Suppose n^2 is even. Then we can write

$$n^2 = 2k$$

for some integer k. Taking square roots on both sides gives

$$n = \sqrt{2k}.$$

Since $2k$ is divisible by 2, the square root must also be even. Therefore n is even.

Interpret. Students first explain the reasoning used in the argument. They identify the key steps of the proof and describe how the conclusion is obtained from the assumption.

Critique. Students are then asked to determine whether the argument is valid. In particular, they examine the step where the square root of $2k$ is taken. Students discuss why this reasoning is problematic and identify the flaw in the argument. Students are then invited to construct a correct proof. One possible approach is to use the contrapositive:

If n is odd, then $n = 2m + 1$ for some integer m. Then

$$n^2 = (2m + 1)^2 = 4m^2 + 4m + 1 = 2(2m^2 + 2m) + 1,$$

which is odd. Therefore, if n^2 is even, then n is even.

Through this activity, students practise evaluating the validity of mathematical arguments and recognising that an apparently plausible proof may contain incorrect reasoning. Such critique tasks encourage students to engage critically with mathematical explanations, particularly when proofs are generated by AI systems.

Discussion and Implications

The examples presented in the previous section illustrate how generative AI can support richer forms of engagement with mathematical proof. Rather than viewing AI primarily as a tool that replaces students' work, the activities demonstrate how AI-generated arguments can become objects for inquiry and discussion in the classroom. Through the practices of generating, interpreting, comparing, and critiquing proofs, students interact with mathematical reasoning in ways that extend beyond the traditional emphasis on proof production.

From this perspective, the presence of generative AI invites a shift in proof pedagogy from a production-oriented approach toward a more inquiry-oriented approach. In many traditional settings, proof tasks are framed primarily as exercises in constructing correct arguments. While such activities remain important, the availability of AI-generated proof means that students can now readily access candidate arguments. This development creates opportunities for tasks that emphasise analysing, evaluating, and refining mathematical reasoning.

This shift also has implications for assessment practices. If students can easily obtain complete proofs from generative AI systems, assessments that focus solely on proof production may not adequately capture students' understanding of mathematical reasoning. Alternative forms of assessment may therefore become increasingly relevant. For example, students may be asked to interpret the reasoning within a proof, compare different proofs of the same statement, or identify and correct flaws in an argument. Such tasks provide insights into students' ability to reason about mathematical arguments rather than simply reproduce them.

More broadly, the ASPI framework highlights the potential of generative AI to function as a tool for mathematical exploration. By prompting AI systems to produce alternative proofs or explanations, students can encounter multiple ways of reasoning about the same mathematical idea. These experiences may deepen students' understanding of the explanatory role of proof and help them appreciate the diversity of mathematical arguments.

At the same time, the integration of generative AI into proof learning raises new questions for mathematics education research. Further studies may investigate how students interact with AI-generated proofs, how such interactions influence students' understanding of mathematical reasoning, and how teachers can design tasks that effectively integrate AI into proof instruction. In this way, the ASPI framework may serve not only as a pedagogical proposal but also as a conceptual lens for examining proof learning in AI-supported environments.

Conclusion

Generative artificial intelligence is rapidly becoming part of the technological landscape in which mathematics education takes place. As AI systems become increasingly capable of producing mathematical arguments, educators face the challenge of reconsidering how proof is taught and learned in school mathematics. Rather than viewing generative AI solely as a threat to traditional proof instruction, this paper has argued that AI can serve as a productive partner in proof learning.

Drawing on research on the functions of proof in mathematics education, the paper proposed the AI-Supported Proof Inquiry (ASPI) framework as a way of conceptualising students' interactions with proofs in AI-supported learning environments. The framework emphasises four practices – generate, interpret, compare, and critique – through which students can engage with mathematical arguments as objects of inquiry.

The classroom task designs presented in this paper illustrate how generative AI can support these practices within the context of secondary mathematics. By interacting with AI-generated proofs, students can explore alternative arguments, reflect on different proof strategies, and develop the ability to evaluate the validity and explanatory power of mathematical reasoning.

References

de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras, 24*, 17-24.

Hanna, G. (2000). Proof, explanation and exploration. An overview. *Educational Studies in Mathematics, 44*(1-2), 5-23. https://doi.org/10.1023/A:1012737223465

Hanna, G., & de Villiers, M. (2008). ICMI Study 19: Proof and proving in mathematics education. *ZDM Mathematics Education 40*(2): 329-336. https://doi.org/10.1007/s11858-008-0073-4

Park, H., & Manley, E. D. (2024). Using ChatGPT as a proof assistant in mathematics pathways course. *The Mathematical Education, 63*(2), 139-163. https://doi.org/10.7468/mathedu.2024.63.2.139

Laborde, C. (2001). Integration of technology in the design of geometry tasks with Cabri-Geometry. *International Journal of Computers for Mathematical Learning, 6*(3), 283-317. https://doi.org/10.1023/A:1013309728825

Stylianides, G. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*(3), 289-321.

Yoon, H., Hwang, J., Lee, K., Roh, K. H., & Kwon, O. N. (2024). Students' use of generative AI for proving mathematical statements. *ZDM Mathematics Education 56*, 1531-1551. https://doi.org/10.1007/s11858-024-01629-0