STIAEM

Slow Tech Innovation

AI, Ethics and New Teaching Methods

Applying Chain-of-Thought Techniques to Display Intermediate Logical Steps

Co-funded by the European Union

2025-1-IT02-KA220-SCH-000365213

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them... Project reference code: 2025-1-IT02-KA220-SCH-000365213

Università di Foggia

Studi Umanistici, Lettere, Beni Culturali, Scienze della Formazione

İSTANBUL - ESENLER İLÇE MİLLÎ EĞİTİM MÜDÜRLÜĞÜ

1474



STIAEM

Slow Tech Innovation AI, Ethics and New Teaching Methods



Co-funded by the European Union

2025-1-IT02-KA220-SCH-000365213

eduOpen



CC BY-NC-SA 4.0 International
This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only. If others modify or adapt the material, they must license the modified material under identical terms.



Università di Foggia
Studi Umanistici, Lettere, Beni Culturali, Scienze della Formazione

















OER 4.2.2: Applying Chain-of-Thought Techniques to Display Intermediate Logical Steps
Reference Module: Module 4 - Prompt Engineering for Inclusive Education
Related Video-Pill: Video 4.2: Advanced Prompting & Chain-of-Thought
Authoring partner: Rhodes Evening High School - Greece
Topic / Thematic Focus: Chain-of-Thought techniques, intermediate logical steps and step-by-step reasoning

Introduction and Learning Objectives

This resource complements, on the methodological level, the pathway begun with OER 4.2.1 and accompanies video-pill 4.2 'Advanced Prompting and Chain-of-Thought'. If the previous resource showed how to stabilise the output with worked examples, this one focuses on the techniques that make the machine externalise the intermediate steps of reasoning: Chain-of-Thought and its developments, self-consistency and trees of thought.
The didactic objective is clear and ambitious: to make the process observable, not just the result. It is the same principle by which, at school, the correct answer is not enough: the student is asked to show the procedure. Applied to AI, this principle transforms the tool from an 'opaque generator of answers' into an 'object of methodological evaluation'. The text links to DigComp 3.0 (Cosgrove & Cachia, 2025) and to the 'AI pedagogy' pillar of the UNESCO Framework for teachers (Miao & Cukurova, 2024).
By the end of the reading the teacher will be able to:

- explain what Chain-of-Thought techniques are and why they bring out the intermediate steps of reasoning;
- use step-by-step reasoning prompts and evaluate their didactic usefulness as a model of procedure;
- reflect critically on the fact that a fluent chain of thought is not a proof of the truth of the result.

1. Making reasoning visible

Chain-of-Thought (literally 'chain of thought') is the technique that invites the model to externalise the logical steps before delivering the answer. Wei and colleagues, who introduced it, describe its core as follows:
«We explore how generating a chain of thought—a series of intermediate reasoning steps—significantly improves the ability of large language models to perform complex reasoning.» (Wei et al., 2022, Abstract)
The pedagogical value of this technique lies in one word: visibility. The machine does not merely give the outcome; it accompanies it with the sequence of steps that justify it. For the teacher, this means being able to assess not only whether the answer is correct, but whether the path leading to it makes sense. It is the same kind of judgement we express on a maths task: the result matters, but so does how one got there.
There is also an immediate benefit for students. An output that shows the steps is not merely a finished product to use: it is a model of procedure to imitate. Asking the machine to 'reason out loud' means obtaining a gradual explanation that the student can follow, compare with their own method and use as scaffolding for their own learning.

2. The 'step by step' prompt

How is reasoning activated? With a targeted instruction. Kojima and colleagues showed that complex examples are not needed: it is enough to add to the request an invitation to proceed one step at a time.

«[...] we show that LLMs are decent zero-shot reasoners by simply adding “Let’s think step by step” before each answer.» (Kojima et al., 2022, Abstract)

The simplicity of the device is surprising, and the results confirm it. In the same work, the addition of that invitation produced a clear improvement in performance on reasoning tasks:

«[...] increasing the accuracy on MultiArith from 17.7% to 78.7% and GSM8K from 10.4% to 40.7% with large-scale InstructGPT model (text-davinci-002) [...]» (Kojima et al., 2022, Abstract)

For the teacher, the lesson is twofold. On the one hand, that the way the request is formulated matters enormously: two almost identical prompts can produce very different performances. On the other, that the didactic value of an output lies not only in the final answer, but in the trace of reasoning that accompanies it. It is this trace, and not only the outcome, that makes the process discussable in class.

This is a point not to be misunderstood. The simplicity of the invitation does not mean that reasoning is guaranteed: it means that the model possesses, in latent form, the ability to break down a problem, and that a suitable formulation is enough to elicit it. The teacher's task, here too, is to choose that formulation and carefully assess its outcome.

3. Self-consistency: multiple paths, one answer

A single chain of thought can follow a wrong path. Self-consistency addresses precisely this risk: instead of generating a single line of reasoning, the model generates several and chooses the most coherent answer among those obtained. Wang and colleagues explain the method as follows:

«It first samples a diverse set of reasoning paths instead of only taking the greedy one, and then selects the most consistent answer by marginalizing out the sampled reasoning paths.» (Wang et al., 2022, Abstract)

The underlying idea is intuitive and recalls common wisdom: if several independent routes lead to the same result, confidence in that result increases. The authors formulate it explicitly:

«Self-consistency leverages the intuition that a complex reasoning problem typically admits multiple different ways of thinking leading to its unique correct answer.» (Wang et al., 2022, Abstract)

On the didactic level, self-consistency offers an excellent starting point for discussion: it is the principle of cross-checking. When we ask the machine to reach the same conclusion by different routes, we are applying the same rigour we ask of students when they compare procedures. Convergence between independent routes is not a proof, but it is an indication of robustness from which to begin verifying.

A classroom experience makes the idea concrete. A history teacher asks the machine for three alternative reconstructions of the causes of an event - economic, political, cultural - and compares with the class the points of convergence and the divergences. The conclusions that emerge from several routes become the axes of a concept map; the points on which the reconstructions diverge become open questions to explore further. Cross-checking is not

used to 'find the right answer': it serves to show that reasoning admits different paths, and that judgement arises from comparison.

4. Trees of thought: explore, evaluate, go back

An evolution of Chain-of-Thought is the Tree of Thoughts. Instead of following a single linear chain, the model explores several branches of reasoning, evaluates them and continues along the best ones. Yao and colleagues describe it as follows:

«[...] we introduce a new framework for language model inference, “Tree of Thoughts” (ToT), which generalizes over the popular “Chain of Thought” approach to prompting language models, and enables exploration over coherent units of text (“thoughts”) that serve as intermediate steps toward problem solving.» (Yao et al., 2023, Abstract)

The core of the method is a change of stance: from an associative and immediate answer to a form of deliberate decision-making. The authors speak of a process that considers different reasoning paths, evaluates its own choices, looks ahead and goes back when needed (Yao et al., 2023). The gain in reliability is documented by an impressive figure: on complex problem-solving tasks, while simple Chain-of-Thought solved only a small share of cases, the tree-based approach solved the majority (Yao et al., 2023).

«[...] while GPT-4 with chain-of-thought prompting only solved 4% of tasks, our method achieved a success rate of 74%.» (Yao et al., 2023, Abstract)

Transposed into a didactic key, the Tree of Thoughts suggests a valuable analogy: exploring several hypotheses, evaluating them and reviewing the choices made is exactly what we ask of a student tackling a complex problem. The machine, orchestrated by a good prompt, dramatises precisely that pathway.

5. The inconvenient truth: a fluent trace is not a proof

We come to the most delicate point, which runs through the entire module. Chains of thought improve performance and make the process visible, but they do not guarantee correctness. Wei and colleagues say it without ambiguity:

«[...] there is no guarantee of correct reasoning paths, which can lead to both correct and incorrect answers.» (Wei et al., 2022, Discussione)

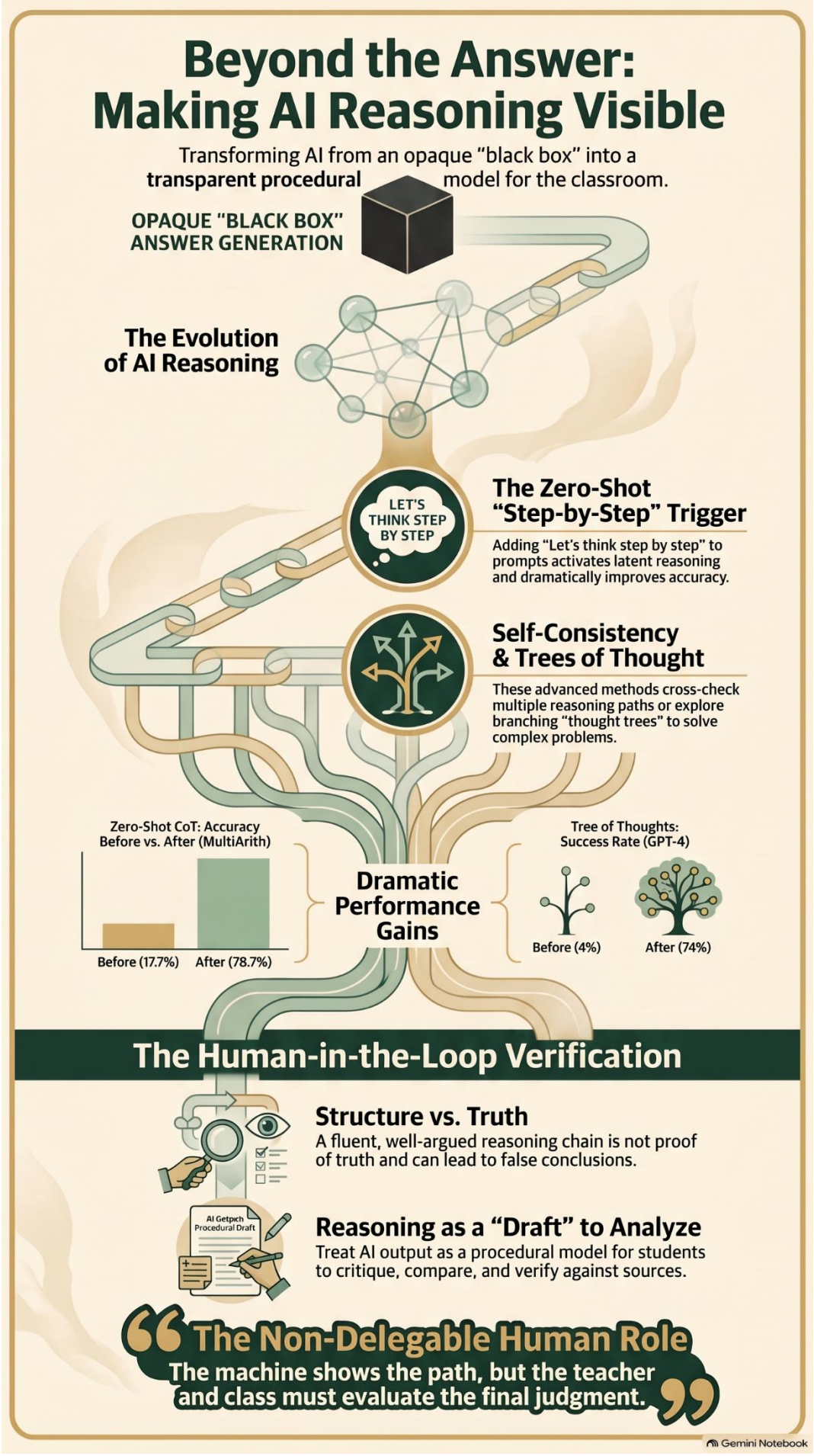
It is the 'inconvenient truth' of these techniques. A well-written, orderly and persuasive line of reasoning can lead to a false conclusion. Indeed, the fluency of the text can make the error more convincing, because a procedure set out with confidence appears more reliable than it is. The teacher's task is therefore twofold: to use the trace of reasoning as material for analysis, and not to confuse the form of thought with its truth.

In operational terms, this means treating the reasoned output as a draft to be examined, not as a solution to be adopted. The key steps are identified, it is verified whether the inferences are justified, and the conclusion is compared with reliable sources. It is the same human-in-the-loop attitude that the STIAEM project recommends for every didactic use of AI: the machine shows the path, the teacher and the class evaluate it.

Technique	What it asks the model to do	Classroom example	Didactic value

Chain-of-Thought	Show the intermediate steps before the answer	'Reason step by step, then write the solution'	Makes the procedure observable; a model to imitate
Zero-shot CoT	Reason without examples, with an explicit invitation	'Let's think step by step: break down the problem'	Activates reasoning on everyday tasks
Self-consistency	Generate several paths and choose the most coherent answer	'Propose three different solutions and compare them'	Teaches cross-checking between independent routes
Tree of Thoughts	Explore, evaluate and review several branches of reasoning	'Evaluate alternative hypotheses and go back if needed'	Trains deliberate decision-making on complex problems

Visual Scaffolding



Conclusions and Metacognitive Reflection Prompts

This resource has shown how explicit-reasoning techniques - Chain-of-Thought, step-by-step prompting, self-consistency and trees of thought - bring out the intermediate steps the machine performs before answering. Their didactic value is the visibility of the process: they turn the output into an object of methodological analysis, not a simple answer to use.

But precisely this visibility requires vigilance. A fluent and well-argued chain of thought is not a proof of truth: it can lead, with the same confidence, to correct or erroneous conclusions. Using these techniques in class therefore means turning the process into a subject of discussion, keeping the final judgement in human hands, according to the human-in-the-loop logic that the STIAEM project adopts as a principle.

To transfer these core ideas into one's own practice, the following prompts for reflection are proposed:

In which type of task in your subject is the machine's trace of reasoning more useful than the final result?

How would you use an 'error' in a formally correct chain of thought to educate students about the distinction between form and truth?

What added value does comparing different reasoning paths on the same problem have for your students?

Technical Metadata Sheet

Metadata Field	Specification / Parameter Value
Identifier	OER 4.2.2
Title	Applying Chain-of-Thought Techniques to Display Intermediate Logical Steps
Creator	Rhodes Evening High School - Greece
Description	The resource addresses the techniques that make the machine externalise the intermediate steps of its reasoning: Chain-of-Thought, self-consistency and tree-of-thought. It shows how step-by-step prompts can be used pedagogically as a model of procedure and how to probe multiple reasoning paths. The critical focus is that a fluent chain of thought is not a proof of the correctness of the result.
Keywords	chain-of-thought; step-by-step prompting; self-consistency; tree of thought; reasoning; procedural modelling; prompt engineering; critical evaluation
Language	EN (English)
Affiliation	Rhodes Evening High School - Greece
Contributor	STIAEM Transnational Scientific Board (peer review, Rhodes validation protocol)
Media Format	Text; Infographics
Genre	Training Manual; Educational Guide
URL	
Citation	Cosgrove, J., & Cachia, R. (2025). DigComp 3.0: European digital competence framework (Fifth edition). Publications Office of the European Union. https://data.europa.eu/doi/10.2760/0001149 Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. Advances in Neural Information Processing Systems, 35. https://arxiv.org/abs/2205.11916 Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://doi.org/10.54675/ZJTE2084 Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., & Zhou, D. (2022). Self-consistency improves chain of thought reasoning in language models. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/2203.11171

	Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35. https://arxiv.org/abs/2201.11903 Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. Advances in Neural Information Processing Systems, 36. https://arxiv.org/abs/2305.10601
Educational Level	Secondary School
Associated Module / MOOC Video Lesson	Module 4 - Prompt Engineering for Inclusive Education / Video 4.2 - Advanced Prompting and Chain-of-Thought
Learning Objectives	explain what Chain-of-Thought techniques are and why they bring out the intermediate steps of reasoning; use step-by-step reasoning prompts and evaluate their didactic usefulness as a model of procedure; reflect critically on the fact that a fluent chain of thought is not a proof of the truth of the result.
Rights	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)
Date Available	14/09/2026
Date Modified	

References

Cosgrove, J., & Cachia, R. (2025). DigComp 3.0: European digital competence framework (Fifth edition). Publications Office of the European Union. https://data.europa.eu/doi/10.2760/0001149

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. Advances in Neural Information Processing Systems, 35. https://arxiv.org/abs/2205.11916

Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://doi.org/10.54675/ZJTE2084

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., & Zhou, D. (2022). Self-consistency improves chain of thought reasoning in language models. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/2203.11171

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35. https://arxiv.org/abs/2201.11903

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. Advances in Neural Information Processing Systems, 36. https://arxiv.org/abs/2305.10601.