



AI Guide

ONLINE COURSE CONTENT

MODULE 2: EVALUATION OF AI RECOMMENDATIONS AND OUTPUTS

2025-1-LT01-KA220-HED-000354784

Artificial Intelligence for Green Understanding, Insights & Decision Empowerment



CAM CONSULTING



SQUAREDEV



Co-funded by
the European Union

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 National Agency of Erasmus+ Programme in Lithuania – Education Exchanges Support Foundation (ŠMPF). Neither the European Union nor the ŠMPF can be held responsible for them.

Table of Contents

<i>Evaluation of AI recommendations and outputs.....</i>	<i>3</i>
A. General Information	3
B. Content for STEM HE Teachers.....	3
Highlights.....	3
Introduction.....	3
Accuracy.....	4
Relevance.....	4
Completeness.....	5
Consistency.....	5
Transparency.....	5
Explainability.....	5
Reliability.....	6
Timeliness.....	6
Fairness and Objectivity.....	6
Actionability.....	6
Common problems.....	7
References.....	7
Green Decision Context.....	8
Videos/Interactive Resources	8



Evaluation of AI recommendations and outputs

A. General Information

Module No	Module2
Subtopic	Evaluation of AI recommendations and outputs
AI Literacy Component	AI output interpretation, uncertainty, and logical inconsistency detection.
Decision Making Skills	Evaluating alternatives; judging reliability; comparing solutions; detecting inconsistency
Learning Objectives for STEM HE Teachers	Critically evaluate AI-generated recommendations and outputs for accuracy, relevance, uncertainty, inconsistency, and contextual fit.
Learning Objectives for STEM Students	Critically evaluate AI-generated environmental recommendations before using them in decisions.
Suggested AI Tools	Chatbots; Consensus; Elicit; Perplexity
Estimated Duration	60 minutes

B. Content for STEM HE Teachers

Highlights

- AI-generated recommendations should support and not replace human decision-making.
- High-quality AI outputs should be evaluated for accuracy, relevance, completeness, consistency, transparency, and reliability before being used in STEM contexts.
- AI outputs may contain uncertainty, hallucinations, logical inconsistencies, bias, or outdated information, even when they appear highly convincing.
- Critical evaluation requires comparing AI recommendations with trusted evidence, alternative solutions, and the specific decision context.
- Human expertise remains essential for interpreting AI recommendations and making informed, ethical, and context-aware decisions, particularly in STEM and sustainability applications.

Introduction

AI has become an increasingly important tool for supporting decision-making across various fields, such as science, engineering, education, healthcare, and environmental management. Modern AI systems are capable of generating recommendations, summaries, predictions, explanations, simulations, and complete reports within seconds. While these capabilities significantly improve productivity, AI-generated outputs should never be accepted uncritically.

Unlike traditional software, generative AI does not retrieve facts directly from a database. Instead, it predicts the most probable continuation of text based on patterns learned during training. Consequently, AI may generate responses that appear highly convincing while containing factual

inaccuracies, outdated information, unsupported assumptions, or fabricated references. This phenomenon is often referred to as hallucination.

For educators and STEM professionals, the ability to critically evaluate AI outputs is becoming as important as the ability to use AI tools themselves. AI should therefore be viewed as a decision-support assistant rather than an autonomous decision-maker.

AI systems may:

- produce incorrect factual statements;
- generate plausible but nonexistent references;
- overlook important contextual information;
- inherit biases from training data;
- provide outdated information;
- misunderstand ambiguous and confusing prompts;
- express excessive confidence even when uncertain.

These limitations become particularly significant in domains involving public safety, sustainability, environmental protection, healthcare, engineering, or public policy, where incorrect recommendations may have serious consequences.

Characteristics of High-Quality AI Recommendations

The quality of AI-generated recommendations determines whether they can be safely and effectively incorporated into professional decision-making. High-quality recommendations should not only provide correct answers but also be relevant, transparent, and appropriate for the specific context in which they will be used [1]. Evaluating these characteristics enables users to distinguish reliable outputs from responses that require further verification or refinement.

Accuracy

Accuracy refers to the degree to which an AI-generated recommendation reflects factual information and produces correct conclusions. An accurate recommendation is supported by valid evidence, consistent with established scientific knowledge, and free from factual errors [2]. Accuracy is particularly important in STEM disciplines, where incorrect information may lead to poor decisions, safety risks, or inefficient resource management.

Because generative AI models predict likely responses rather than retrieving verified facts, users should validate important claims against trusted sources such as scientific publications, government reports, technical documentation, or official databases. Numerical values, scientific terminology, legal requirements, and cited references should receive particular attention during verification.

Relevance

A recommendation may be factually correct while still being unsuitable for the user's needs. Relevance describes how well the AI output addresses the specific problem, context, and objectives defined in the prompt. High-quality recommendations remain focused on the requested task without introducing unnecessary or unrelated information.

Context plays a critical role in determining relevance. Factors such as geographical location, available resources, organizational constraints, environmental conditions, or regulatory requirements may significantly influence whether a recommendation is applicable. Users should therefore evaluate whether the AI has correctly interpreted the context and whether additional prompt refinement is necessary.

A response may be technically correct but irrelevant if it:

- answers a different question;
- includes unnecessary information;
- ignores important constraints;
- misunderstands the requested context.

Completeness

Completeness refers to the extent to which the recommendation provides all essential information needed to support informed decision-making. Rather than offering a single answer, a complete recommendation explains the reasoning behind the proposed solution, identifies assumptions, discusses possible alternatives, and highlights any limitations or uncertainties [3].

Incomplete outputs often omit critical details such as implementation requirements, potential risks, prerequisite conditions, or expected trade-offs. In complex decision-making scenarios, missing information can be as problematic as incorrect information. Users should therefore consider whether additional questions are needed to obtain a more comprehensive response.

Consistency

Consistency describes the logical coherence of the recommendation both internally and across multiple related responses. A consistent AI output avoids contradictions, conflicting statements, and incompatible conclusions. Terminology, equations, numerical values, assumptions, and recommendations should remain stable throughout the response [4].

Consistency is also important when interacting with AI over multiple prompts. If repeated questions produce substantially different answers without explanation, users should investigate the reasons for these discrepancies and verify the information independently. Consistent responses generally indicate that the model has interpreted the task reliably, although consistency alone does not guarantee correctness.

The generated information should not contradict itself. Examples include:

- conflicting numerical values,
- incompatible recommendations,
- inconsistent terminology,
- contradictory conclusions.

Transparency

While many modern AI systems function as complex "black boxes," high-quality outputs should clearly communicate the assumptions, inputs, and reasoning that contributed to the final recommendation whenever possible [3].

Transparent recommendations distinguish between verified facts, assumptions, interpretations, and predictions. They acknowledge uncertainty where appropriate and avoid presenting speculative information as established fact. Transparency enables users to critically assess the recommendation and determine whether additional verification is required.

Users should understand:

- how conclusions were reached;
- which assumptions were made;
- what evidence supports the recommendation;
- where uncertainty exists.

Explainability

Closely related to transparency is explainability, which focuses on how clearly the AI communicates the reasoning behind its conclusions. Rather than simply providing an answer, explainable recommendations describe the logical steps that connect the available information to the proposed solution [3].

Explainability is especially valuable in educational settings because it supports learning and helps users develop critical thinking skills. In professional environments, understandable explanations increase user confidence, facilitate peer review, and simplify the identification of potential errors or

questionable assumptions. When recommendations influence significant decisions, explainability becomes an essential component of responsible AI use.

A useful AI recommendation should explain why a conclusion was reached rather than only presenting the final answer. Explainable outputs improve:

- trust,
- learning,
- error detection.

Reliability

Reliability describes the degree to which an AI system produces dependable and stable recommendations under similar conditions. Reliable AI consistently generates outputs of comparable quality when provided with equivalent inputs and behaves predictably across repeated interactions [5].

Reliability also involves recognizing the limitations of the model. A reliable system should indicate when available information is insufficient, when uncertainty is high, or when expert consultation is advisable. Users should avoid relying solely on a single AI-generated recommendation for high-stakes decisions and should instead combine AI support with independent verification and professional expertise.

Timeliness

For many applications, particularly those involving environmental monitoring, healthcare, finance, or public policy, the usefulness of a recommendation depends on the underlying information. Timeliness refers to whether the recommendation reflects the most recent data, regulations, scientific discoveries, or real-world conditions.

Since some AI models rely on historical training data, users should verify that recommendations involving rapidly changing information are based on up-to-date sources. When current information is critical, AI-generated outputs should be supplemented with real-time databases, official reports, or recent publications.

Fairness and Objectivity

High-quality AI recommendations should be as impartial as possible and avoid systematically favoring particular individuals, groups, regions, or viewpoints without objective justification. Fairness involves minimizing the influence of biases that may exist within training data or be introduced through prompt formulation.

Users should evaluate whether recommendations are supported by evidence rather than stereotypes or assumptions. Considering multiple perspectives and consulting diverse information sources can help reduce the impact of unintended bias.

Actionability

Finally, a high-quality recommendation should be actionable. Beyond describing a problem or presenting general advice, it should provide practical guidance that users can implement. Actionable recommendations typically include clear next steps, identify required resources, specify potential constraints, and, where appropriate, suggest methods for monitoring outcomes.

In professional practice, actionable recommendations are often more valuable than purely descriptive responses because they support informed decision-making and facilitate effective implementation.



Co-funded by
the European Union

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 National Agency of Erasmus+ Programme in Lithuania – Education Exchanges Support Foundation (SMPF). Neither the European Union nor the SMPF can be held responsible for them.



Common problems

Many AI systems generate responses that include references to scientific articles, books, reports, websites, or legislation. The presence of citations often increases users' confidence in the output, giving the impression that the information has been verified or comes from respected sources. However, AI-generated citations should not automatically be considered trustworthy.

Unlike traditional search engines or academic databases, generative AI models produce text by predicting likely sequences of words rather than retrieving and verifying references from a curated repository. As a result, AI may generate citations that appear credible but contain inaccuracies, incomplete information, or entirely fabricated sources. Consequently, every citation included in an AI response should be independently evaluated before being used in academic work, professional reports, or decision-making.

One of the most common issues is the generation of references that do not exist. AI may combine the names of real authors, realistic article titles, existing journals, and plausible publication years to create citations that appear authentic but cannot be found in any academic database.

For example, an AI might cite:

Smith, J., & Johnson, A. (2023). Artificial Intelligence for Sustainable Water Management. Journal of Environmental Analytics, 18(4), 245–261.

Although the citation follows the correct academic format, the article, journal issue, or even the journal itself may not exist.

Sometimes the cited publication exists, but important bibliographic details are incorrect. Errors may include:

- incorrect author names;
- wrong publication year;
- incorrect journal title;
- incorrect volume or issue number;
- invalid page numbers;
- incorrect Digital Object Identifier (DOI).

Even small errors may make it difficult to locate the original source.

AI may correctly identify an existing publication but incorrectly attribute information to it. A cited article may not actually contain the statement or conclusion that the AI claims it does.

For instance, an AI response may state:

According to the IPCC (2023), solar energy has zero environmental impact.

While the stated source reports extensively discuss renewable energy, they do not make such an absolute statement. The AI has incorrectly attributed a simplified or exaggerated claim to a reputable source.

References:

- [1] Wang, S., Zhang, X., Wang, Y., & Ricci, F. (2024). Trustworthy recommender systems. *ACM Transactions on Intelligent Systems and Technology*, 15(4), 1-20.
- [2] Boljam, A. M., Mohammed, H. U. H., Alluri, V. R. R., Saini, V., & Bojja, S. G. R. (2025, July). Impact Analysis of Generative AI on the Accuracy and Scalability of Machine Learning Models. In *2025 International Conference on Computing Technologies & Data Communication (ICCTDC)* (pp. 1-7). IEEE.
- [3] White, M., Haddad, I., Osborne, C., Liu, X. Y. Y., Abdelmonsef, A., Varghese, S., & Hors, A. L. (2024). The model openness framework: Promoting completeness and openness for reproducibility, transparency, and usability in artificial intelligence. *arXiv preprint arXiv:2403.13784*.
- [4] Song, Y., Dhariwal, P., Chen, M., & Sutskever, I. (2023). Consistency models.

[5] Huang, Y., Gao, C., Wu, S., Wang, H., Wang, X., Zhou, Y., ... & Zhang, X. (2025). On the trustworthiness of generative foundation models: Guideline, assessment, and perspective. arXiv preprint arXiv:2502.14296.

Green Decision Context

AI recommendations are increasingly used to support green and sustainability-related decisions, including:

- selecting energy-efficient technologies and reducing energy consumption
- recommending sustainable resource management practices (water, materials, waste)
- supporting environmental monitoring through analysis of sensor and satellite data
- evaluating alternative solutions based on environmental impact and carbon footprint
- assisting policymakers, engineers, and organizations in sustainability planning

When evaluating AI recommendations in green contexts, users should verify whether the recommendations:

- are based on accurate and up-to-date environmental information
- consider local conditions, regulations, and sustainability constraints
- explain the assumptions and trade-offs behind the proposed solution
- account for both short-term benefits and long-term environmental impacts
- are supported by reliable scientific evidence and trusted environmental sources

Videos/Interactive Resources

- Evaluating GenAI Output -
<https://www.youtube.com/watch?v=gQx973zsOOw>
- How to Evaluate AI Agents: Comprehensive Strategies for Reliable, High-Quality Agentic Systems -
<https://www.youtube.com/watch?v=gsMRpZdMEIQ>



Co-funded by
the European Union

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 National Agency of Erasmus+ Programme in Lithuania – Education Exchanges Support Foundation (SMPF). Neither the European Union nor the SMPF can be held responsible for them.

