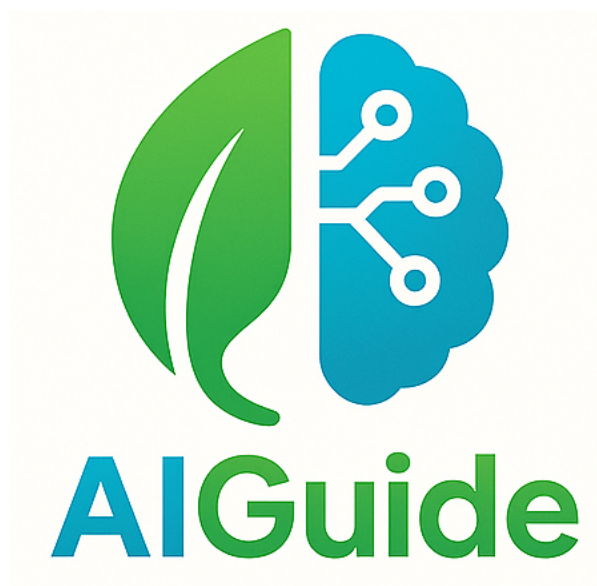




ONLINE COURSE CONTENT SAMPLE

WORK PACKAGE 3 ACTIVITY 2

2025-1-LT01-KA220-HED-000354784

Artificial Intelligence for Green Understanding, Insights & Decision Empowerment



CAM CONSULTING



SQUAREDEV



Co-funded by
the European Union

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Document Identity

Lead Organization	☐ VDU ☒ EGE ☐ URAK ☐ EDUJI ☐ CAM ☐ SQUAREDEV
Confidentiality Status	☒ Internal ☐ Public
Related WP	☐ WP2 ☒ WP3 ☐ WP4 ☐ WP5
Related Activity Name	<i>WP3A3-Development of Online Course Content</i>

Document Version

Version	Date	Organization
V1	01.07.2026	URAK
V2		
V3		
V4		

*This document may change without notice.

**The content in this document is only a sample and aims to create a common ground within the consortium during the workshop. Therefore, the document should not be used or shared.

Table of Contents

1. Evaluation of AI recommendations and outputs	4
A. General Information	4
B. Content for STEM HE Teachers.....	4
Highlights.....	4
Introduction.....	4
Green Decision Context	8
Real World Examples	9
Example Projects	9
Recommended Readings.....	9
Videos/Interactive Resources.....	9
2. Fact-checking AI-generated information A. General Information	10
B. Content for STEM HE Teachers.....	10
Highlights.....	10
Introduction.....	10
Green Decision Context	16
Real World Examples.....	17
Example Projects	17
Videos/Interactive Resources	17
2. Awareness of limitations, bias and transparency.....	17
A. General Information	17
B. Content for STEM HE Teachers.....	18
Highlights.....	18
Introduction.....	18
Green Decision Context	20
Real World Examples (Choose from best practices report)	21



Example Projects (Choose from best practices report).....	21
---	----



1. 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 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.

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

Real World Examples

Example Projects

Recommended Readings

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>



2. Fact-checking AI-generated information

A. General Information

Module No	Module2
Subtopic	Fact-checking AI-generated information
AI Literacy Component	Misinformation detection, hallucination awareness, source triangulation, and scientific evidence verification.
Decision Making Skills	Evidence verification, rejecting unsupported claims, and validating information.
Learning Objectives for STEM HE Teachers	Apply fact-checking procedures to AI-generated claims using reliable scientific and technical sources.
Learning Objectives for STEM Students	Verify AI-generated environmental claims using reliable sources.
Suggested AI Tools	Consensus; Elicit; ResearchRabbit; Connected Papers; Perplexity
Estimated Duration	60 minutes

B. Content for STEM HE Teachers

Highlights

- AI-generated information should not be accepted without verification, as generative AI predicts likely text rather than retrieving verified facts.
- Fact-checking is a core AI literacy skill, combining critical thinking, information literacy, source evaluation, and evidence-based reasoning before AI outputs are used in education.
- AI can generate several types of factual errors, including hallucinated facts, fabricated citations, incorrect statistics, false quotations, outdated information, and statements that lack important context.
- The level of verification depends on the risk of the information. High-risk domains, such as healthcare, engineering, environmental data, public policy, and safety, require particularly detailed fact-checking.
- Effective fact-checking requires verifying claims using independent and authoritative sources, checking that references are genuine, information is current, and important context has not been omitted.
- Established information literacy frameworks, such as the SIFT Method and the CRAAP Test, provide practical approaches for evaluating the credibility and reliability of AI-generated information before it is used or shared.

Introduction

The increasing integration of AI into education has fundamentally changed the way teachers prepare lessons, develop instructional materials, assess student work, and support learning. Generative AI systems such as ChatGPT, Claude, Gemini, Microsoft Copilot, and similar large language models (LLMs) are capable of producing coherent explanations, summaries, lesson plans, quizzes, programming code, and even scientific literature reviews within seconds. Their ability to



generate human-like text has made them valuable educational assistants, reducing the time required for routine tasks and enabling teachers to focus on pedagogical activities.

However, despite their impressive capabilities, generative AI systems are not designed as factual databases. Their primary objective is to generate statistically probable sequences of words rather than retrieve verified facts from authoritative sources. As a result, AI-generated responses may contain inaccuracies, fabricated references, outdated information, or misleading interpretations presented with a high degree of confidence. These limitations create significant challenges, particularly in educational environments where teachers are responsible for providing accurate, evidence-based information to students.

For this reason, fact-checking AI-generated information has become one of the core competencies of AI literacy [1]. Rather than accepting AI outputs at face value, educators must develop the ability to critically evaluate the accuracy, reliability, completeness, and credibility of AI-generated content before incorporating it into teaching materials or recommending it to learners.

Fact-checking should therefore be viewed not merely as a technical skill but as a component of digital information literacy, critical thinking, and responsible AI use. Educators who understand both the strengths and limitations of generative AI are better positioned to use these technologies safely while simultaneously teaching students how to become critical consumers of AI-generated information.

Why AI-Generated Information Cannot Be Automatically Trusted

Unlike traditional search engines that retrieve existing documents, generative AI systems operate by predicting the most likely sequence of words based on patterns learned during their training process [2]. During training, large language models analyze enormous collections of books, websites, articles, and other textual resources. Through this process they learn statistical relationships between words, phrases, and concepts rather than storing verified facts.

Consequently, an AI model does not "know" whether a statement is true or false. Instead, it estimates which sequence of words is most likely to follow a user's prompt. While this mechanism often produces remarkably accurate responses, it also explains why AI occasionally generates convincing but entirely incorrect information.

Several factors contribute to factual inaccuracies [3]:

- incomplete or outdated training data;
- ambiguous or poorly formulated prompts;
- conflicting information within the training corpus;
- lack of access to current events or newly published research (for models without real-time retrieval);
- probabilistic text generation rather than factual verification.

One of the best-known consequences of this process is the phenomenon known as an AI hallucination [4]. Hallucinations occur when an AI system confidently generates information that appears plausible but has no basis in reality. These may include invented scientific papers, fabricated quotations, imaginary organizations, incorrect statistics, or non-existent legal regulations. For example, an AI system may produce a citation that follows the correct academic formatting style and includes realistic author names, journal titles, publication years, and page numbers, despite the referenced article never having existed. Such fabricated references are particularly problematic in education because they may easily pass unnoticed unless carefully verified.

It should be noted, that hallucinations are not software bugs but an inherent consequence of the probabilistic nature of the LLMs. Therefore, educators should approach AI-generated information as draft material requiring verification rather than authoritative knowledge.



Understanding Fact-checking

Fact-checking is the systematic process of verifying whether claims correspond to reliable evidence obtained from trustworthy and independent sources. Within the context of AI-generated information, fact-checking extends beyond simply asking whether an answer is correct. It also includes evaluating the quality of supporting evidence, identifying missing context, assessing the credibility of cited sources, and determining whether the information is sufficiently current for the intended purpose.

Effective fact-checking addresses several key questions:

- Is the factual claim accurate?
- Can the statement be confirmed by independent sources?
- Are the cited references genuine?
- Is the information up to date?
- Has important context been omitted?
- Does the response distinguish facts from opinions or predictions?
- Is uncertainty communicated appropriately?

Effective FACT-CHECKING

addresses several key questions:



Fact-checking therefore combines elements of critical thinking, source evaluation, information literacy, and evidence-based reasoning [5]. For educators, it becomes an essential quality assurance process before AI-generated materials are incorporated into classroom activities, presentations, assignments, or assessments.

Types of AI-Generated Information That Require Verification

Not all AI-generated content carries the same level of risk. The amount of verification required depends largely on the consequences of incorrect information.

Low-risk information

Some responses involve general knowledge that is relatively stable over time. Examples include definitions of technical terms, grammar explanations, mathematical formulas, programming syntax,



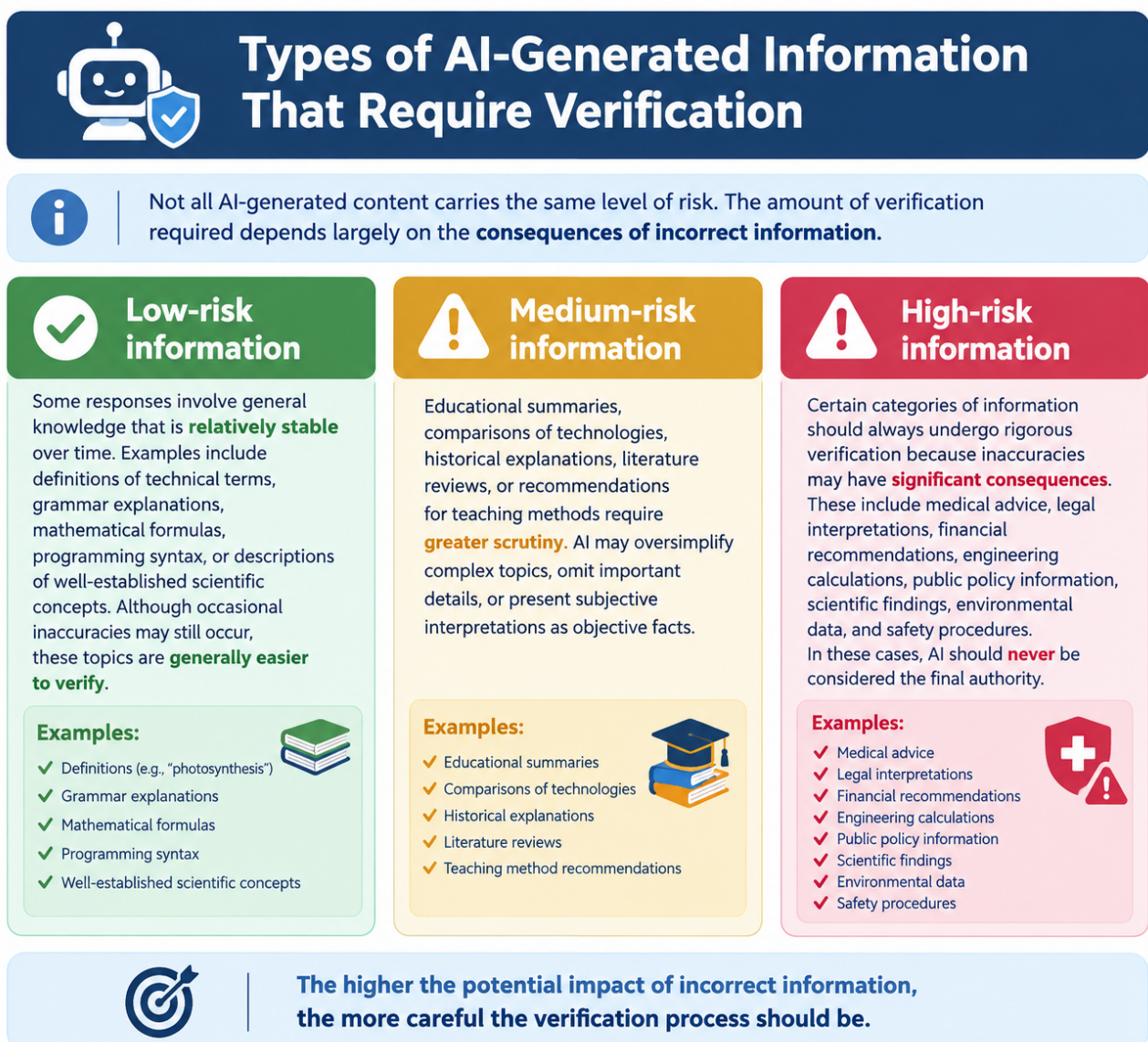
or descriptions of well-established scientific concepts. Although occasional inaccuracies may still occur, these topics are generally easier to verify.

Medium-risk information

Educational summaries, comparisons of technologies, historical explanations, literature reviews, or recommendations for teaching methods require greater scrutiny. AI may oversimplify complex topics, omit important details, or present subjective interpretations as objective facts.

High-risk information

Certain categories of information should always undergo rigorous verification because inaccuracies may have significant consequences. These include medical advice, legal interpretations, financial recommendations, engineering calculations, public policy information, scientific findings, environmental data, and safety procedures. In these cases, AI should never be considered the final authority.



Common Types of AI Errors

Understanding typical AI errors enables educators to recognize situations where additional verification is necessary [6].



Hallucinated facts

The AI generates information that sounds convincing but is entirely fictional.

Example:

"The International Association of Digital Education established AI Literacy Standards in 2019."

Although plausible, such an organization or publication may not exist.

Fabricated citations

One of the most common problems in academic writing is the generation of non-existent references.

AI may produce a citation such as:

Smith, J., & Brown, L. (2023). Artificial Intelligence in Modern Education. Journal of Educational Innovation.

Despite appearing authentic, neither the journal nor the article may exist.

Incorrect numerical information

Statistics are frequently generated without proper evidence. For example:

"Seventy-eight percent of teachers regularly use AI in classrooms."

Unless accompanied by a verifiable source, such statistics should not be trusted.

False quotations

Large language models often attribute statements to famous individuals that they never actually said. For example:

"Albert Einstein once said, 'Creativity begins where algorithms end.'"

Searching reliable quotation databases reveals that no such quotation exists.

Outdated information

Rapidly changing topics such as AI technologies, legislation, software versions, or public health recommendations may no longer reflect current knowledge.

Missing context

A statement may be technically correct while still misleading. For example:

"Solar energy is the cheapest source of electricity."

This statement depends on geographical location, climate conditions, storage requirements, infrastructure costs, and the methodology used in economic comparisons.

Without this context, learners may develop an oversimplified understanding.



Common Types of AI Errors

Understanding typical AI errors enables educators to recognize situations where additional verification is necessary

Hallucinated facts

The AI generates information that sounds convincing but is entirely fictional.

Example:
"The International Association of Digital Education established AI Literacy Standards in 2019."

Although plausible, such an organization or publication may not exist.

Fabricated citations

One of the most common problems in academic writing is the generation of non-existent references. AI may produce a citation such as:

Example:
"Smith, J., & Brown, L. (2023). Artificial Intelligence in Modern Education. Journal of Educational Innovation."

Despite appearing authentic, neither the journal nor the article may exist.

Incorrect numerical information

Statistics are frequently generated without proper evidence. For example:

Example:
"Seventy-eight percent of teachers regularly use AI in classrooms."

Unless accompanied by a verifiable source, such statistics should not be trusted.

False quotations

Large language models often attribute statements to famous individuals that they never actually said. For example:

Example:
"Albert Einstein once said, 'Creativity begins where algorithms end.'"

Searching reliable quotation databases reveals that no such quotation exists.

Outdated information

Rapidly changing topics such as AI technologies, legislation, software versions, or public health recommendations may no longer reflect current knowledge.

Missing context

A statement may be technically correct while still misleading. For example:

Example:
"Solar energy is the cheapest source of electricity."

This statement depends on geographical location, climate conditions, storage requirements, infrastructure costs, and the methodology used in economic comparisons.

Without this context, learners may develop an oversimplified understanding.

Frameworks for Evaluating AI-Generated Information

Several well-established frameworks from information literacy can be applied directly to AI-generated content.

The SIFT Method

Developed by digital literacy expert Mike Caulfield, the SIFT method encourages critical evaluation before accepting information [7].

- Stop - Pause before believing or sharing AI-generated content.
- Investigate the source - Identify where the information originates and whether the source is trustworthy.
- Find better coverage - Search for independent confirmation from multiple reliable sources.
- Trace claims to the original source - Whenever possible, consult the original publication rather than relying on summaries or secondary interpretations.

The SIFT method is particularly valuable because it emphasizes verification before engagement, reducing the spread of misinformation.

The CRAAP Test

The CRAAP framework evaluates information according to five criteria [8]:

- Currency – Is the information current?
- Relevance – Does it address the intended purpose?
- Authority – Who produced the information?
- Accuracy – Is it supported by evidence?
- Purpose – Why was the information created?

Originally developed for evaluating online information, this framework is equally applicable to AI-generated outputs.





Frameworks for Evaluating AI-Generated Information

Several well-established frameworks from information literacy can be applied directly to AI-generated content.



The SIFT Method

Developed by digital literacy expert Mike Caulfield, the SIFT method encourages critical evaluation before accepting information.

S

II

Stop

- Pause before believing or sharing AI-generated content.

I

🔍

Investigate the source

- Identify where the information originates and whether the source is trustworthy.

F

📄

Find better coverage

- Search for independent confirmation from multiple reliable sources.

T

🔗

Trace claims to the original source

- Whenever possible, consult the original publication rather than relying on summaries or secondary interpretations.



The SIFT method is particularly valuable because it emphasizes verification before engagement, reducing the spread of misinformation.



The CRAAP Test

The CRAAP framework evaluates information according to five criteria.

C

📅

Currency

- Is the information current?

R

🎯

Relevance

- Does it address the intended purpose?

A

👤

Authority

- Who produced the information?

A

🛡️

Accuracy

- Is it supported by evidence?

P

🎯

Purpose

- Why was the information created?



Originally developed for evaluating online information, this framework is equally applicable to AI-generated outputs.

References

- [1] Zhang, Y., Fang, W., & Pan, W. (2026). Fact-Checking Misinformation in the Age of AI: Audience Responses to Human, AI, and Hybrid Sources. *International Journal of Human–Computer Interaction*, 1-17.
- [2] Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic markets*, 33(1), 63.
- [3] Uddin, M., Arfeen, S. U., Alanazi, F., Hussain, S., Mazhar, T., & Arafatur Rahman, M. (2026). A critical analysis of generative AI: challenges, opportunities, and future research directions. *Archives of Computational Methods in Engineering*, 33(2), 1763-1793.
- [4] Maleki, N., Padmanabhan, B., & Dutta, K. (2024, June). AI hallucinations: a misnomer worth clarifying. In *2024 IEEE conference on artificial intelligence (CAI)* (pp. 133-138). IEEE.
- [5] Wolfe, R., & Mitra, T. (2024, June). The impact and opportunities of Generative AI in fact-checking. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1531-1543).
- [6] Mahmood, A., Fung, J. W., Won, I., & Huang, C. M. (2022, April). Owning mistakes sincerely: Strategies for mitigating AI errors. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1-11).
- [7] Caulfield, M., & Wineburg, S. Learning to Verify.
- [8] Blakeslee, S. (2004). The CRAAP test. *Loex Quarterly*, 31(3), 4.

Green Decision Context

AI-generated information increasingly supports green and environmental decision-making, including:

- interpreting environmental monitoring data (air, water, soil, biodiversity)
- summarizing climate change research and sustainability reports
- providing information on renewable energy technologies and circular economy practices



- explaining environmental regulations, standards, and sustainability policies
- supporting education on climate action and sustainable development

Before using AI-generated information in green contexts, users should verify:

- scientific claims against trusted environmental organizations and peer-reviewed sources
- numerical data such as carbon emissions, energy consumption, and environmental indicators
- whether environmental regulations and sustainability standards are current
- that cited reports, legislation, and scientific publications actually exist
- whether important environmental trade-offs or local conditions have been omitted

Real World Examples

Example Projects

Videos/Interactive Resources

- Fact checker – Manus - <https://manus.im/playbook/fact-checker>
- Fact-checker – FullFact - <https://fullfact.ai/>

2. Awareness of limitations, bias and transparency

A. General Information

Module No	Module2
Subtopic	Awareness of limitations, bias and transparency
AI Literacy Component	AI bias, AI limitations, black-box problem, transparency, and explainability.
Decision Making Skills	Critical judgement, bias detection, fairness and accountability evaluation.
Learning Objectives for STEM HE Teachers	Explain AI limitations, bias mechanisms and transparency issues relevant to STEM and green decisions.
Learning Objectives for STEM Students	Identify bias, limitations and lack of transparency in AI-supported outputs.
Suggested AI Tools	Model cards; dataset documentation; chatbots;
Estimated Duration	60 minutes



B. Content for STEM HE Teachers

Highlights

Introduction

An essential aspect of AI literacy is understanding that AI systems are probabilistic models rather than repositories of verified knowledge. They predict the most likely sequence of words based on patterns learned from vast amounts of training data. Consequently, AI-generated answers may contain factual inaccuracies, incomplete information, logical inconsistencies, outdated knowledge, or unintended biases.

Teachers, researchers, and students must therefore develop awareness of three fundamental concepts:

- limitations – what AI systems can and cannot do;
- bias – how AI may systematically favor certain viewpoints;
- transparency – understanding how AI generates answers, what information it uses, and how confident users should be in its output.

Developing this awareness enables users to critically evaluate AI-generated information rather than accepting it uncritically.

Understanding the Limitations of AI Systems

One of the most common misconceptions is that AI "knows" information in the same way humans do. In reality, modern language models do not possess human understanding or factual knowledge. Instead, they generate responses by predicting statistically probable word sequences learned during training. For example, when asked:

Who discovered penicillin?

the AI generates the answer "Alexander Fleming" because this association frequently appears in its training data, and not because it independently verifies historical facts.

This distinction is crucial because AI cannot correctly distinguish between truth and falsehood unless supported by reliable information during training or connected to external knowledge sources.

Although AI often appears capable of sophisticated reasoning, much of its performance results from recognizing patterns rather than genuine logical understanding. AI may successfully solve many reasoning problems but can also fail on relatively simple tasks when [1]:

- multiple logical steps are required;
- assumptions are ambiguous;
- contextual information is missing;
- mathematical precision is essential.

For example, an AI may correctly explain Newton's laws but incorrectly solve a multi-step physics problem if intermediate reasoning is flawed.

Perhaps the best-known limitation of generative AI is hallucination [2]. An AI hallucination occurs when the system confidently produces information that is false, fabricated, or unsupported by evidence. Examples include [2]:

- inventing scientific references;
- fabricating quotations;
- creating non-existent legal cases;
- generating incorrect statistics;
- describing fictional events as real.



A particularly dangerous aspect is that hallucinated answers are often presented with high linguistic confidence, making them difficult for inexperienced users to recognize. For example:

"According to Smith et al. (2022)..."

may appear entirely credible even though no such publication exists.

Hallucinations become more likely when users ask highly specialized, obscure, or recent questions for which reliable information is limited.

AI models are trained on data collected during specific time periods. Unless connected to external search systems or live databases, they cannot automatically know:

- recent scientific discoveries;
- current legislation;
- breaking news;
- newly released products;
- recent policy changes.

Consequently, responses may accurately reflect historical knowledge while omitting recent developments.

Although modern AI systems can process long conversations, they still have practical context limits [3]. Long interactions may lead to [4]:

- forgotten earlier instructions;
- inconsistent answers;
- contradictions;
- loss of important details.

For lengthy projects, users should periodically restate objectives and verify that AI continues following the intended requirements.

Bias in AI-Generated Answers

Bias refers to systematic tendencies that cause AI-generated responses to favor particular perspectives, assumptions, or groups over others [5]. Bias is not always intentional. Instead, it often reflects patterns present within the data used to train the model [5]. Because AI learns from billions of documents written by humans, it inevitably inherits some characteristics of human language, culture, and society.

Several factors contribute to AI bias [6]:

- Training data bias – if certain viewpoints appear more frequently in training data, AI is more likely to reproduce them;
- Historical bias – training datasets include historical documents;
- Selection bias – not all information available on the Internet is included in AI training;
- Prompt bias – the phrasing of a user's question can itself introduce bias.

Types of Bias in AI Response [7]:

- Cultural bias – responses may reflect assumptions common in particular regions or societies while overlooking alternative cultural perspectives;
- Language bias – high-resource languages generally receive better support than less commonly spoken languages;
- Confirmation bias – if users ask leading questions, AI may reinforce the assumptions embedded in those questions;
- Availability bias – frequently discussed topics receive richer and more detailed responses than subjects with limited documentation;
- Popularity bias – widely accepted opinions often dominate AI outputs even when legitimate minority viewpoints exist.



Transparency in AI answers

In the context of AI-generated answers, transparency refers to the extent to which users understand how an AI system produces its responses, what information the responses are based on, and what limitations or uncertainties may affect their reliability. Unlike traditional software, where the logic behind outputs can often be explicitly traced, modern generative AI systems operate through complex neural networks that learn statistical relationships from vast amounts of data. As a result, the internal decision-making process is not fully interpretable, making transparency particularly important for responsible AI use.

From a user's perspective, transparency means being informed about the nature of AI-generated content. Users should understand that AI systems do not "know" facts in the human sense, nor do they independently verify every statement they generate. A transparent AI system should communicate the limitations of its answers rather than presenting every response as unquestionably correct.

An important aspect of transparency is the communication of confidence and uncertainty. Well-designed AI systems increasingly acknowledge when information may be uncertain, incomplete, or dependent on interpretation. For example, an AI may indicate that evidence on a particular topic is limited, that multiple scientific viewpoints exist, or that recent developments could have changed the available knowledge. Such statements help users distinguish between well-established facts and areas where additional verification is recommended. Transparency therefore encourages informed decision-making rather than blind trust in AI-generated outputs.

References:

- [1] Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., ... & Mirjalili, S. (2023). Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. *Authorea preprints*, 1(3), 1-26.
- [2] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2), 1-55.
- [3] Kuratov, Y., Bulatov, A., Anokhin, P., Rodkin, I., Sorokin, D., Sorokin, A., & Burtsev, M. (2024). Babilong: Testing the limits of llms with long context reasoning-in-a-haystack. *Advances in Neural Information Processing Systems*, 37, 106519-106554.
- [4] Wang, X., Salmani, M., Omid, P., Ren, X., Rezagholizadeh, M., & Eshaghi, A. (2024). Beyond the limits: A survey of techniques to extend the context length in large language models. *arXiv preprint arXiv:2402.02244*.
- [5] Wei, X., Kumar, N., & Zhang, H. (2025). Addressing bias in generative AI: Challenges and research opportunities in information management. *Information & Management*, 62(2), 104103.
- [6] Ranjan, R., Gupta, S., & Singh, S. N. (2024). A comprehensive survey of bias in LLMs: Current landscape and future directions. *arXiv preprint arXiv:2409.16430*.
- [7] Tiwari, S. (2025). Biases and Fairness in LLMs. In *Generative AI: Techniques, Models and Applications* (pp. 229-242). Cham: Springer Nature Switzerland.

Green Decision Context

AI is increasingly used to support green and sustainability-related decisions, such as:

- analyzing environmental and climate data
- recommending sustainable farming and resource management practices
- supporting renewable energy planning and energy-efficiency measures
- assessing environmental impacts of projects and technologies
- assisting in biodiversity conservation and ecosystem management



When using AI for green decision-making, users should understand that:

- AI may generate incorrect or oversimplified explanations of environmental processes.
- AI may reflect biases present in its training data, favoring certain regions, technologies, or viewpoints.
- AI may not include the latest environmental regulations, climate research, or sustainability standards.
- AI recommendations should be interpreted in the context of local environmental conditions and expert knowledge.
- Transparency about the assumptions, uncertainties, and limitations of AI outputs is essential for responsible environmental decision-making.

Real World Examples (Choose from best practices report)

Example Projects (Choose from best practices report)

