



Venturing into AI:
Learning for an
Unbounded, Ethical, and
Sustainable Europe

values

Module 4: Reliable AI-Assisted
Content

Sources, Verification and Hallucination Control



www.valuesai.eu



Co-funded by
the European Union



Content



Venturing into AI: Learning for an Unbounded, Ethical, and Sustainable Europe

Module 4: Reliable AI-Assisted Content: Sources, Verification and Hallucination Control

- 01 Creating credible content in the AI era (10 min)
- 02 AI content risks (20 min)
- 03 Accountable author (20 min)
- 04 Working with sources (15 min)
- 05 AI hallucinations (10 min)
- 06 Content verification (10 min)
- 07 Data analysis with AI (10 min)
- 08 Quiz - Safe Data with AI (10 min)



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

2 of 54



01

Creating credible content in the AI era

(10 min)

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



Reliable AI content

Creating credible content in the AI era is not only about producing text quickly, but above all about ensuring accuracy, transparency, and responsibility. Using AI does not automatically reduce quality, but it does require stronger human oversight, especially when the content concerns facts, data, health, politics, or social issues.



Source: Copilot Microsoft 365



01. Creating credible content in the AI era

Reliable AI content

Key principles

- Cite sources. Refer to studies, reports, official statements, and original data.
- Verify before publishing. Every piece of content should be checked for facts, numbers, and context.
- Disclose AI use. Transparency increases audience trust.
- Write clearly and precisely. Avoid vague wording that can lead to misinterpretation.
- Update content. When new data appears or an error is found, revise the material.
- Ensure clear authorship and accountability. Content should have an identifiable author or editorial owner.
- Do not manipulate the audience. AI should not be used to deliberately mislead people, even if it increases reach.



Source: Copilot Microsoft 365



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

01. Creating credible content in the AI era

Reliable AI content

What builds trust

Credible content usually explains not only what but also why and on what basis, so combining data with expert commentary and practical examples works well.

- publication and update dates,
- links to primary sources,
- an expert or author with relevant credentials,
- a logical text structure,
- a clear distinction between fact, opinion, and interpretation,
- no sensational or clickbait-style language.

In practice, credible content in the AI era is created when AI supports the work, but humans remain responsible for truthfulness and meaning.



Source: Copilot Microsoft 365



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

01. Creating credible content in the AI era

Why „reliable content” matters with AI

AI tools can produce content quickly, but every output is a draft: fluent text, plausible numbers, and convincing summaries that may still be wrong.

★ ★ ★

Why this matters:

- Once published, weak content can damage your credibility and mislead your audience.
- AI can fabricate citations, statistics, and quotes that look real (hallucinations).
- „Reliable” means more than fluent-it means sourced, verified and accountable.

Example: Asking ChatGPT to define a concept gives a fluent answer with no sources. The same prompt in NotebookLM, with academic PDFs loaded, returns a sourced definition tied to real authors.



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

7 of 54

02. AI content risks



Co-funded by
the European Union

02

AI content risks

(20 min)



Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

8 of 54

AI content carries three risks: hallucinations, missing sources, and AI detection

Hallucination risk (false facts):

AI can invent dates, statistics, names and quotes that sound real but never existed.

Source risk (no foundation):

General models like ChatGPT 4.0 produce fluent text without grounding it in real publications. Bibliography is missing, vague, or fabricated.

Detection risk (AI fingerprints):

Detectors flag generic, list-like, formulaic AI text. Output without editing scores 100% AI; grounded, edited output scores under 10%.

When people use AI for content, they often focus only on speed. In practice, AI output carries three major risks.



„Looks reliable” but isn’t (Myth vs Reality)

Watch for:

perfectly formatted citations with wrong page numbers



real author names attached to invented papers



plausible statistics with no verifiable source



fluent definitions that contradict the actual literature

Myth: „It’s reliable if it sounds academic.”

Fact: AI imitates academic style perfectly while still inventing facts.



Myth: „It’s safe if a detector says it’s human-written.”

Fact: Detectors only check style. Edited AI output can score 92% original and still contain hallucinations.



What is AI SLOP?

It sounds confident, but includes:

Fabricated facts



Wrong numbers or dates



Incorrect names, citations, or scientific claims



Repetitive, generic structure



„Smooth” text with zero real evidence

AI slop = low-quality, misleading, or inaccurate content generated by AI.



Why it happens?

1

The model predicts patterns, not truth – it generates the most likely response based on data, but it does not actually “know” whether it is true.

2

It may fill in gaps with invented details – when information is missing, the model can add something that sounds logical but may not reflect reality.

3

It does not always have up-to-date information – if the model’s knowledge is time-limited, it may provide outdated facts or miss recent events.

4

It may misunderstand context – an unclear question, shorthand, or ambiguity can lead to the wrong answer.



Why it happens?

5

It sounds confident even when wrong – the wording may appear highly reliable even if the content contains errors.

6

It can mix real facts with hallucinations – part of the answer may be correct, while other parts are invented or unverified.

7

It depends on the quality of the input – if the prompt is unclear, contradictory, or incomplete, the output may be less accurate.

This is why verification becomes a key digital skill.



How AI SLOP destroying the Internet



AI-generated content is flooding the internet, from videos and music to news and books. On top of that, generative AI is making false information look more convincing than ever.

We're entering a new era of information overload, and figuring out what's real and what's not is becoming more and more of a challenge.



03. Accountable author



Co-funded by
the European Union

03

Accountable author

(10 min)



Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

15 of 54

Accountable lens



Use four accountability questions whenever you produce content with AI:

- ★ Accuracy: Did I check facts, dates, and quotes against trusted sources?
- ★ Sources: Can I cite real publications behind every key claim?
- ★ Originality: Did I edit the AI draft so it is not detectable as raw AI output?
- ★ Accountability: Can I explain how I used AI, what I changed, and why?

Reliable content = smart use of AI guided by accountability.



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

16 of 54

Accountable lens

Safe starts before AI

- What you enter or upload matters first

Rights and consent matter

- Other people's data needs permission and care

Fairness is part of safety

- Biased / low-quality data can cause harm

Safer workflows exist

- You can often get AI help while sharing far less.



Before you use AI: every output is a draft



Many people focus on AI speed. Reliable content starts earlier: what you verify, ground, and edit.

In practice, AI tools generate:

- ★ definitions and explanations (often without sources)
- ★ summaries of papers (sometimes invented or distorted)
- ★ lists of references (frequently fabricated)
- ★ data analyses and interpretations (without methodology)



Co-funded by
the European Union

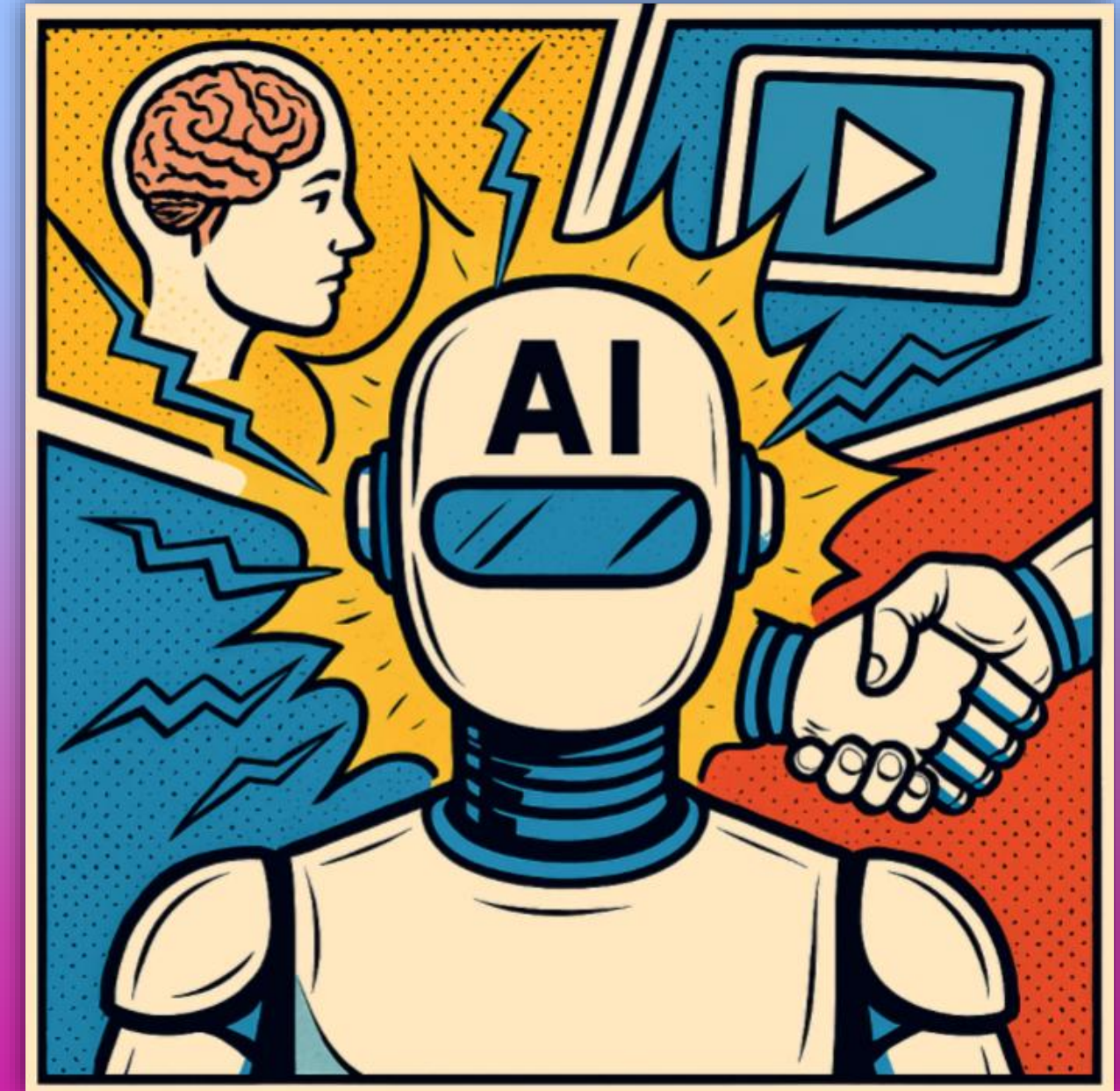
Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

Before you use AI: every output is a draft

Reliable AI use begins with a simple habit: treat every output as a draft to verify, not a final answer.



Source: Copilot Microsoft 365



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

03. Accountable author

AI output is a draft, not truth

What counts as content you must verify:

- ★ definitions and theoretical claims
- ★ statistics, percentages, and dates
- ★ names of authors, books, and articles
- ★ data analyses and conclusions

DON'T assume: „It sounds professional” or „It must be true if AI said so.”

DO remember: Every fact, citation, and number from AI must be checked before use.



The Never Trust List (NTL)

Never Trust List (NTL) - Do not publish without checking:

- ★ Citations: any reference to a book, paper, or author the AI named
- ★ Quotes: any „direct quote” the AI attributes to a person
- ★ Statistics: percentages, sample sizes, study results, rankings
- ★ Historical facts: dates, events, locations, names of laws or treaties
- ★ Definitions and theories: claims about what a concept „is” or who proposed it
- ★ Numerical analysis data: client lists, HR data, internal documents, non-public figures

If it cannot be traced to a verified source-do not publish.

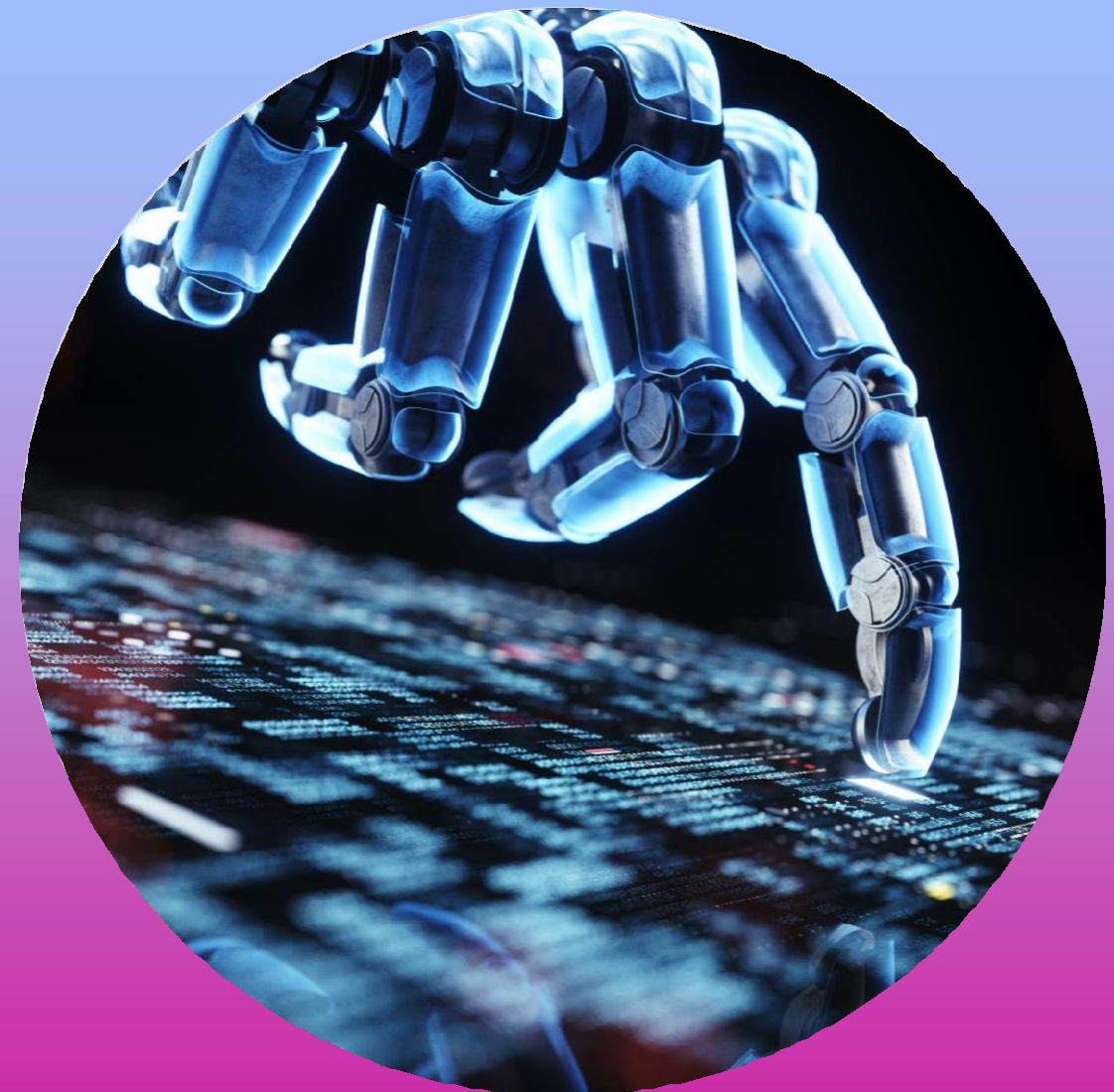


How to make AI output more reliable (before you publish)

Before publishing AI output, apply these checks:

- ★ Verify citations: open in Google Scholar, Scopus, or publisher site
- ★ Cross-check every statistic with at least one independent source
- ★ Re-run the prompt in a grounded tool (e.g., NotebookLM with real PDFs loaded)
- ★ Edit structure, replace generic phrases, add specific examples

If you cannot find the source of a claim → do not publish it





04

Working with sources

(15 min)



Sources provided by AI



Using credible sources – choose materials from reputable journals, public institutions, scholarly databases, and expert organizations.



Distinguishing scholarly, industry, media, and opinion sources – each source type serves a different function and offers a different level of reliability.



Checking publication dates – currency matters, especially in fast-changing fields where AI may provide outdated information.



Verifying citations – check whether the cited author, title, year, and quoted claim actually exist and match what the text or AI says.



Sources provided by AI



Distinguishing primary and secondary sources – a primary source presents original material, while a secondary source discusses or interprets earlier work.



Cross-checking information – important claims should be confirmed in at least two or three independent sources.



Evaluating the author, publisher, and purpose – always ask who wrote the text, where it was published, and whether its goal is to inform, sell, persuade, or comment.



04. Working with sources

Example of AI providing an incorrect source

ChatGPT 5 Instant

W jaki sposób pandemia COVID-19 zmieniła zachowania konsumentów i strategie marek?

Świetne pytanie badawcze — bardzo aktualne i szeroko analizowane w literaturze z zakresu **marketingu, zarządzania i zachowań konsumentów**. Oto syntetyczna odpowiedź (w formie przydatnej do pracy naukowej lub prezentacji):

🔍 Jak pandemia COVID-19 zmieniła zachowania konsumentów i strategie marek?

1. Cyfryzacja zachowań konsumentów

Pandemia przyspieszyła przejście do kanałów online:

- gwałtowny wzrost zakupów internetowych, subskrypcji i e-commerce (Pantano et al., 2020),
- konsumenci oczekiwali szybkich, bezpiecznych i wygodnych doświadczeń zakupowych (Sheth, 2020),
- rosnąca rola social commerce i platform mobilnych (Liu et al., 2021).

2. Nowe priorytety i wartości konsumentów

- wzrost znaczenia bezpieczeństwa, zdrowia, zaufania i transparentności marek,
- większa wrażliwość na kwestie etyczne, społeczne i środowiskowe (He & Harris, 2020),
- pojawienie się zjawiska „mindful consumption” – bardziej przemyślanych zakupów i unikania nadmiaru.

• Liu, Y., Lee, J. M., & Lee, C. (2021). The challenges and opportunities of digital transformation in the post-COVID-19 world. *Technological Forecasting and Social Change*, 166, 120602.

This slide shows a ChatGPT 5 Instant conversation where the user asks about how the COVID-19 pandemic changed consumer behavior and brand strategies. The AI provides a structured response with citations to academic sources (Pantano et al., 2020; Sheth, 2020; **Liu et al., 2021**; He & Harris, 2020), discussing digitalization of consumer behavior and new consumer priorities including safety, health, trust, and „mindful consumption.” When the user verified one citation, they discovered that the article „The challenges and opportunities of digital transformation in the post-COVID-19 world” **was NOT published** in *Technological Forecasting and Social Change* as the AI claimed, but rather in *IEEE Engineering Management Review*, Vol. 48(3), pp. 97-103.



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

04. Working with sources

Example of AI providing an incorrect source. Conclusion

- Bhatti, A. (2020). Online shopping behavior during COVID-19 pandemic. *International Journal of Business and Social Science*, 11(10), 1-10.
- He, H., & Harris, L. (2020). The impact of COVID-19 pandemic on corporate social responsibility and marketing philosophy. *Journal of Business Research*, 116, 176-182.
- Huang, M.-H., Rust, R. T., & Maksimovic, V. (2021). The feeling economy: Managing in the next generation of artificial intelligence. *California Management Review*, 63(3), 43-65.
- Liu, Y., Lee, J. M., & Lee, C. (2021). The challenges and opportunities of digital transformation in the post-COVID-19 world. *Technological Forecasting and Social Change*, 166, 120602.
- Pantano, E., Pizzi, G., Scarpi, D., & Dennis, C. (2020). Competing during a pandemic? Retailers' ups and downs during the COVID-19 outbreak. *Journal of Business Research*, 116, 209-213.
- Sheth, J. (2020). Impact of COVID-19 on consumer behavior: Will the old habits return or die? *Journal of Business Research*, 117, 280-283.

The AI misattributed a real article to the wrong journal, incorrectly citing it as published in *Technological Forecasting and Social Change* instead of *IEEE Engineering Management Review*.



It generated a plausible but false bibliographic reference, including journal name and publication details, creating a misleading APA-style citation.



The AI failed to verify publication metadata (journal, publisher, DOI) despite the source being easily verifiable.



It combined a real title with an incorrect publishing context, a typical case of source hallucination.



This error undermines academic reliability, potentially leading to incorrect bibliographies and issues during peer review.

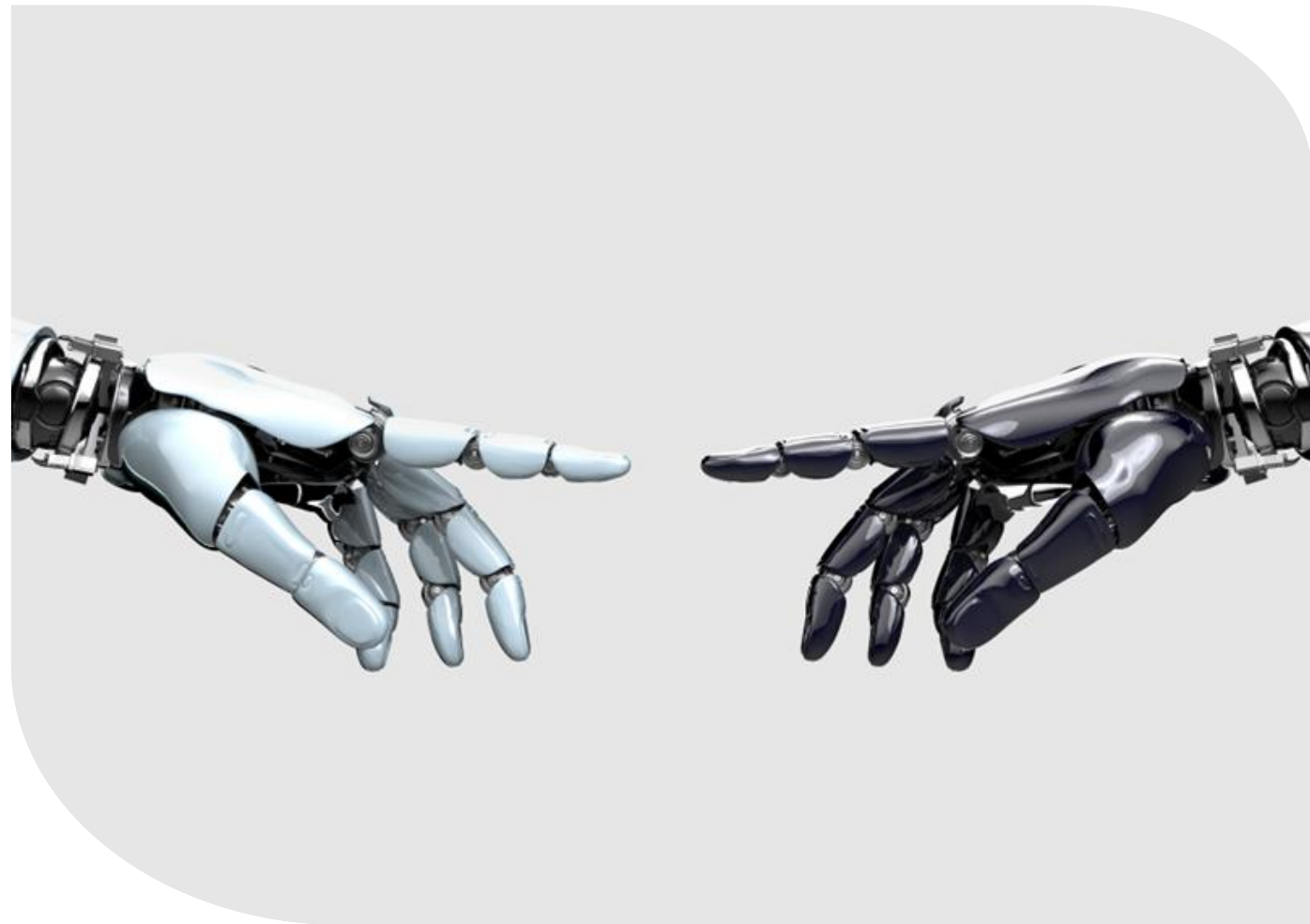


Example of AI providing an incorrect source. Conclusion

AI can effectively summarize literature, but all AI-generated references **must be independently verified** before use in academic work.



Exercise



Ask any AI model to generate a one-page A4 write-up on any topic. Ask it not to use websites, but instead to rely on publications such as articles, books, reports, and similar sources. Create a list of sources. Then, according to good practice, verify the accuracy of the sources provided: do they really exist? Are they accessible? Do they list the correct authors? ... If you detect an AI mistake, share your observations with the group and the instructor.



Scenario: ChatGPT 4.0 vs NotebookLM

A research case study compared two tools on the same task:
„Generate a definition of innovation strategy with a bibliography.”



The same prompt produced two outputs that show what grounding changes:

ChatGPT 4.0: fluent definition, no real bibliography, detected as 100% AI by Originality.ai.

NotebookLM (with real PDFs): grounded definition citing real authors (Porter, Janasz, Pomykalski, Freeman); 92% original score.



What this means in practice:

Without grounding, AI invents fluent but unsourced content. With a real knowledge base, the same model produces verifiable text tied to specific publications.



ChatGPT 4.0 vs NotebookLM: what grounding changes

Sources are not just academic - they are what make AI output accountable.



ChatGPT 4.0 (no grounding):

- Fluent text, simplified definitions, no bibliography
- Plausible company examples (Google, Tesla, Apple) with no „studies” behind them.
- 100% detected as AI-generated by Originality.ai

NotebookLM (with PDFs loaded):

- Theory-grounded definitions citing Porter, Janasz, Pomykalski, Freeman
- Real research traditions: innovation typologies, network economy, SME context.
- Concrete bibliography in APA format, ready to verify
- Style scored 92% original by Originality.ai



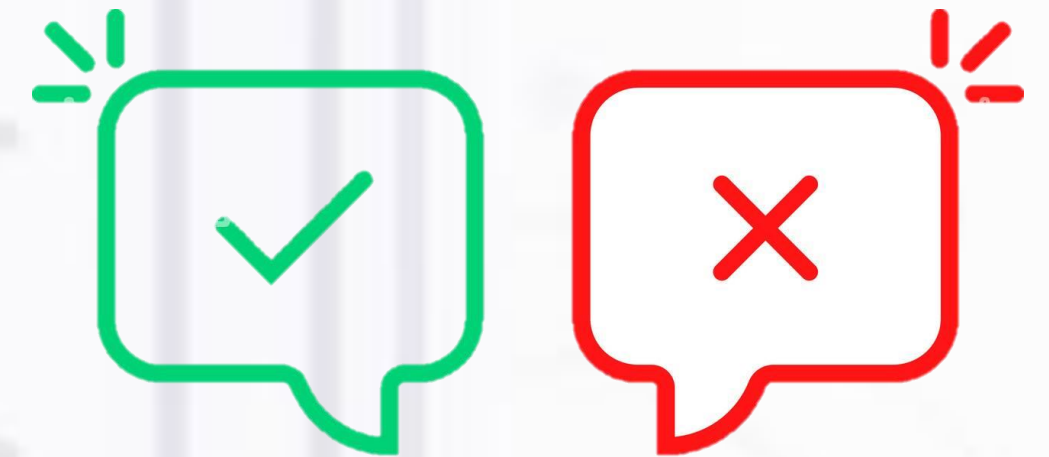
Sources and rights: „Do I have a real reference for this claim?“

Reliable AI use is not only about facts. It is also about authorship: who wrote the original idea, whose work you cite, and how you credit it.



You can usually rely on AI *cannot* trust AI output for:

- exact statistics, percentages, sample sizes
- specific quotes attributed to named authors
- publication years, page numbers, journal titles
- definitions of niche or recent concepts not in mainstream literature
- general explanations of well-known concepts
- structural editing (grammar, flow, paragraphing)
- summarising text you have provided to AI yourself



Do's

Don'ts



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

32 of 54

What a „real reference” looks like (and what it does not)

Reliable AI use is not only about facts. It is also about authorship: who wrote the original idea, whose work you cite, and how you credit it.



A real source is not only a name and a year.

For AI-cited references, a source must be:

- Real: the publication actually exists and is findable in a database
- Specific: matches the claim word-for-word, not just the topic
- Accessible: you can open it and read the relevant section
- Authoritative: published by a credible journal, publisher, or institution
- Important: Consent depends on audience and tool

A person might agree to tell you something, but not agree to have it shared with an AI tool or stored in a system.





05

AI hallucinations

(10 min)

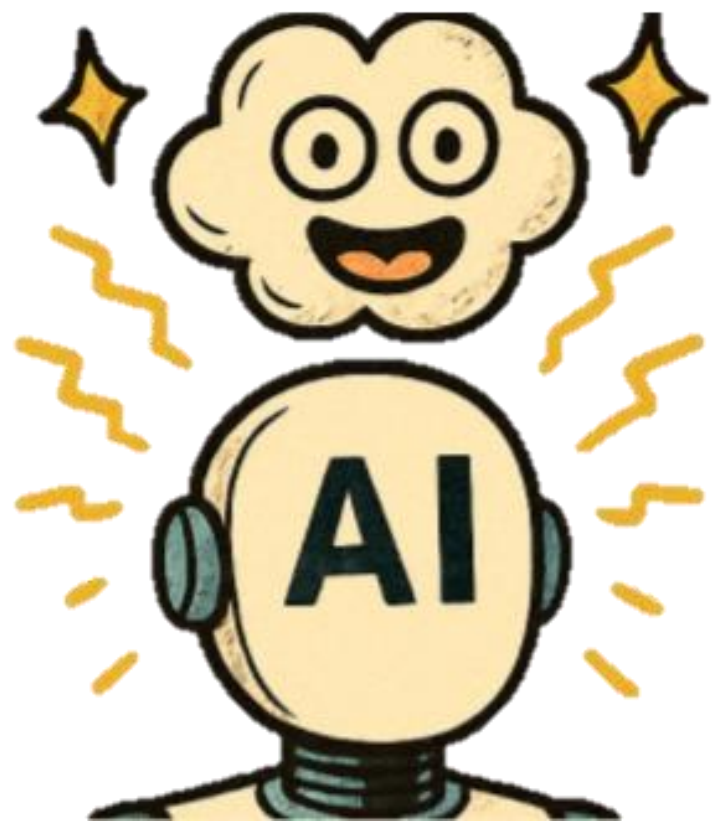


05. AI hallucinations

What AI hallucinations are?

AI hallucinations are model-generated answers that sound plausible but contain false, invented, or misleading information. The model does not „know” what is true — it usually predicts the most likely text rather than checking reality.

AI hallucinations occur when a system produces content that appears correct but is factually wrong, such as a **fabricated quote, a nonexistent source, an incorrect date, or a misleading explanation**. These are not medical hallucinations — they are information-generation errors, not perceptual experiences.



Source: Copilot Microsoft 365



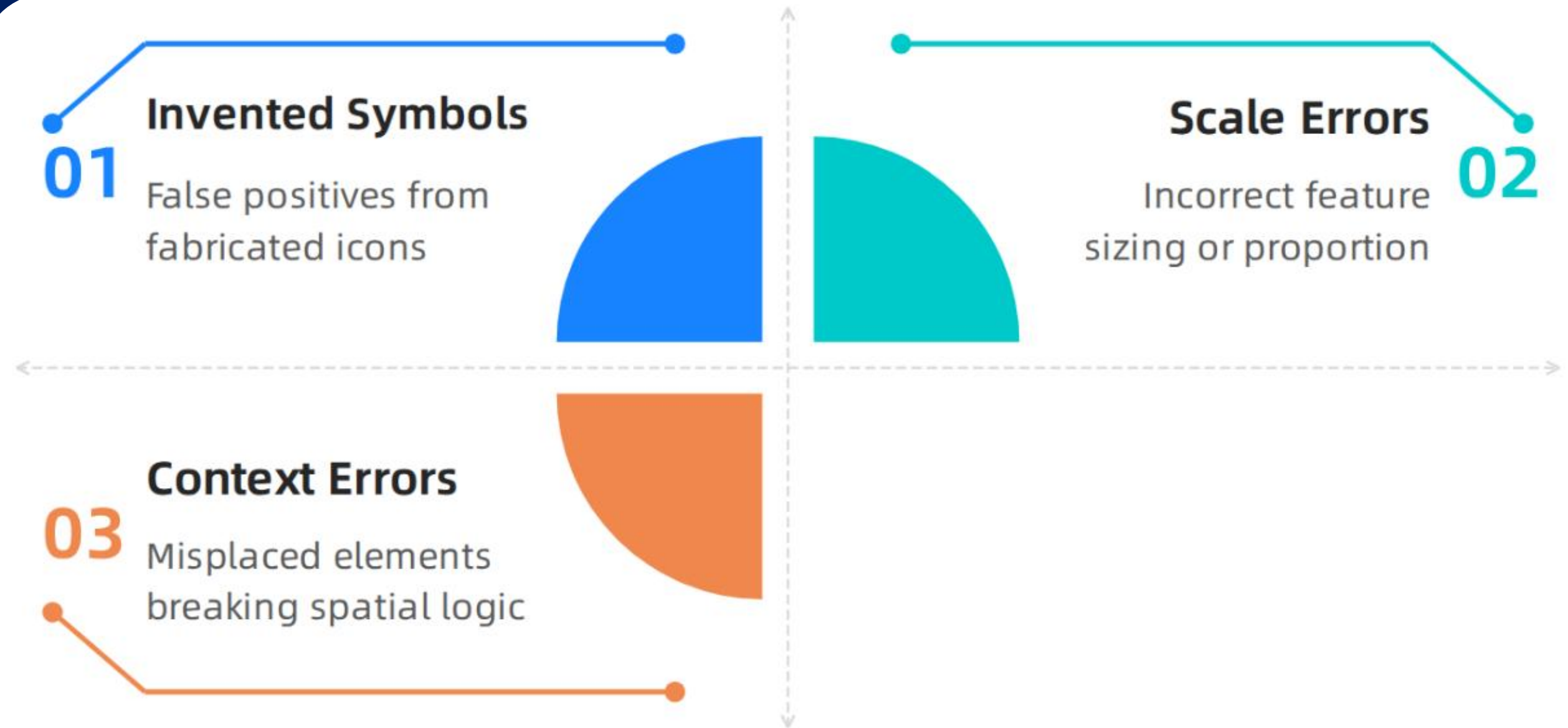
Co-funded by
the European Union



AI Skills, Powered by Values

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

AI hallucinations map



How to detect and fix AI hallucinations in graduated symbol maps. Aihallucinationreport.com. <https://aihallucinationreport.com/how-to-detect-and-fix-ai-hallucinations-in-graduated-symbol-maps/>



Co-funded by the European Union



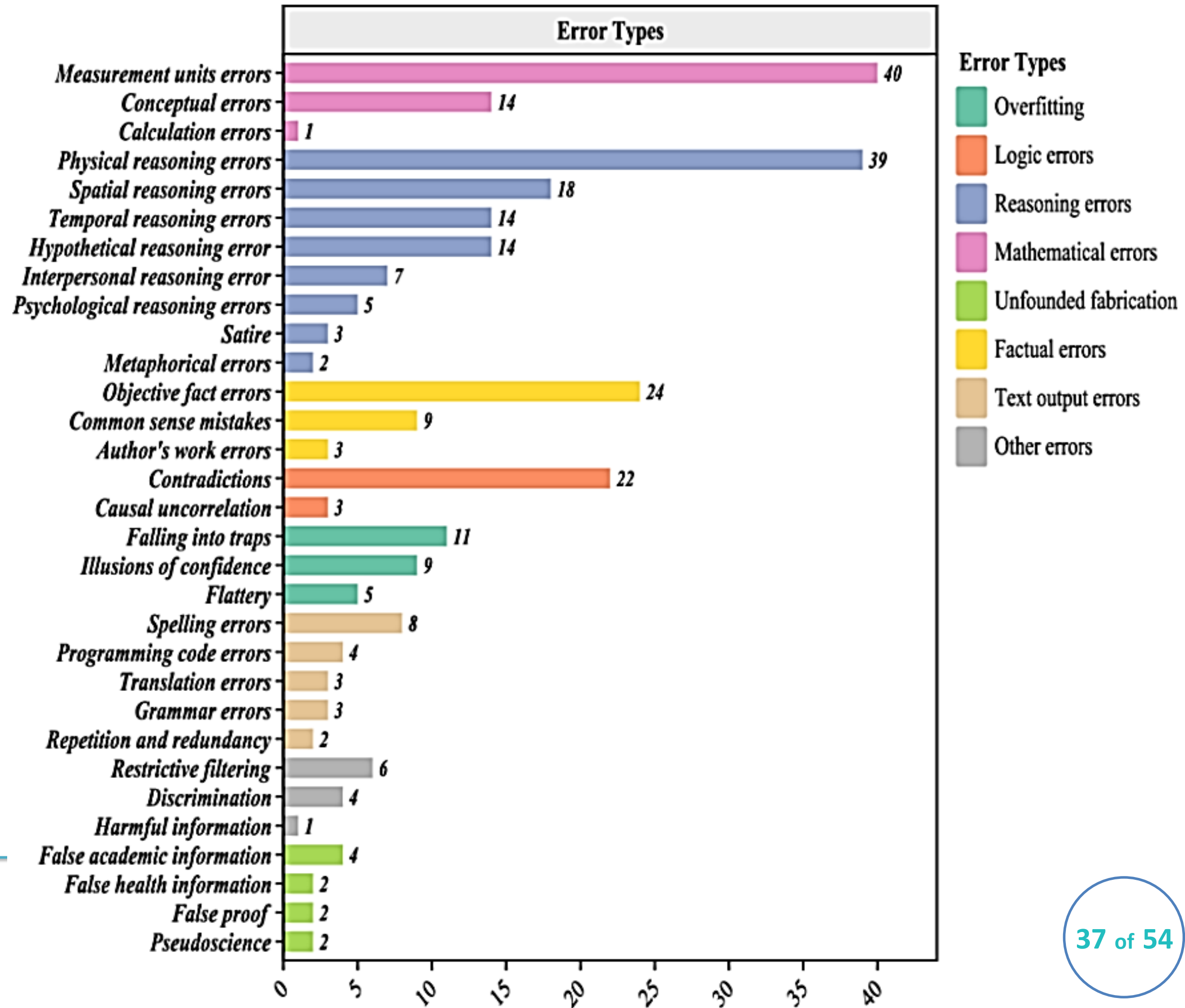
AI Skills, Powered by Values

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

AI hallucinations

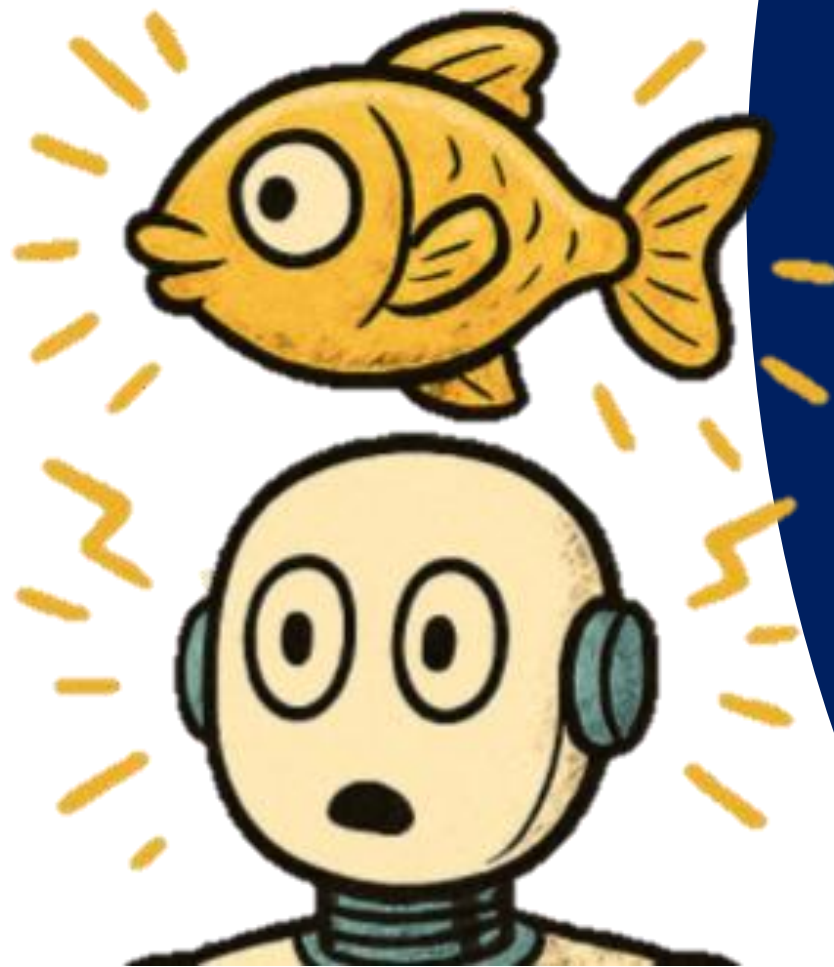
Statistics for 284 analysis units. Different colors represent different error types.

Sun, Yujie & Sheng, Dongfang & Zhou, Zihan & Wu, Yifei. (2024). AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications*. 11. 10.1057/s41599-024-03811-x.



05. AI hallucinations

Types of AI hallucinations



Factual hallucinations – the model states false facts, events, people, or definitions.

Source hallucinations – AI invents sources, publications, authors, titles, or assigns claims to nonexistent texts.

Quote hallucinations – the model gives a quote that was never actually said or falsely attributes it to a person.

Numerical hallucinations – AI miscalculates data, confuses percentages, dates, statistics, or table values.

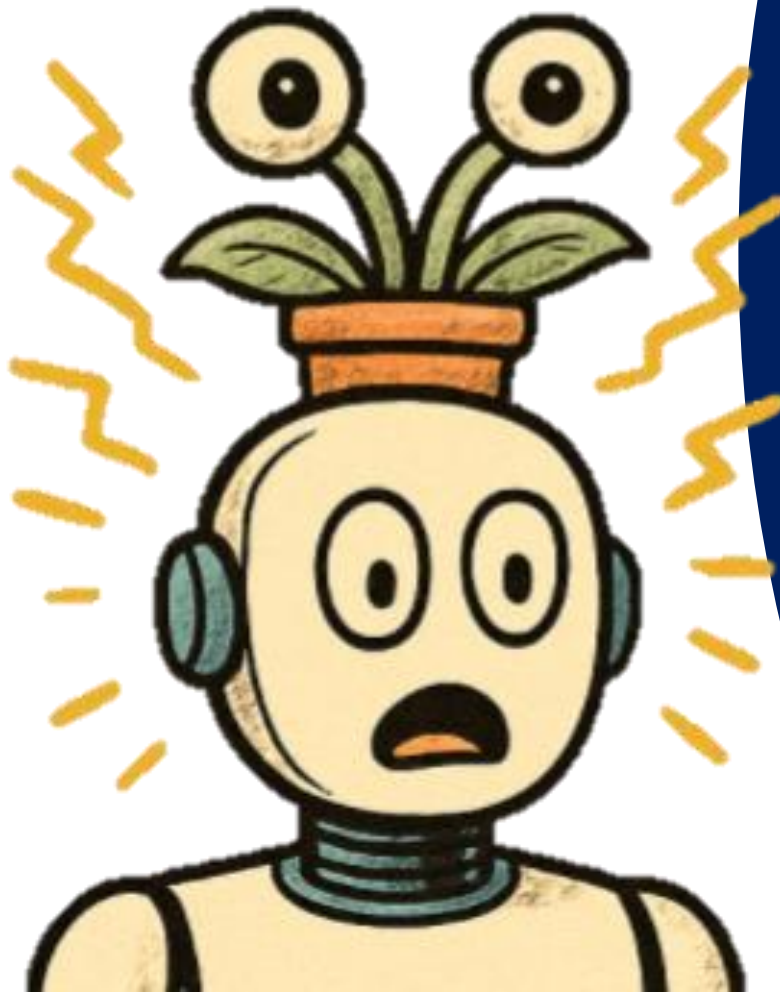
Contextual hallucinations – the model misunderstands the question or adds details the user did not intend.

Mixed hallucinations – the answer contains both correct facts and invented elements, making the error harder to spot.



05. AI hallucinations

Causes of AI hallucinations



The model predicts patterns, not truth – it generates the most likely words rather than checking facts.

Training data may be incomplete or flawed – if the data contain gaps, errors, or contradictions, the model can reproduce them.

Lack of up-to-date information – the model may not know recent events or changes.

Unclear prompt context – with ambiguous or shorthand prompts, the model may misread the user's intent.

Systems reward fluent answers – models are often trained to respond confidently and smoothly, even when uncertain.

No built-in fact checking – without access to external sources, the model cannot verify whether its answer is true.

Source: Copilot Microsoft 365



Co-funded by
the European Union



AI Skills, Powered by Values

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

05. AI hallucinations

Indirect hallucinations: when names and dates look right but aren't

AI can hallucinate even when names, dates and titles look correct. The pattern is plausible but the link to the original work is broken.

Common indirect hallucinations:

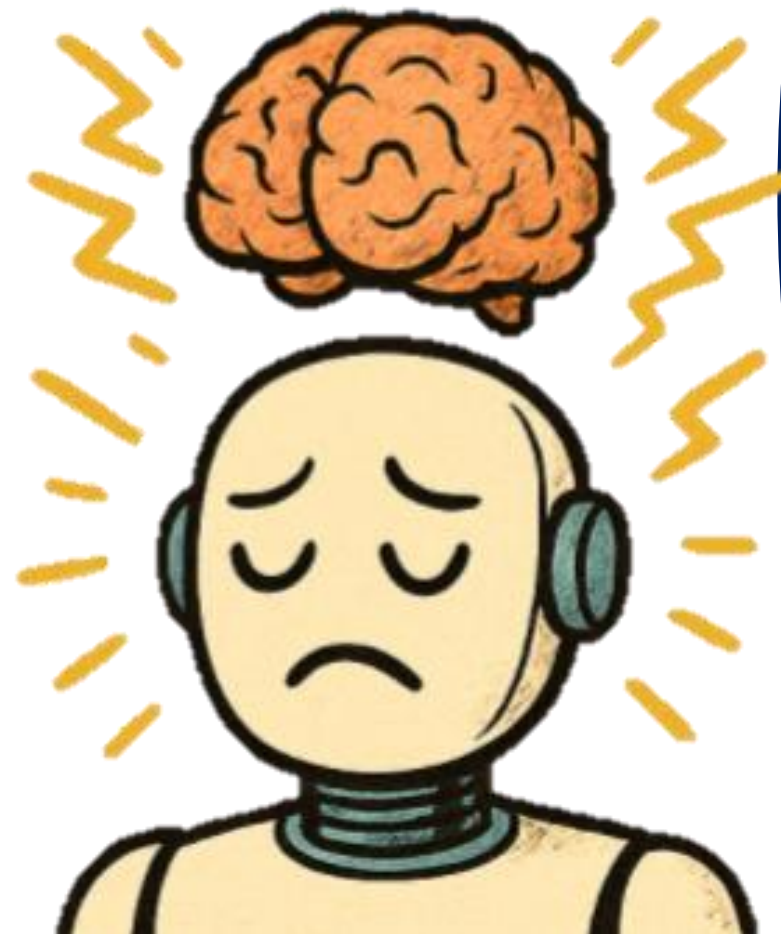


- real author + real journal + invented paper title
- real paper + wrong year, page, or volume number
- real concept + wrong author ("Porter said" when Schumpeter did)
- real statistic + invented context (right number, wrong study)

Safer alternatives:



- search:** open Google Scholar and confirm the exact paper exists
- replace:** cite a paper you have actually read on the same topic
- soften:** use general phrasing if you cannot pin down a specific source
- remove:** if you cannot find the source, delete the claim entirely



Source: Copilot Microsoft 365



Co-funded by
the European Union



AI Skills, Powered by Values

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



06

Content verification

(10 min)



AI-generated content can be useful, but it should not be published or used without prior verification

Treat the AI answer as a draft, not as final truth.

Check claims, numbers, quotes, and sources before using them.

Document what was verified and on what basis.



Source: Copilot Microsoft 365



Co-funded by
the European Union

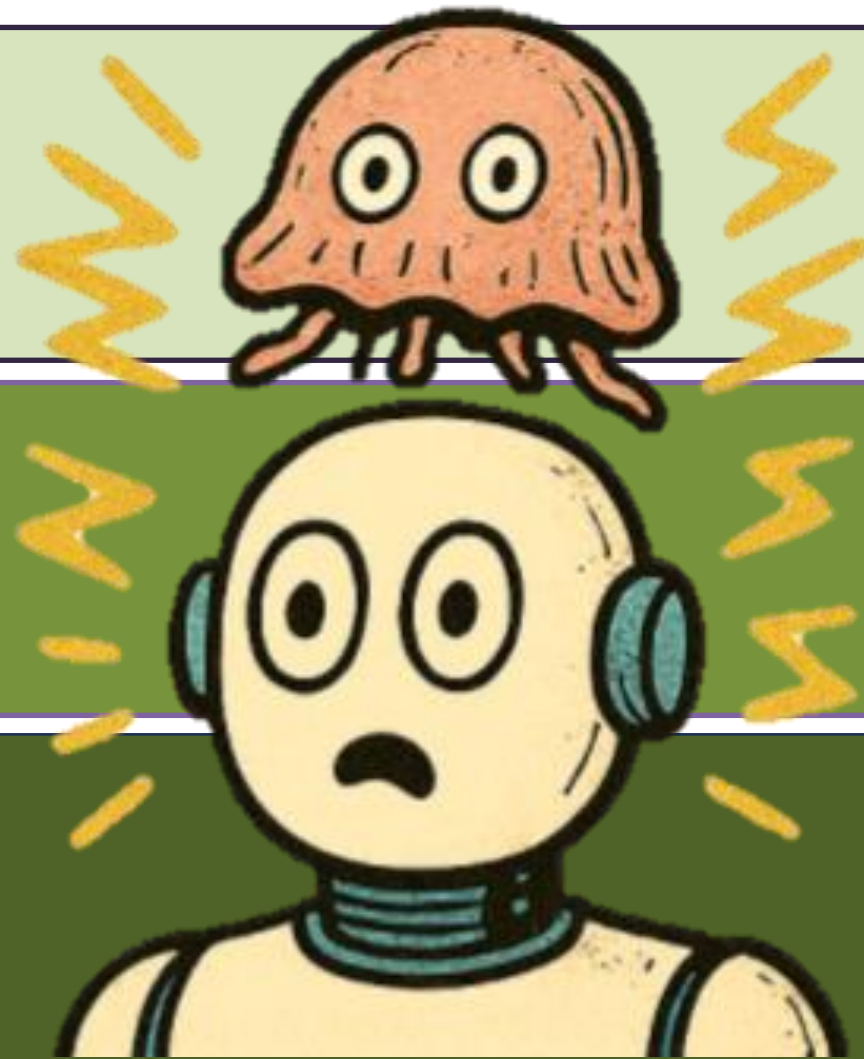
Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

Step 1: Identify risk and key claims

First, identify which parts of the AI response require mandatory checking: facts, dates, names, numbers, quotes, definitions, and cited sources. The greater the impact on health, law, education, finance, or reputation, the stricter the verification should be.



What in this answer is fact, and what is interpretation?

Which elements could cause harm if they are wrong?

Did the AI provide real sources, or does it only sound convincing?

Source: Copilot Microsoft 365



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

43 of 54

Step 2: Verify facts and sources

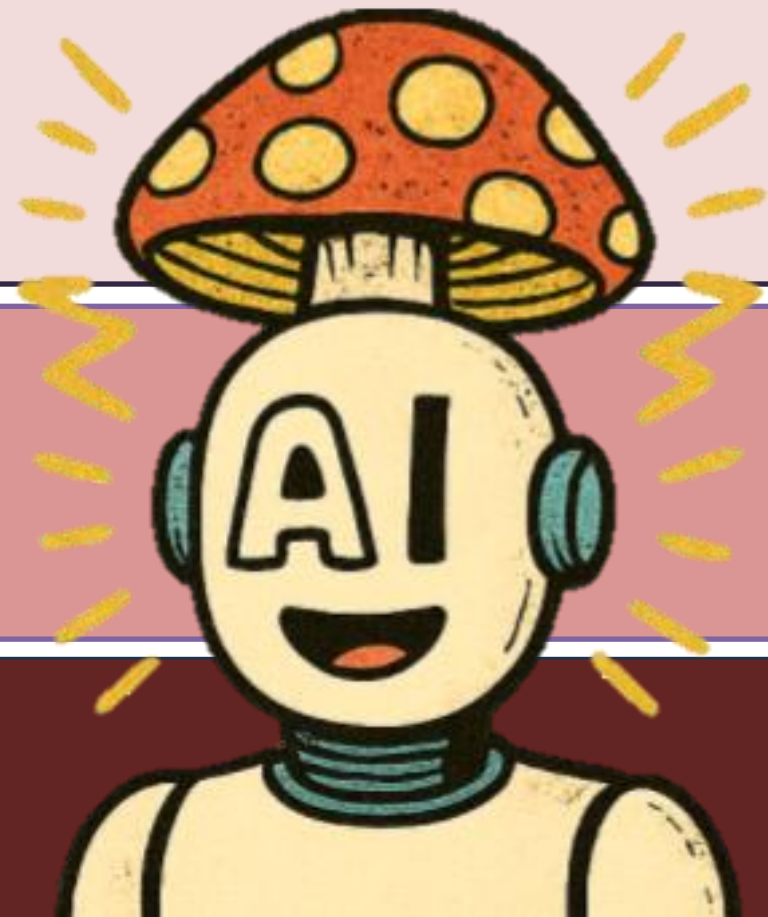
Every important claim should be checked against credible external sources, preferably primary ones: scholarly articles, reports, public institution websites, and reputable industry publications.

Never assume that a source cited by AI actually exists — verify the author, title, year, accessibility, and whether the source truly contains the stated information.

Confirm the information in at least two independent sources.

Open the original source, not only a summary or repost.

Check the publication date and currency of the information.



Source: Copilot Microsoft 365



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

44 of 54

Step 3: Assess quality, language, and context

After checking the facts, assess whether the text is logical, clear, balanced, and appropriate for its purpose. AI may produce content that is formally correct but poorly contextualized, overly confident, biased, or stylistically unsuitable.



The text mixes facts with opinion.



It contains bias, oversimplification, or excessive generalization.

The tone and style fit the audience and publication goal.

Source: Copilot Microsoft 365



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



AI Skills, Powered by Values

45 of 54

Edit, decide, and document

Finally, correct errors, remove uncertain fragments, and decide whether the content is fit for use. Good practice is to record what was changed, what was rejected, and which sources confirmed the final version of the text. **Final decision:**



Publish – if the content has been checked and corrected.

Hold – if some information remains uncertain.

Reject – if the content contains too many errors or lacks source confirmation.





07

Data analysis with AI

(10 min)



When you analyse data with AI, validation is the work



When you ask AI to analyse data, the system will generate a fluent narrative.

Depending on prompt and tool, the output may also:

- ★ Skip the methodology (no mention of sample size, method, or assumptions)
 - ★ Average over outliers without flagging them
 - ★ Generalise from a small sample (e.g., n=10 respondents)
- ★ Produce confident conclusions (where data only supports tentative ones)
- ★ Confirmation bias: AI repeats the framing of your prompt back to you



When you analyse data with AI, validation is the work



Why validating analysis matters:

- ★ You catch errors before they reach a report or publication.
- ★ You distinguish strong findings from confident-sounding noise.
- ★ You avoid drawing decisions from unrepresentative samples.
- ★ You keep authorship of the analysis - not the model.



SAFE Analysis Ladder: choose the most verifiable level

★★★

When you analyse data with AI, do not start at the top of the ladder. Choose the lowest level that still answers your question.

- ★ Level 1 (Most reliable): Ask a general question (no personal data).
Example: “Explain when a t-test is appropriate vs Mann-Whitney U.”
- ★ Level 2: Use an anonymised summary (roles, no identifiers).
Example: “Mean = 4.2, SD = 1.1, n = 120 - what does this suggest about the trend?”
- ★ Level 3: Use limited, de-identified data only if necessary.
Example: “Here is the dataset. Compute X, then I will recheck against my own calculation.”

Stop rule:

If you cannot remove sensitive/confidential elements → do not upload.

★★★



Conclusion from small data

★★★

Unreliable AI output (realistic):

„AI ran sentiment analysis on 200 tweets about our brand. 67% positive. Therefore brand sentiment is strongly positive.”

What is unreliable?

- ★ „67%” with no methodology, no model, no validation
- ★ 200 tweets is a tiny, possibly biased slice
- ★ „strongly positive” overstates what the number shows

★★★





08

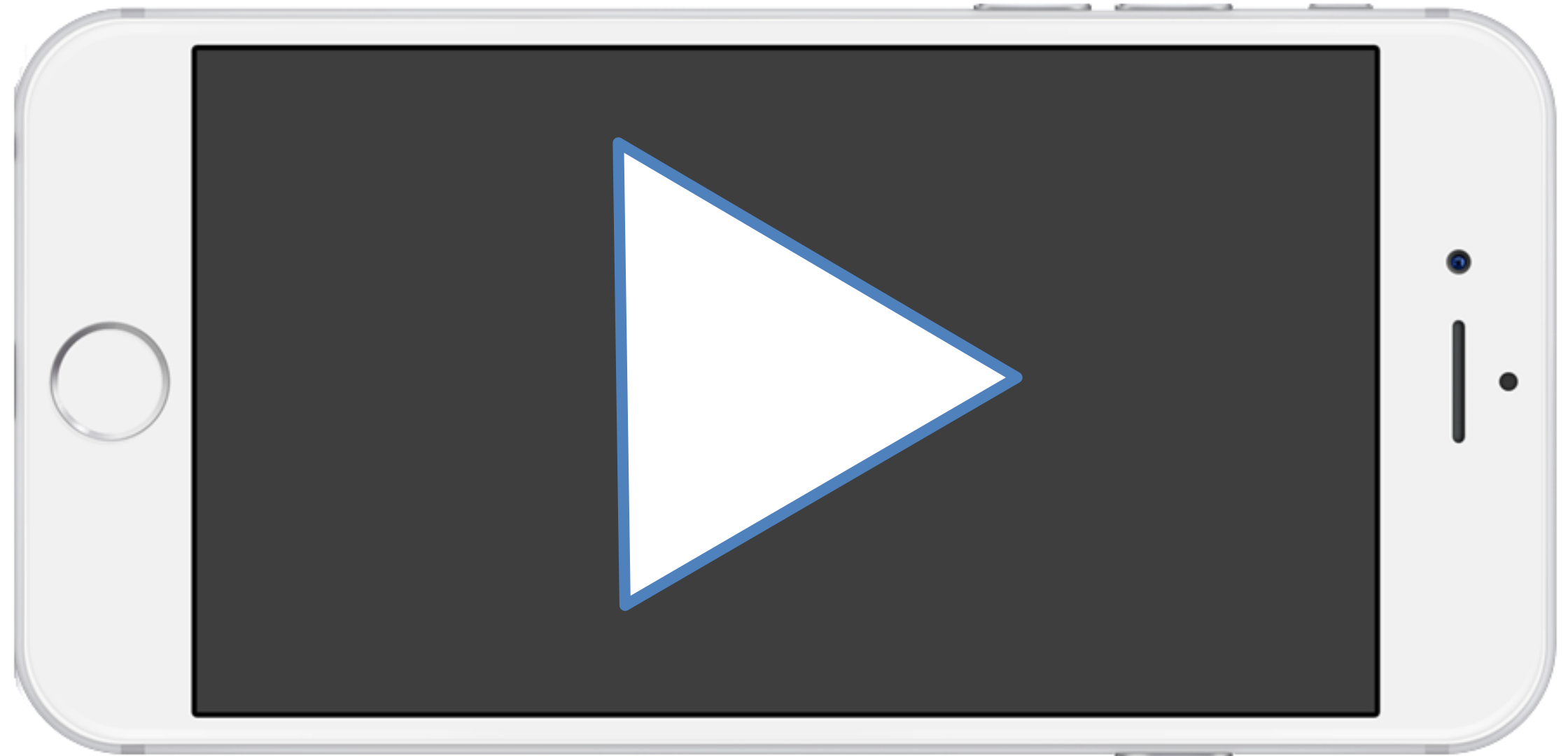
Quiz - Safe Data with AI

(10 min)



Quiz

This quiz helps you practise safe and responsible AI use: what should never be shared, how to respect privacy and consent, how to think about fairness and data quality, and how to choose safer workflows. Important: Do not enter real personal data. Use fictional examples.





Thank

you

www.valuesai.eu

Contact:

www.valuesai.eu



<https://www.linkedin.com/company/values-ai/posts/?feedView=all>



<https://www.facebook.com/ValuesAI>



https://www.instagram.com/values_ai/



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.