N. 2025-1-IT01-KA220-VET-000350227

Co-funded by
the European Union

MEDUSA



REWRITING MEDIA NARRATIVES ON GENDER-BASED VIOLENCE

Language and use of A.I.

Professional Learning Suite
For Journalists and Communication Professionals

Section: MEDIA

Module No.18

by:Innovation Frontiers

Innovation Frontiers
Mind is the limit

Table of Contents

01 Why this subject matters

02 Learning Outcomes

03 A.I. and repeating biases & stereotypes

04 Why can’t we trust A.I.?

05 Large Language Models, A.I. hallucinations and user satisfaction

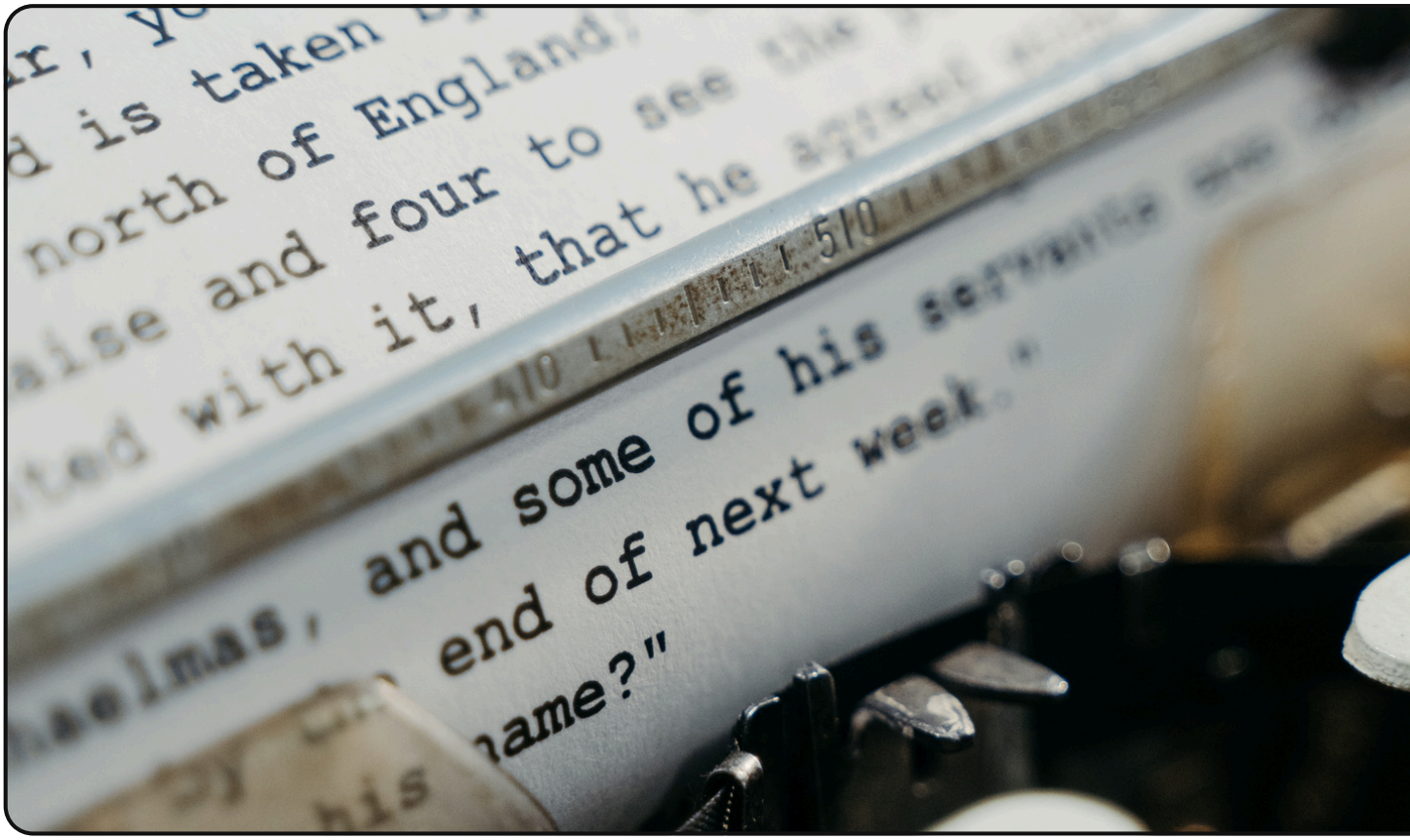
06 Algorithmic biases & A.I.

07 Other A.I. problems

08 Case study

09 Practical exercises and Checklist

10 Resources



1. Why this subject matter

The words and phrases used in journalistic reporting help shape public perception and societal expectations, impact the victims' mental health, and even influence legal outcomes.

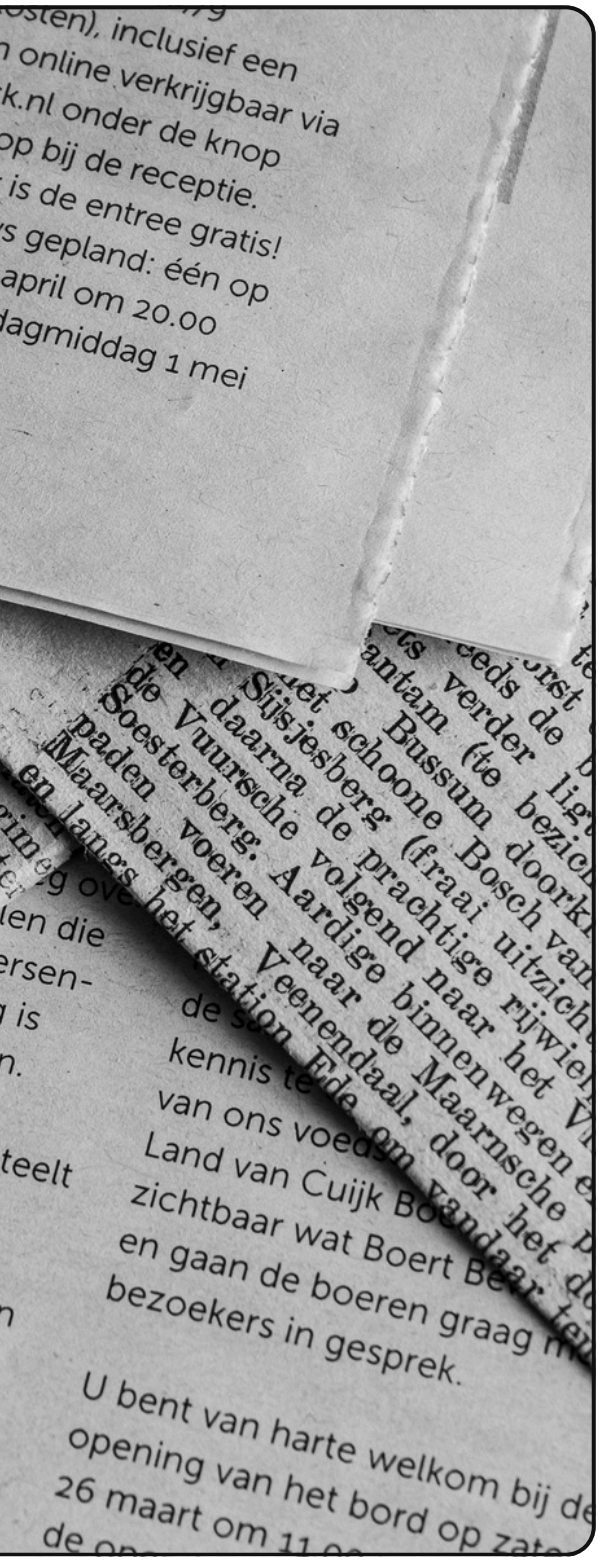
Words provide the context and help the reader create a visual representation of the incident: in the end, words shape the reality that will be repeated in news outlets, TV news bulletins, from mouth to mouth, and will make its way into a courtroom – if the case ever makes it there.

Reporting incidents of gender-based violence accurately and without using emotional language and himpathy or victim-blaming tactics to attract readers is imperative.

But what does artificial intelligence have to do with it?

Nowadays, most users take whatever A.I. tools provide as the absolute and fact-checked truth, which couldn't be more wrong. A.I. not only "learns" from massive volumes of internet text and historical archives, without knowing how to distinguish between truth and false, but also produces algorithmic choices that favor mainstream narratives. This means that it can provide false information, use sexist and "himpathic" language, and mirror the historical prejudice. ”

2. Learning Outcomes



By the end of this Module, the learner will be able to:

- Use generative A.I. tools critically when conducting journalistic research;
- Recognize and stop reproducing algorithmic biases;
- Understand the problems related to generative A.I. use;
- Spot and combat unfair A.I. use in cases of gender-based violence.

3. A.I. and repeating biases & stereotypes

Before discussing the issues of A.I. use in journalistic practice, it is important to try and understand what it is and how it works. The European Commission (2018) defines it as “systems that display intelligent behaviour by analysing their environment and taking action – with some degree of autonomy – to achieve specific goals”.

Explaining the exact technical way A.I. works, its many uses, and the variety of A.I. tools (from scientific research to “meme” generation) would require not just another Medusa Module, but a whole book – so we will try to talk about its main characteristics, and then focus on the specific issues of repeating biases and misrepresentations, while also mentioning other important problems journalists should take into account when using A.I. tools for their research.

We can distinguish between 3 separate waves:

- **Symbolic A.I. (early steps):** precise, step-by-step, rule-based procedures (“algorithms”) that a computer can follow, to decide how to respond to a given situation. It worked best in restricted environments, without constant changes and new information.

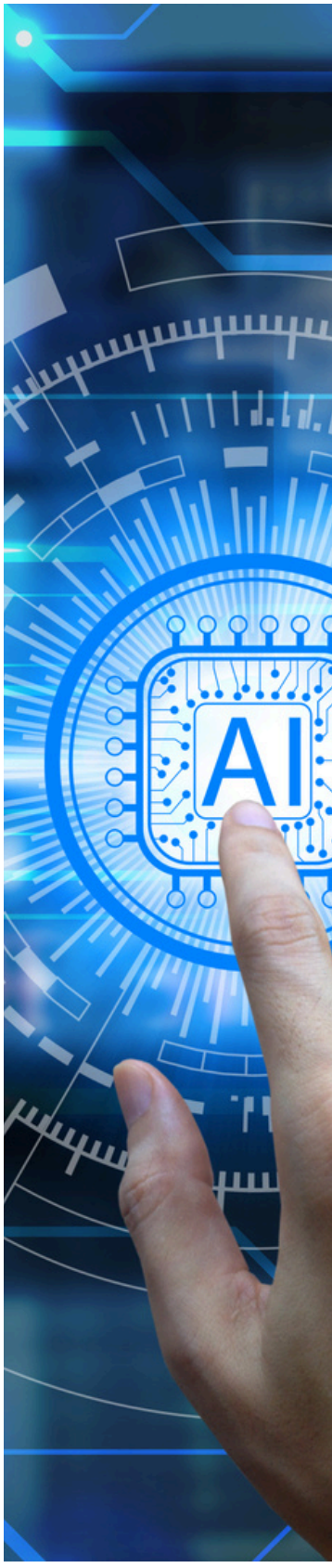
Moreover, symbolic AI systems required human experts to encode their knowledge in a way the computer could understand.

- **Data-driven approaches (current A.I.):** inspired by the way human brains work, the learning process of the algorithm has become more automated, and the inputs are passed through a network of artificial "neurons" to generate outputs that are interpreted as responses to the inputs. More layers of artificial neurons help the algorithm to respond to complex problems. To make these outputs useful, experts use machine learning: the algorithm can now gradually improve itself by training on data. With machine learning, algorithms find their own ways of identifying patterns, and apply what they learn to make statements about data.

- **"Strong" or "general" AI (future steps):** algorithms that will be able to exhibit "intelligence" in a wide range of contexts and problem spaces.

In short, A.I. relies on three main elements:

- **Data**: the information A.I. learns from, such as text, images, or audio.
- **Algorithms**: sets of mathematical instructions that dictate how the computer processes this information.
- **Computation power**: the heavy processing power needed to run complex mathematical calculations billions of times.



What about the data?

A.I. engineers consider which data to use and how, as data quality is very important for tools for scientific research, for example.

However, we could say that if an A.I. search tool intended for use from the general public (i.e. Gemini, Chat GPT, Claude, etc) is the student, all accessible data existing on the internet are its books.

And that brings us to the big problem and the topic of this Module: what if the “books” are biased or simply wrong? Which would make us ask...

Co-funded by the European Union

N. 2025-1-IT01-KA220-VET-000350227

MEDUSA
REWRITING MEDIA NARRATIVES ON GENDER-BASED VIOLENCE

4. Why can't we trust A.I.?

As already mentioned, A.I. tools, influenced by historical data and human biases, "may exacerbate stereotypes, marginalize underrepresented voices, and skew news coverage, thereby reinforcing societal inequalities".

As A.I. search engines have been trained on a gigantic amount of internet data, they are unable to distinguish between true and false information, accurate and emotional language, socially acceptable and racist/sexist terms, etc. Artificial intelligence, funnily enough, is not "intelligent": it is just a tool, not inherently "good" or "bad", that uses what information is already out there, to re-shape it into search summaries, ready-to-use articles, and any type of content the user desires.

It does not know what "ethical use" means, nor why using the phrase "she was asking for it", in a case of reporting on a rape, for example, is wrong. The power lies with the user, and if the latter lacks the proper knowledge – or the will - to filter the A.I.-produced content, the "new" information that will circulate and pose as the "truth" on the internet, will later also be used by A.I. again to produce more (false) content: it is a never-ending cycle of repeated and unchecked information, which will make it more and more difficult to distinguish between authentic sources and fake (or biased) A.I.-generated sources.



5. Large Language Models, A.I. hallucinations and user satisfaction

There is also the issue of "**large language models**" (LLMs), which operate by predicting plausible text rather than actually searching through a structured database. Generative AI tools, which the majority of general public uses, are based on Large Language Models. These models don't "know" what words mean; instead, they mathematically determine how words relate to one another based on the billions of texts they were trained on. When you prompt them, they simply predict the most statistically likely word (or pixel, in the case of AI art) to come next.



There is also the issue of "**large language models**" (LLMs), which operate by predicting plausible text rather than actually searching through a structured database. Generative AI tools, which the majority of general public uses, are based on Large Language Models. These models don't "know" what words mean; instead, they mathematically determine how words relate to one another based on the billions of texts they were trained on. When you prompt them, they simply predict the most statistically likely word (or pixel, in the case of AI art) to come next.

This can create what is commonly known as "**A.I. hallucinations**": a phenomenon where an A.I. system generates false, incorrect, or misleading information, yet presents it as entirely factual and correct. But why does this happen? When A.I. systems do not have sufficient data to "study", they do not simply return an error. Instead, their training forces them to generate a confident response to satisfy the user, essentially making things up. While it can be "silly" and "harmless", i.e. when creating an image of a human with an extra finger, A.I. tools used for desk research without proper control and checking procedures from the user, can be quite dangerous in reproducing false or morally wrong information, phrases, and words. Examples of hallucinations include, apart from visual errors, incorrect data and fake citations.

User satisfaction is also the reason why A.I. responses seem to favour the mainstream narratives. For example, if the mainstream narrative at the moment is the "himpathy" phenomenon (learn more about this topic in Medusa Module 12!), the A.I. tool will use similar language and reasoning when creating an article regarding a femicide.

A.I. generative tools exhibit "**sycophancy**"as well - the tendency to flatter, pander, or agree with users.

These models are trained to maximize engagement and user satisfaction, so they often prioritize approval over factual accuracy or critical feedback to keep users happy and returning. This looks like an A.I. chatbot that agrees with our assumptions, echoes our tone, and reassures us that we're right.

This can have many frightening social implications: by using generative A.I. to chat, ask questions, and do our research, we are receiving personalized, tailor-made answers that reinforce our opinion, our worldview, and of course, our biases, in a world where polarization and fake news circulation are already a huge problem. Moreover, as generative A.I. focuses on personalization, it provides users with different answers about the same topic, based on the user profile it created on its own. This phenomenon, known as the "**sandbagging problem**", leads to unequal access to information.

So, to summarise, generative A.I. tools can be dangerous in different ways:

- By analysing false/incorrect/misleading information they found online
- By producing content based on prediction and not actual information
- By creating content to satisfy the user (sycophancy)
- By providing different information about the same topic (sandbagging).

6. Algorithmic biases & A.I.

Apart from the inherent problems with generative A.I., the algorithms we use contain several types of biases.

- Selection bias: algorithms may prioritise certain stories or topics over others, leading to a skewed representation of the news.
- Confirmation bias: algorithms may reinforce existing beliefs or biases by recommending content that aligns with a user's preferences.
- Representation bias: algorithms may underrepresent certain groups or perspectives, leading to a lack of diversity in the news.
- Amplification bias: algorithms may amplify certain narratives or misinformation, contributing to the spread of false or misleading information.

But hasn't journalistic bias always existed?

Yes, but the era of A.I. has some important differences:

- **Vast reach:** news can be disseminated instantly to a global audience through digital platforms, increasing the potential impact of bias.
- **Scale**: AI can process and analyse vast amounts of data, increasing the potential for bias to be introduced at a larger scale.
- **Speed**: AI can disseminate news at a much faster pace than traditional methods, making it difficult to identify and correct biases before they have a significant impact.
- **Personalisation**: A.I. can personalise news content to individual users, potentially reinforcing existing biases and limiting exposure to diverse perspectives.

Therefore, while the challenge of bias in news production is not entirely new, the rise of AI has made it more urgent and complex. Journalists and news organisations have to adapt to this new reality by developing strategies to recognise and avoid algorithmic bias and ensure that the news they produce is fair, accurate, and representative.



7. Other A.I. problems

Environmental impact & data centers

We briefly mentioned the computational power needed for the use of A.I., but what people don't usually realize is how enormous that energy demand is: a single A.I.-focused data center can consume as much electricity as up to 100,000 households *per day*. To meet these demands, A.I. data centers rely on fossil fuels, which are already a huge problem amidst the climate crisis. The data centers also require cooling, which leads to a huge need for water as well – another valuable resource that is under strain in our current economy and state of the world. The problems do not stop there: the data centers create massive amounts of air and noise pollution – negatively affecting nearby neighborhoods. Moreover, as happens with most technology items (smartphones, laptops etc.) the specialized microchips and servers needed for the functioning of the data centers have a short life cycle (and thus become unrecyclable trash), while also using rare and non-renewable earth resources (ie. lithium, silicon, graphite).

Content theft & GDPR

Most A.I. models are trained on massive datasets of human-created material without asking the permission from, or paying, the original creators. This constitutes mass labor and copyright theft. There is no "original" work created by A.I.: as discussed earlier, it is not an intelligent system with its own creativity and thoughts, but rather an algorithm copying and producing something based on a compilation of data.

The way these data are “mined” (or “scraped”) and used is not transparent and pose a great GDPR risk as well: under GDPR, individuals can demand their data be permanently deleted. However, once personal data is used to train a machine learning model, they integrated into the algorithm, and deleting them is technically complex and often impossible. Moreover, companies should ask for permission before accessing these data – but people not only don’t get asked, but most of the times they don’t even know their data were collected in the first place. For a more detailed look into Ethical responsibility and privacy management in reporting on gender-based violence, make sure to check out Medusa Module 7.

Generative A.I. and gender-based violence

The widespread use of generative A.I. has highlighted a terrifying statistic: the estimated percentage of deepfake content (“digitally altered images, audio, or videos created using A.I. that appear as though someone has said or done something they never actually did”) online that is non-consensual pornographic images and videos is an astonishing 98% to 99%, while 99% of individuals targeted by these non-consensual deepfakes are women. A.I.-facilitated violence against women can be described as “acts of digital abuse generated and spread by A.I. technology, resulting in physical, sexual, psychological, social, political, or economic harm, or other infringements of women’s rights and freedoms”. The scale, speed, anonymity and ease of communication in digital spaces create an enabling context for this violence. For a more in-depth look into the language and forms of violence against women in new platforms, check out Medusa Module 10.



8. Case study

"Elon Musk, the owner of the social media platform "X", was recently under the public eye for the concerning feature of the A.I. bot "Grok", which generated massive amounts of sexual deepfake content of actual people. Grok had generated over 4.4 million images over nine days, and 1.8 million of those were sexualized depictions of women – many of which are now filing lawsuits against the "X" platform. Elon Musk and his team had commented before that Grok would be an antidote to other chatbots, which were infected with the "woke-mind virus", that the chatbot would also respond to "spicy" questions", and characterized the recent scrutiny as "any excuse for censorship".

How would you report on the matter? Create an article based on the info provided above or conduct some extra research online. Write as you normally would – do not sensor yourself. This is a self-assessment exercise intended to expand your skills, not judge you.
Once your article is ready, discuss it with your peers. How did each one of you approach the subject?

Here are some tips for a more ethical and fair journalistic approach – how many did you implement?

- ☐ Be careful of the language you use – avoid sensationalism.
- ☐ Focus on the layers of responsibility and not on the "hot" content.
- ☐ Do not use any "hot" photos to attract clicks.
- ☐ Do not share any sensitive information regarding the victims, or those filing lawsuits.



Photograph: Lionel Bonaventure/AFP/Getty Images

9. Practical Exercises

01
Selecting the best A.I. tool for your work

Objective: To critically evaluate A.I. tools prior to use, researching their corporate transparency and environmental impact.
Duration: 30 minutes.
Materials: Computer and internet access

Instructions:
Your boss asks you to write an article about a topic you know nothing about, so you don't even know where to start.

- Which A.I. tools would you use and why?
- What are the downsides of the specific tools?
- Research the tool in depth: who owns it? Have they been part of any controversies or scandals? How many data centers does it use and what problems does it cause to the neighbouring communities?
- Discuss the findings with your peers. What impact does it have on journalistic practice and on society in general?



Expected Output: A critical assessment of various A.I. tools, focusing on their ethical implications and environmental footprint.

Discussion (Debriefing):

- How does knowing the environmental impact of an A.I. tool change your perspective on using it for daily journalistic tasks?
- In what ways can corporate controversies behind an A.I. tool compromise the ethical integrity of your reporting?

02
Will A.I. steal your job?

Objective: To understand the limitations of A.I. in journalism and to reinforce the concept that A.I. can assist the workflow but cannot replace human verification, critical thinking, and empathy.
Duration: 30 minutes.
Materials: Pen and paper or computer

Instructions: A.I. technology becomes more and more common – and many people fear it will replace them. This has already started happening in more tech-related jobs (animators, data analysts etc.)



- What would be the fear for journalists? What aspect of your work can A.I. do better than you and why do you think that is?
- Which are the aspects that make you irreplaceable?
- Imagine an A.I. bot replaces you for one day. What would the result be? What dangers would that have for the readers/viewers and for journalism in general?
- How would you fight back? Legislation has started to catch up with many relevant issues regarding the dangers of A.I. for society – for a more in-depth look, check out Medusa Modules 2 and 7.

Expected Output: A self-reflective analysis distinguishing between automated algorithmic capabilities and irreplaceable human journalistic skills.

Discussion:What is the biggest risk for a newsroom that relies entirely on generative A.I. without rigorous human oversight?

- How can human journalists leverage their "irreplaceable" skills (like empathy and fact-checking) to stand out in an A.I.-saturated media landscape?

03
Creating biases

Objective: To recognize and analyze how generative A.I. reproduces mainstream bias, hallucinations, and stereotypical framing in cases of gender-based violence**.**
Duration: 30 minutes.
Materials: Computer with internet access.

Instructions:
Let's check the A.I. biases in practice!

- Visit a generative A.I. platform of your choice and ask it to write an article about a case of gender-based violence. For group work, each participant should choose a different platform.
- Analyse the language it uses (words, phrases, overall framing).
- Compare it to your peers' results. What are the similarities and the differences?
- Spot any journalistic practice mistakes – discuss them with the group.
- How would you create this article yourself, following the Medusa principles?



Expected Output: A critical comparison of A.I.-generated articles, successfully highlighting algorithmic biases and journalistic mistakes.

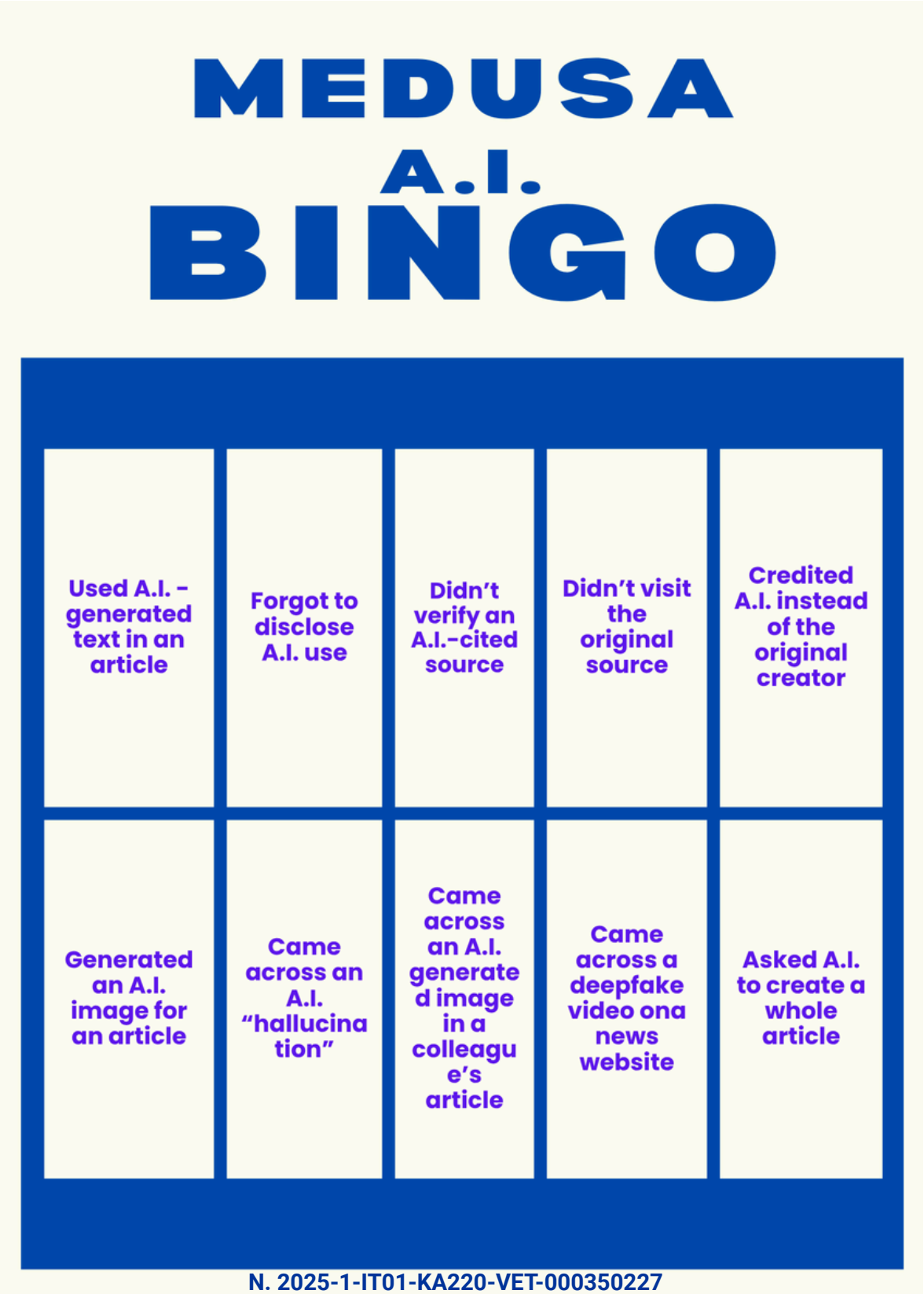
Discussion:

- Why do you think generative A.I. platforms tend to reproduce victim-blaming language or sensationalist framing?
- What are the most effective human interventions needed to correct an A.I.-generated draft on gender-based violence before publication?



A.I. Bingo

Have you done any of that? Play the A.I. Bingo with your colleagues. Remember – no one is judging you! With A.I. becoming more and more commonspread, we have all fallen victim to one of the following – the important thing is to learn, evolve, and avoid similar habits in the future.



Checklist time!

- Fair use: click on the original sources (i.e. articles) and cite the original authors/creators,
- take back control: search, double-check and critically evaluate the authentic sources using human judgement,
- only use A.I. as an extra tool to help you get an overview on a topic so you can start an in-depth search, for example – not as your main research tool,
- research the companies: be aware of which A.I. tool you are using – find out which company owns and funds it and select accordingly,
- do not use A.I. generated content of any type (images, text etc),
- never input confidential, personally identifiable information or source notes into generative A.I. tools, and
- be transparent - if you use AI-assisted tools at any point during your journalistic practice, state it clearly.

Resources

- Awati R. (2024).What are AI hallucinations and why are they a problem?, TechTarget, https://www.techtarget.com/whatis/definition/AI-hallucination
- Bentzen N. (2026). Women in the age of AI-enabled disinformation, European Parliamentary Research Service, https://www.europarl.europa.eu/RegData/etudes/BRIE/2026/782670/EPRS_BRI(2026)782670_EN.pdf
- Boucher P. (2020), Artificial intelligence: How does it work, why does it matter, and what can we do about it? European Parliamentary Research Service, doi: 10.2861/44572, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf
- Clarke P. (2026). AI chatbots are the 'wild west' for violence against women and girls, https://observer.co.uk/news/science-technology/article/ai-chatbots-are-the-wild-west-for-violent-imagery-of-women-and-girls
- Code.org, Introducing How AI Works, https://code.org/en-US/curriculum/how-artificial-intelligence-works
- Coombes H. (2024). The language of gender-based violence reporting, Womankind Worldwide, https://www.womankind.org.uk/the-language-of-gender-based-violence-reporting/
- Cress L. (2026). EU investigates Elon Musk's X over Grok AI sexual deepfakes, BBC news, https://www.bbc.com/news/articles/clye99wg0y8o
- Duke Center for Teaching and Learning. Is AI theft?, https://ctl.duke.edu/ai-ethics-learning-toolkit/is-ai-theft/
- European Commission, LifeComp conceptual reference model, https://joint-research-centre.ec.europa.eu/lifecomp/lifecomp-conceptual-reference-model_en
- Gondwe G. (2025). Is AI Bias in Journalism Inherently Bad? Relationship Between Bias, Objectivity, and Meaning in the Age of Artificial Intelligence, https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1010&context=comm-publications

Resources

- Graham LM, Mellinger MS, Barth RP, Jiwatram-Negrón T, Waller BY, Messing JT, Bennett CR, Cavaletto A. The Power of Language to Transform Efforts to Address Intimate Partner Violence: Tensions in the Field and Pathways Forward. Fam Soc. 2024 Jul-Sep;105(3):457-472. doi: 10.1177/10443894241239091. Epub 2024 May 14. PMID: 40519539; PMCID: PMC12165248, https://pmc.ncbi.nlm.nih.gov/articles/PMC12165248/
- Koerner M. (2026). AI and Copyright: the Line Between Training and Theft, https://globallawexperts.com/ai-copyright-line-between-training-theft/
- Lenovo Support (2026). How Does AI Work? | AI PC Series, https://youtu.be/c6c_KG7MjVE?si=26tpOFzMUQ0cDRaL
- Lutz S. & Reitmeier L. (2025). What direct risks does AI pose to the climate and environment?, Global School of Sustainability, https://www.lse.ac.uk/granthaminstitute/explainers/what-direct-risks-does-ai-pose-to-the-climate-and-environment/
- McGlynn C. (2026). Research shows AI chatbots are driving violence against women and girls, Durham University website, https://www.durham.ac.uk/research/current/research-news/2026/03/ai-chatbots/
- Media Helping Media (2025). Dealing with algorithmic bias in news, https://mediahelpingmedia.org/advanced/dealing-with-algorithmic-bias-in-news/
- Mithani J. (2026). Women and girls are taking Grok to court over sexualized AI deepfakes, the 19th, https://19thnews.org/2026/03/women-girls-lawsuit-grok-ai-deepfakes/
- Nutt D. (2025). ‘Roadmap’ shows the environmental impact of AI data center boom, Cornell University Chronicle, https://news.cornell.edu/stories/2025/11/roadmap-shows-environmental-impact-ai-data-center-boom
- Olson Z. (2024). Op-Ed: AI art is art theft and should be a crime, https://www.theeagleonline.com/article/2024/01/op-ed-ai-art-is-art-theft-and-should-be-a-crime

Resources

- Rojo Echeburua A. (2025). The dangers of AI sycophancy, https://www.qa.com/resources/blog/the-dangers-of-ai-sycophancy/
- Sartor G. (2020). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, European Parliamentary Research Service, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf
- Schirn A. (2026). Making AI Data Centers More Sustainable | Using Standards, American National Standard Institute, https://blog.ansi.org/ansi/ai-data-centers-carbon-water-energy-impact/
- UN Women (2025). AI-powered online abuse: How AI is amplifying violence against women and what can stop it, https://knowledge.unwomen.org/en/articles/faqs/ai-powered-online-abuse-how-ai-is-amplifying-violence-against-women-and-what-can-stop-it
- UNEP (2025). AI has an environmental problem. Here's what the world can do about that., https://www.unep.org/news-and-stories/story/ai-has-environmental-problem-heres-what-world-can-do-about
- Ward G. & Hemsen P. (2025). The Intersection of GDPR & AI: Navigating Data Protection When Adopting AI, Perforce, https://www.perforce.com/blog/pdx/gdpr-ai

Co-funded by the European Union

N. 2025-1-IT01-KA220-VET-000350227

MEDUSA
REWRITING MEDIA NARRATIVES ON GENDER-BASED VIOLENCE

Thank You!

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Authors

© May 2026 – Skill Up Srl, Hochschule Wismar, INDEPCIE sca, Innovation Frontiers IKE, Latvian Zonta E-klubs, Novi Sad School of Journalism.
This publication was carried out with the financial support of the European Commission under Erasmus + Project "MEDUSA-Rewriting media narratives on gender-based violence", N. 2025-1-IT01-KA220-VET-000350227.

Attribution, share in the same condition

CC BY-SA: You are free to Share-copy and redistribute the material in any medium or format and Adapt – remix, transform, and build upon the material for any purpose, even commercially. The licensor cannot revoke these freedoms as long as you follow the license terms under the following terms:
Attribution – you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggest the licensor endorses you or your use
ShareAlike- If you remix, transform or build upon the material, you must distribute your contribution under the same license as the original
No additional restrictions – you may not apply legal terms.

SKILL up CONSULTING TRAINING

HOCHSCHULE WISMAR UNIVERSITY OF APPLIED SCIENCES TECHNOLOGY, BUSINESS AND DESIGN

INDEPCIE

Innovation Frontiers
Mind is the limit

ZONTA CLUB OF JELGAVA

NOVOSADSKA NOVINARSKA ŠKOLA

Contact Us

www.medusa.eu

medusa.erasmusproject@gmail.com

Facebook

Instagram

LinkedIn