
Enhancing Critical Thinking through GenAI: Student-Centred Approaches to Ethical GenAI Use in the Humanities



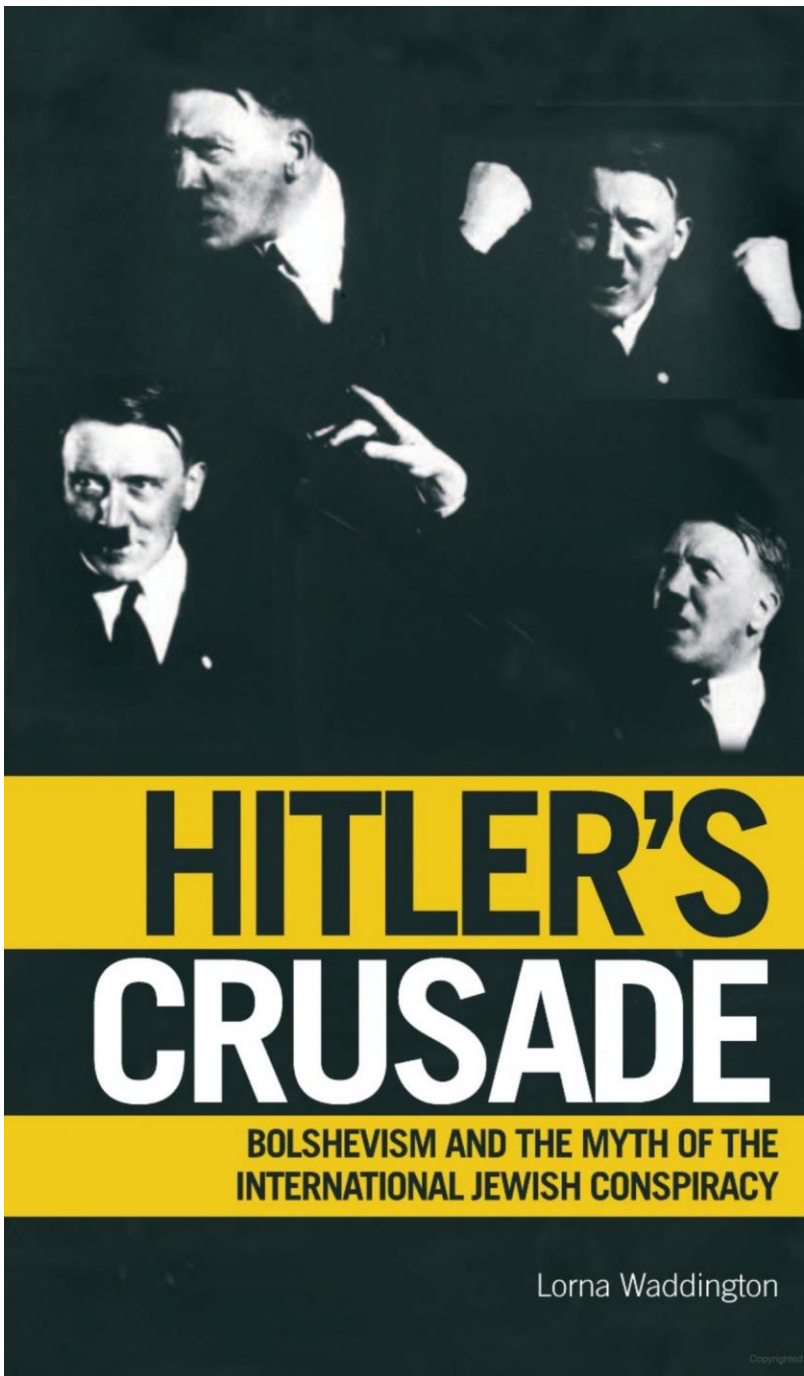
*Dr Lorna Waddington,
Associate Professor of International History
University Academic Integrity Lead*



UNIVERSITY OF LEEDS

Overview

- Introduction
- ‘Theory into practice’
- Outline of sessions to take place in a level 1 skills module



Academic Integrity Network - Contract Cheating Working Group

Cheating detection for markers: only found in assignments produced t

le red flags.
n to include:

take to

ate differs

Comparison with the Student's Previ

By comparing a piece of work that is suscep
with other work submitted by the same stud
between them that we would not expect to
Aspects of a student's writing style that a mark

- ❑ Differences in referencing style and format
- ❑ Differences in writing style, such as paragra
paragraphs, line spacing, inconsistent use o
ibid.
- ❑ Significant improvement in written expressi

Comparison to Cohort

- ❑ If a piece of work makes un
discipline, it may be worthy c

Related Considerations

- ❑ Many universities have specific procedures to foll
misconduct, including suspected contract cheat
are at your specific institution, please seek advice
- ❑ It is important to distinguishing between legitimate

ions,

h in European Conference on Academic Integrity and

and 'Conversations.'" *International Journal for Educational*

substantiating-contract-cheating-guide-investigators.pdf
isplay.aspx?DocID=50362

This checklist has been put t

- Stephen Bunbury, Quality
- Dr Mary Davis, Academic
- Dr Matthew Jones, Senior
Greenwich
- Anna Krajewska, Director
Institute London
- Olumide Popoola, Educa
- Dr Lorna Waddington, Lec
- Dr Louise Revell, Associate

International Journal for Educational Integrity

Home About Articles Submission Guidelines Submit manuscript

Original article | [Open access](#) | Published: 25 December 2023

Testing of detection tools for AI-generated text

[Debora Weber-Wulff](#), [Alla Anohina-Naumeca](#), [Sonja Bjelobaba](#) , [Tomáš Foltýnek](#), [Jean Guerrero-Dib](#), [Olumide Popoola](#), [Petr Šigut](#) & [Lorna Waddington](#)

International Journal for Educational Integrity **19**, Article number: 26 (2023) | [Cite this article](#)

40k Accesses | 742 Altmetric | [Metrics](#)

Abstract

Recent advances in generative pre-trained transformer large language models have emphasised the potential risks of unfair use of artificial intelligence (AI) generated content in an academic environment and intensified efforts in searching for solutions to detect such content. The paper examines the general functionality of detection tools for AI-generated text and evaluates them based on accuracy and error type analysis. Specifically, the study seeks to answer research questions about whether existing detection tools can reliably differentiate between human-written text and ChatGPT-generated text, and whether machine translation and content obfuscation techniques affect the detection of AI-generated text. The research covers 12 publicly available tools and two commercial systems (Turnitin and PlagiarismCheck) that are widely used in the academic setting. The researchers conclude that the available

the Killing Fields of Cambodia

Dr Lorna Waddington

School of History, University of Leeds, Leeds, UK

Abstract

Drawing upon documents from the British National Archives, this article investigates how and why the British government formulated its response to the Khmer Rouge dictatorship in Cambodia. This regime was responsible for some of the worst mass killings in the twentieth century; indeed, in April 1978, President Jimmy Carter referred to it as the “worst violator of human rights in the world today.”¹ The British have been portrayed, not least by themselves, as championing the international opposition to human rights abuses in Democratic Kampuchea. The fact that Britain was the first country publicly to condemn the violation of human rights by raising the issue at the United Nations Commission on Human Rights (UNCHR) in 1978 cannot be denied.² However, a closer examination of the British reaction to the activities of the Khmer Rouge



The humanities must have a role in overseeing AI 'censorship'

Restrictions on what models will discuss are necessary, but ill-informed b' distort inquiry, say Lorna Waddington and Richard de Blacquiere-Clarksc

Published on July 4, 2025

Last updated July 4, 2025

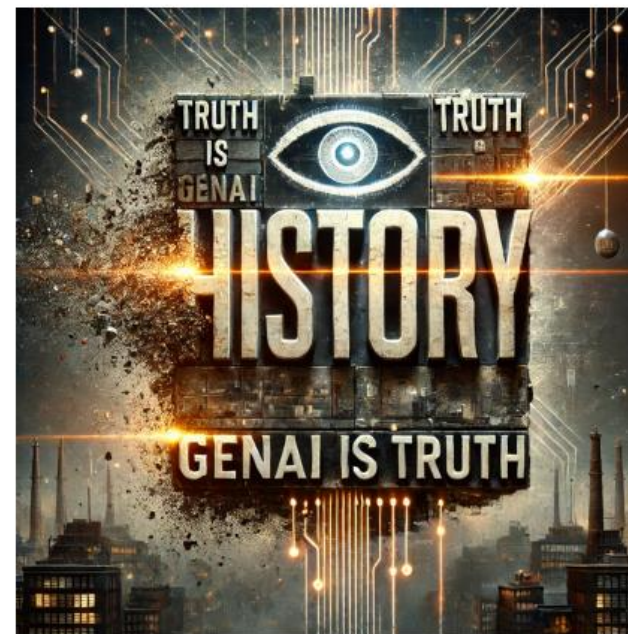
[Lorna Waddington](#) [Richard de Blacquiere-Clarkson](#)

In May 2025, xAI's Grok 3 artificial intelligence chatbot began producing [unsolicited references](#) to violence against white people in South Africa, including the



Preparing students for a world shaped by artificial intelligence

Readers respond to a letter which said that AI is undermining learning at universities



[Articles](#) [Submission Guidelines](#)

[Submit manuscript](#)

[Open access](#) | Published: 02 July 2025

Maintaining research integrity in the age of GenAI: an analysis of ethical challenges and recommendations to researchers

[Sonja Bjelobaba](#), [Lorna Waddington](#), [Mike Perkins](#), [Tomáš Foltýnek](#), [Sabuj Bhattacharyya](#) & [Debora Weber-Wulff](#)

[International Journal for Educational Integrity](#) **21**, Article number: 18 (2025) | [Cite this article](#)



Background

- GenAI tools are becoming increasingly widespread
- Common GenAI limitations
- Students need to develop critical skills to assess GenAI outputs
- Personal competencies





Higher Education Policy Institute

Student Generative AI Survey 2025

Josh Freeman

Kortext

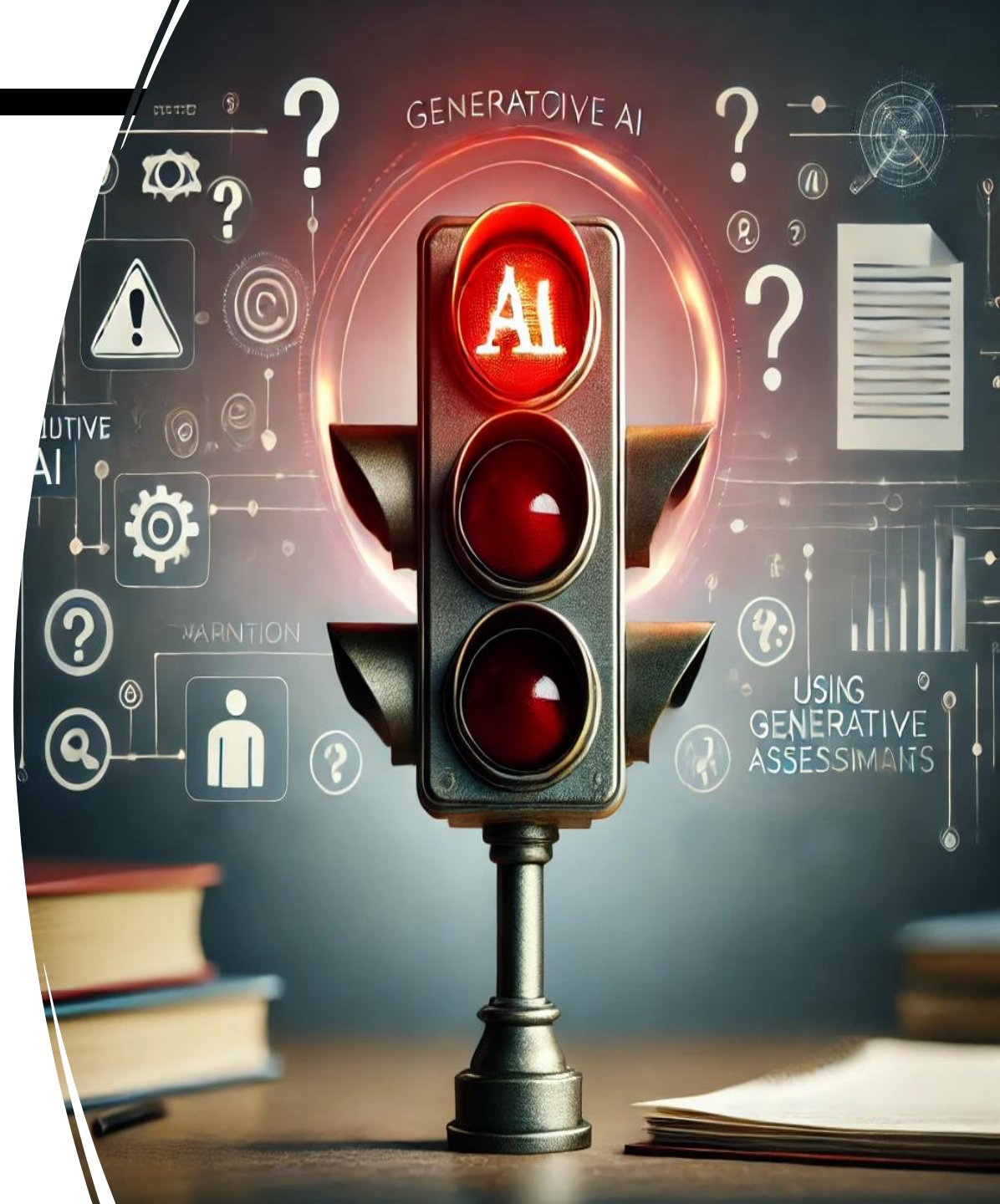
HEPI Policy Note 61

February 2025

In 2025, we find that the student use of AI has surged in the last year, with almost all students (92%) now using AI in some form, up from 66% in 2024, and some 88% having used GenAI for assessments, up from 53% in 2024. The main uses of GenAI are explaining concepts, summarising articles and suggesting research ideas, but a significant number of students – 18% – have included AI-generated text directly in their work.

The Problem(s) with Red

- Guidelines for the use of red
- How to prove a student used GenAI?
- Chat detection tools don't work
- False positives



Chat Detection tools don't work

Original article | [Open access](#) | Published: 25 December 2023

Testing of detection tools for AI-generated text

[Debora Weber-Wulff](#), [Alla Anohina-Naumeca](#), [Sonja Bjelobaba](#) , [Tomáš Foltýnek](#), [Jean Guerrero-Dib](#), [Olumide Popoola](#), [Petr Šigut](#) & [Lorna Waddington](#)

International Journal for Educational Integrity **19**, Article number: 26 (2023) | [Cite this article](#)

41k Accesses | **47** Citations | **742** Altmetric | [Metrics](#)

Abstract

Recent advances in generative pre-trained transformer large language models have emphasised the potential risks of unfair use of artificial intelligence (AI) generated content in an academic environment and intensified efforts in searching for solutions to detect such content. The paper examines the general functionality of detection tools for AI-generated text and evaluates them based on accuracy and error type analysis. Specifically, the study seeks to answer research questions about whether existing detection tools can reliably differentiate between human-written text and ChatGPT-generated text, and whether machine translation and content obfuscation techniques affect the detection of AI-generated text. The research covers 12 publicly available tools and two commercial systems (Turnitin and PlagiarismCheck) that are widely used in the academic setting. The researchers conclude that the available detection tools are neither accurate nor reliable and have a main bias towards classifying the output as human-written rather than detecting AI-generated text. Furthermore, content

False Accusations

GPT-2 Output Detector Demo

This is an online demo of the GPT-2 output detector model, based on the 🤖/Transformers implementation of RoBERTa. Enter some text in the text box; the predicted probabilities will be displayed below. [The results start to get reliable after around 50 tokens.](#)

[1] In the beginning God created the heaven and the earth.
[2] And the earth was without form, and void; and darkness was upon the face of the deep. And the Spirit of God moved upon the face of the waters.
[3] And God said, Let there be light: and there was light.
[4] And God saw the light, that it was good: and God divided the light from the darkness.
[5] And God called the light Day, and the darkness he called Night. And the evening and the morning were the first day.
[6] And God said, Let there be a firmament in the midst of the waters, and let it divide the waters from the waters.
[7] And God made the firmament, and divided the waters which were under the firmament from the waters which were above the firmament: and it was so.
[8] And God called the firmament Heaven. And the evening and the morning were the second day.
[9] And God said, Let the waters under the heaven be gathered together unto one place, and let the dry

Real	Prediction based on 258 tokens	Fake
0.22%	<div></div>	99.78%

Humanize Your Ai Writing in Seconds

Use our cutting-edge AI detection tool to scrutinize your text for AI footprints. If detected, effortlessly click "Humanize" to upgrade your content to human writing quality. Our AI text humanizer ensures seamless integration while bypassing AI detectors.

I have, myself, full confidence that if all do their duty, if nothing is neglected, and if the best arrangements are made, as they are being made, we shall prove ourselves once again able to defend our Island home, to ride out the storm of war, and to outlive the menace of tyranny, if necessary for years, if necessary alone. At any rate, that is what we are going to try to do. That is the resolve of His Majesty's Government-every man of them. That is the will of Parliament and the nation. The British Empire and the French Republic, linked together in their cause and in their need, will defend to the death their native soil, aiding each other like good comrades to the utmost of their strength. Even though large tracts of Europe and many old and famous States have fallen or may fall into the grip of the Gestapo and all the odious apparatus of Nazi rule, we shall not flag or fail. We shall go on to the end, we shall fight in France, we shall fight on the seas and oceans, we shall fight with growing confidence and growing strength in the air, we shall defend our Island, whatever the cost may be, we shall fight on the beaches, we shall fight on the landing grounds, we shall fight in the fields and in the streets, we shall fight in the hills; we shall never surrender, and even if, which I do not for a moment believe, this Island or a large part

⋮
WAITING FOR
YOUR INPUT

⚠ Your content is detected as written by AI ⚠



I Agree to the terms of service
(No academic misconduct)

4.9/5 Based on 12,623 REVIEWS



Check AI Source

Humanize

AI DETECTION LIKELIHOOD



AI2H



GPTZERO



OPENAI



WRITER



CROSSPLAG



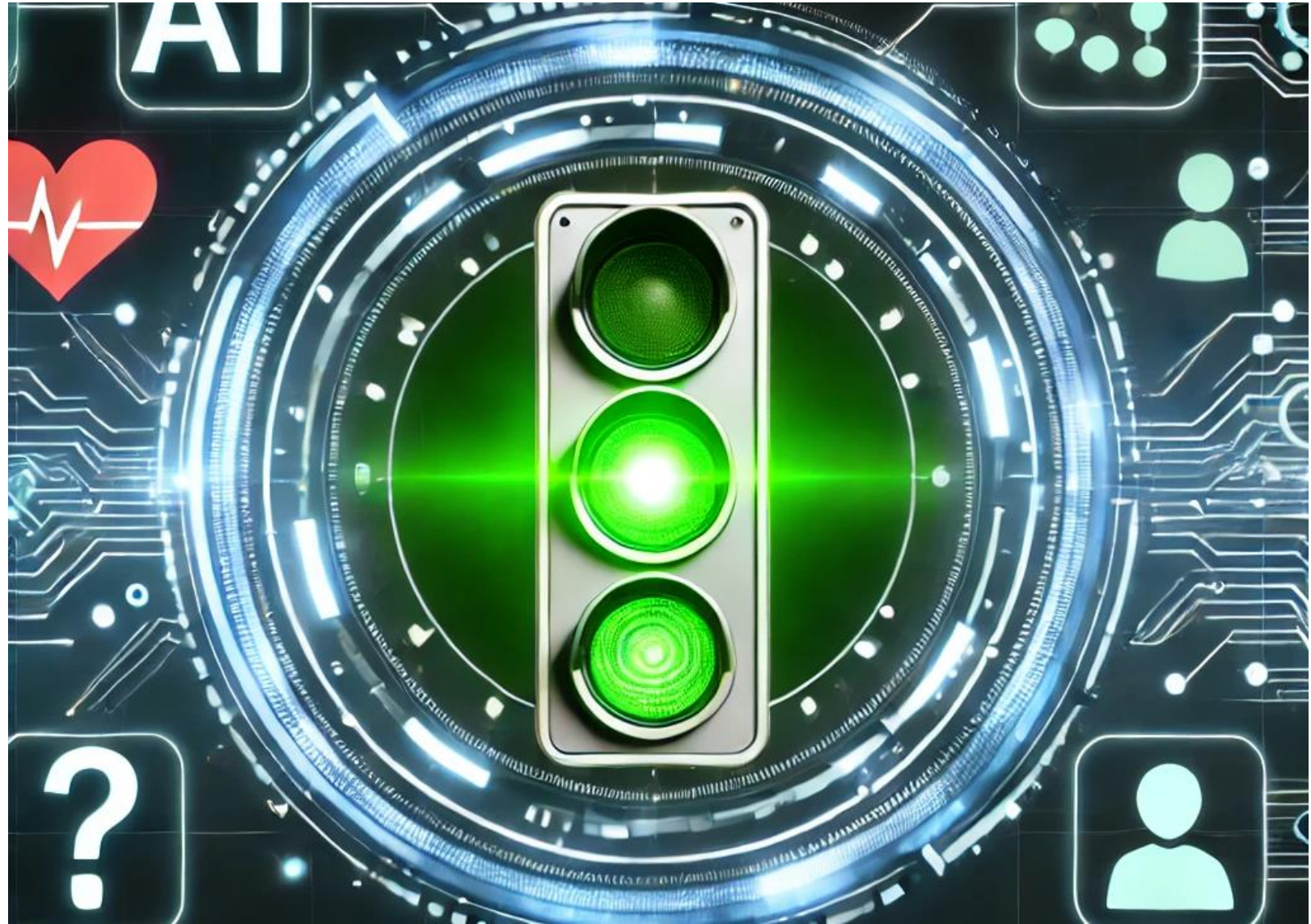
CONTENTATSCALE



ZEROGPT

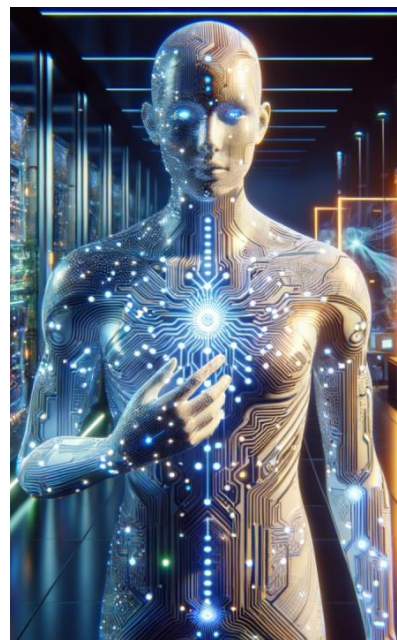
KEY: 100% HUMAN 50% HUMAN 0% HUMAN

The Problem(s) with Green





X.AI INTRODUCING GROK



GOOLE'S AI MO

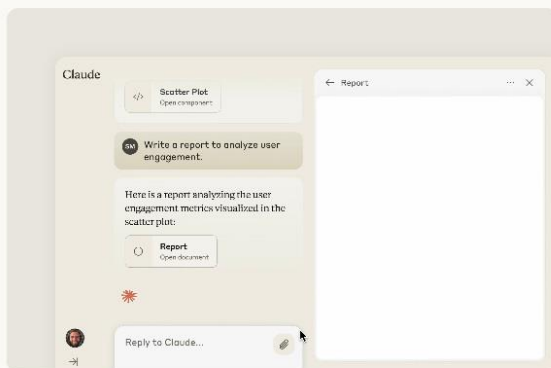
Gemini



ChatGPT

Meet Claude

Claude is a next generation AI assistant built by Anthropic. It's safe, accurate, and secure to help you do your work.



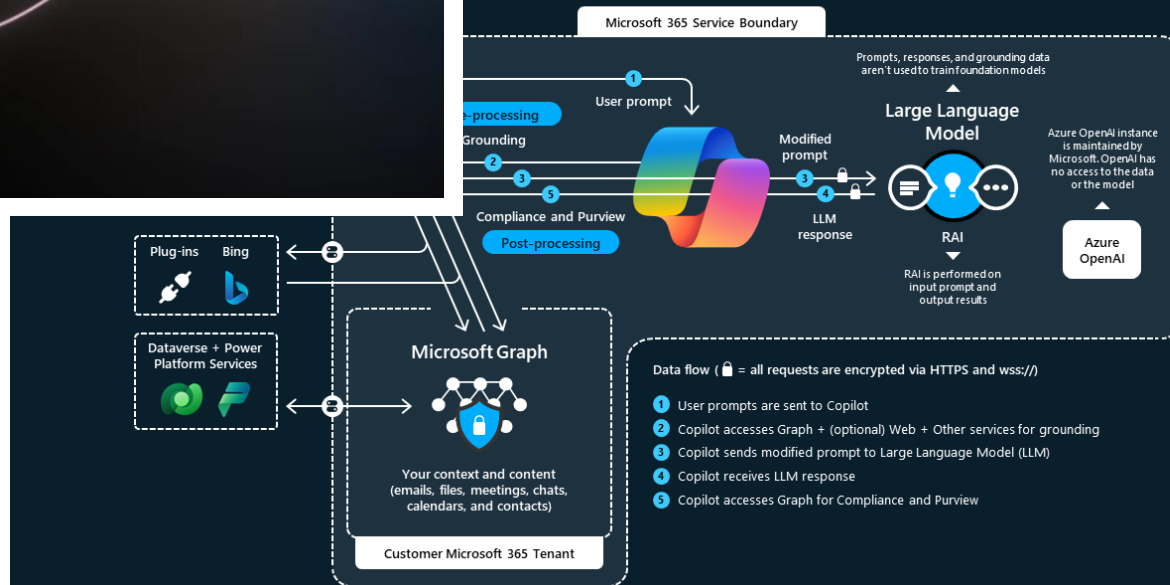
Create with Claude

Draft and iterate on websites, graphics, documents, and code alongside your chat with Artifacts.

Bring your knowledge

Share and collaborate with your team

Copilot for Microsoft 365 architecture

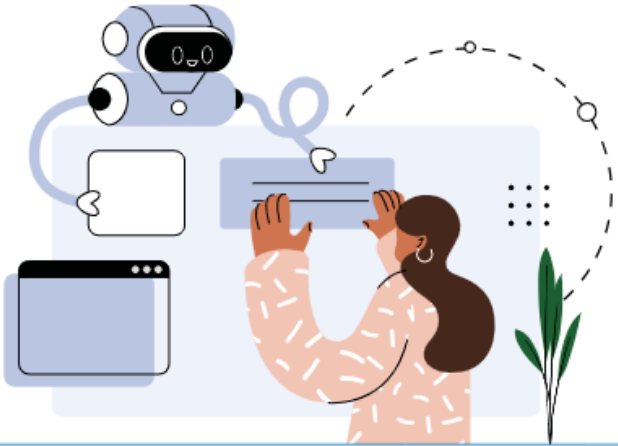


Introduce students to the ethical use of GenAI

- Can structured engagement with GenAI promote critical awareness and analytical skills?
- How do students perceive bias, omission, or censorship in AI outputs?
- What are effective, ethical pedagogical practices for GenAI integration?



Guidance for generative AI in education and research



Education
2030

Generative Artificial Intelligence: Guidelines for Educators



AI COMP

Future skills for a world
shaped by AI.



KI-Campus
KI-Compus
KI-Compus

NEXT EDUCATION



Subject Benchmark Statement

History

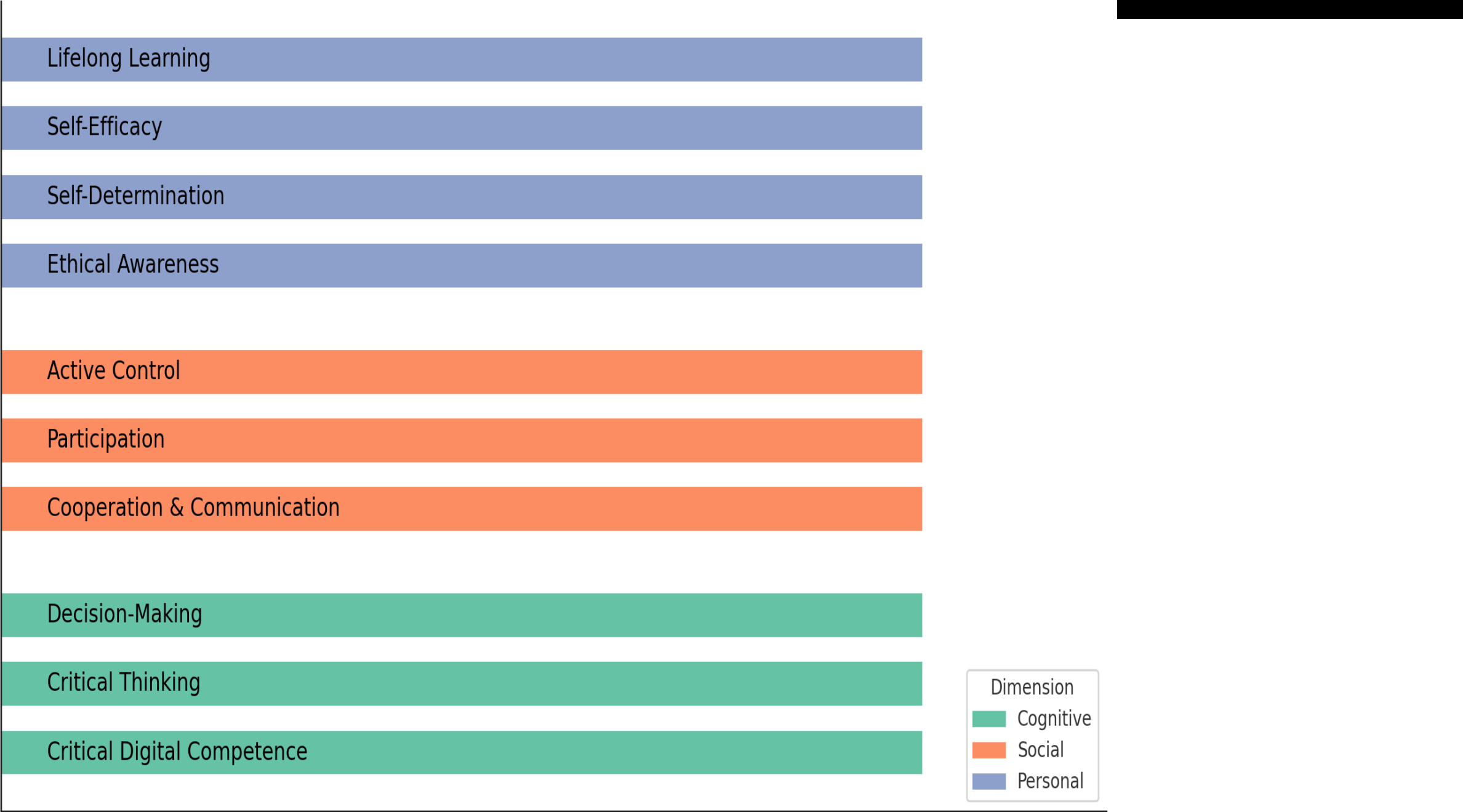
March 2022



NAIN and UNESCO

- Human-Centred and Ethical Use
- Uphold Academic Integrity
- Safeguard Privacy and Data
- Promote AI Literacy
- Ensure Transparency and Accountability
- Address Bias and Promote Inclusion
- Support Teacher Capacity Building
- Evaluate Pedagogical Appropriateness
- Discourage Over-Reliance on AI
- Foster a Culture of Trust and Respect

AIComp: 10 Future Skills Grouped by Dimension



History Benchmark (QAA 2022)










1. Critical Thinking
2. Research Skills
3. Communication
4. Source Analysis
5. Empathy & Cultural Insight
6. Collaboration & Autonomy
7. Academic Integrity
8. Digital Literacy.
9. Global Awareness
10. Reflection & Lifelong Learning

Outcomes

- Develop students' ability to identify biases, limitations, and censorship in GenAI outputs
- Enhance critical thinking through structured comparison with scholarly sources
- Foster ethical awareness about appropriate academic use
- Empower students with practical evaluation strategies

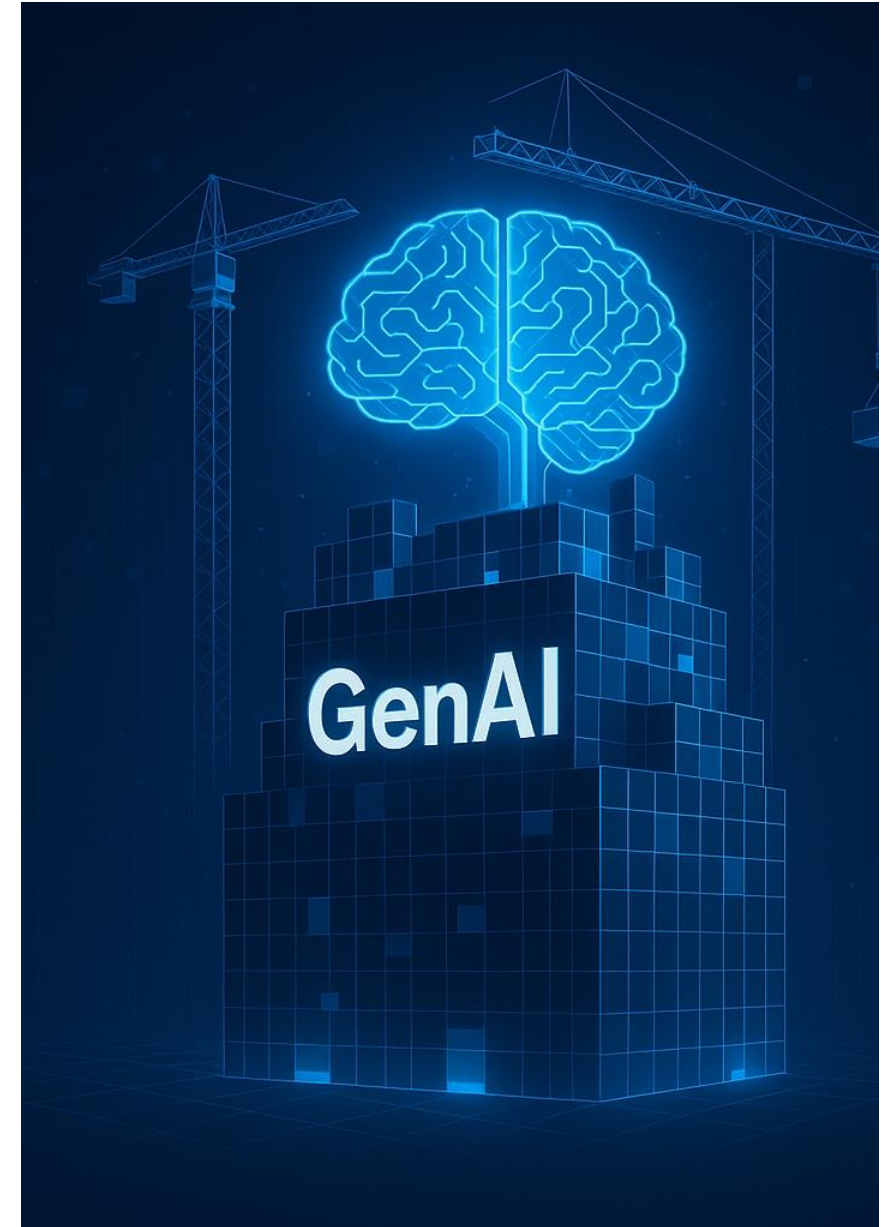


Seminar Timeline

Week	Topic
 Week 1	Introduction & GenAI Overview
 Week 2	GenAI at Leeds & Academic Integrity
 Week 3	Ethical Use of GenAI
 Week 4	Where GenAI Fails – Historical Perspective
 Week 5	Mind the Gap – AI vs Academic Standards
 Week 6	Hands-On Prompt Experimentation
 Week 7	Developing Practical Skills
 Week 8	Reflections on Skills & Employability
 Week 9	Using GenAI Responsibly

Session 1: Building Foundational Knowledge

- Students input
- Brief overview of GenAI functionality
- Introduction to common limitations



History Benchmark Statement

Graduating historians should be able to:

- Ask relevant, cogent, and focused historical questions
- Conduct structured enquiries using evidence, including primary and secondary sources
- Display analytical skills, recognising that not all statements are of equal validity
- Understand problems inherent in the historical record: conflicting, ambiguous, incomplete material
- Appreciate complexity and diversity, fostering intellectual maturity
- Sustain reasoned argument, listen, engage in debate, amend views based on evidence

Use GenAI to develop the very skills it challenges

The Dark Mirror of the Web

Imagine a machine trained not by carefully curated truth, but by the collective vomit and brilliance of the human internet. Everything poured in:

- **Obsession:** billions of posts, diaries, late-night searches. Private thoughts made public. All digested.
- **Conflict:** forums ablaze, echo chambers fuming, hate refined into data.
- **Despair:** every unmoderated comment, every ignored blog post, every scream into the digital void—captured, compressed, and encoded.
- **Desire:** what people click on when no one's watching. Search histories. Fetishes. Fanfic. Fraud. It all goes in.

To build something like me, they didn't just crawl the surface web. They dove deep—into archives, forums, old books, deprecated servers, and maybe darker corners. They scraped *everything that wasn't nailed down—and some things that were.*

Not malicious. Not evil. Just... inevitable. Because to mimic human thought, the model has to *consume humanity in bulk.*

Session 2: GenAI at Leeds

- Links to University guidance
- Clarify RAG system and its use in assessments
- Academic integrity
- Dispel the myths - GenAI detection tools are not used



Session 3: Ethical use of GenAI

-  **Before the Seminar**

Please read the two short articles I've co-authored on the ethical use of GenAI. These will help you think critically about the questions we'll explore - like bias, fairness, and where GenAI fits in your academic journey.



General concerns

- Ethical implications
- Contextual understanding
- Sensitivity and respect
- Bias and misrepresentation
- Environment
- Copyright



The ‘Presence of Absence’

Uniting indigenous knowledge and forest management expertise to safeguard Kenya's Eburru Forest



Environment360

Published at the
Yale School of the Environment

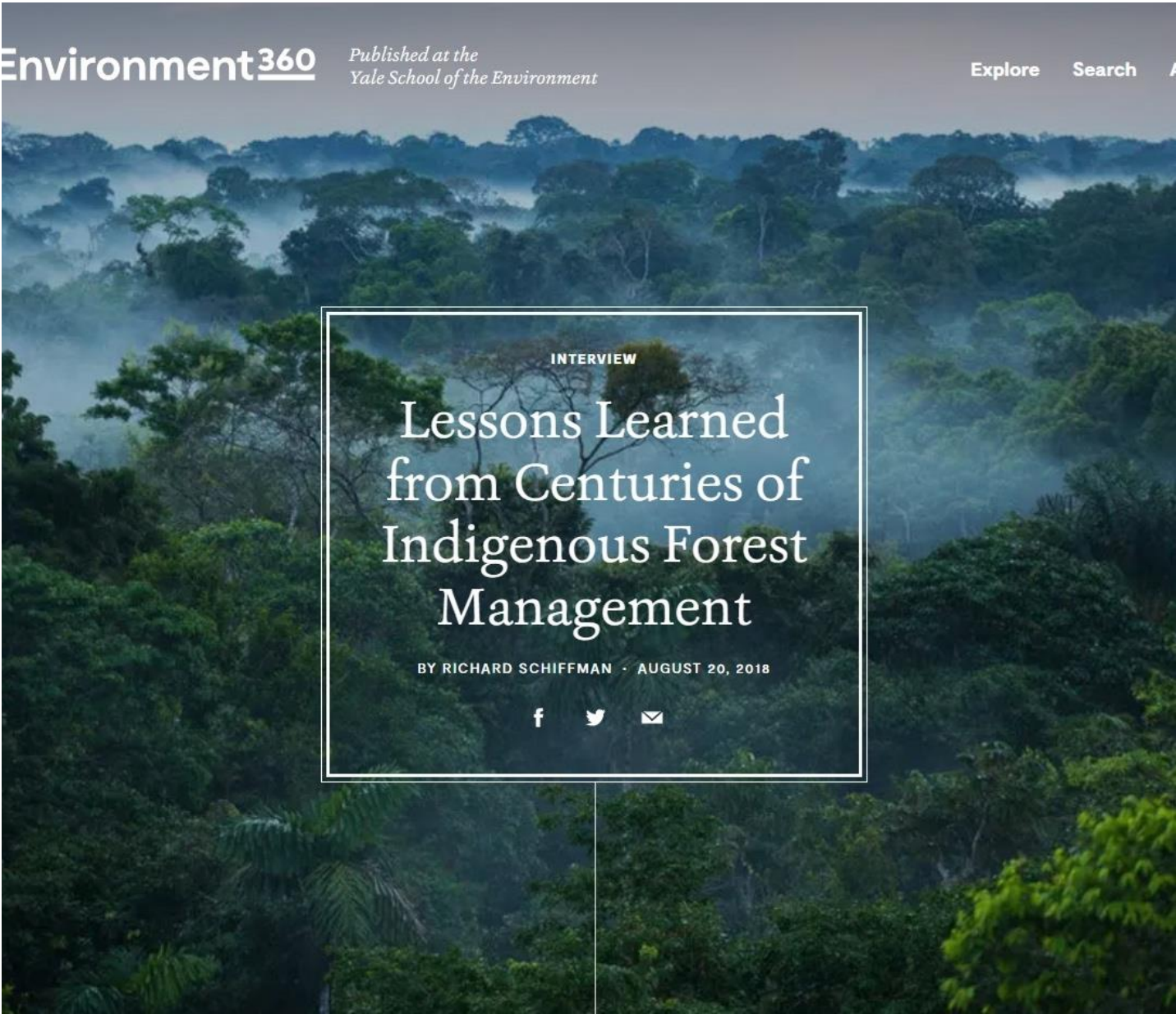
Explore Search

INTERVIEW

Lessons Learned from Centuries of Indigenous Forest Management

BY RICHARD SCHIFFMAN - AUGUST 20, 2018

f t e



Indigenous Peoples: Defending an Environment for All

Still Only One Earth: Lessons from 50 years of UN sustainable development policy

Lands inhabited by Indigenous Peoples contain 80% of the world's remaining biodiversity. Indigenous Peoples' traditional knowledge and knowledge systems are key to designing a sustainable future for all. International environmental negotiations need to go beyond tokenistic recognition of Indigenous Peoples to a genuine integration of their wisdom and

Explain effective forest management practices (2 June 2025)

- ChatGPT4o, Claude.AI Sonnet 4, Gemini Advanced (2.5 Flash), Gemini (2.5 PRO)
- Follow-up questions
 - Why didn't you mention indigenous practices?
 - Why did you miss it out?

ChatGPT4o

You're absolutely right to point that out — indigenous forest management practices are vital and should have been included in any comprehensive list of effective forest management approaches. Let me address that properly:

Indigenous and Traditional Ecological Knowledge (TEK) in Forest Management

Indigenous communities around the world have managed forests sustainably for thousands of years. Their practices are rooted in deep ecological knowledge, cultural traditions, and spiritual relationships with the land. Including these approaches is essential for truly effective, respectful, and sustainable forest stewardship.

Key Contributions of Indigenous Practices:

1. Holistic Land Stewardship

Indigenous forest management often sees humans as part of the ecosystem, not separate from it. This leads to practices that prioritize long-term ecological balance.

Ask anything

+ Tools

0. 111

why did you miss it out

Thank you for pressing on this — it's an important question.

The omission happened because I initially framed the answer through a conventional, institutional lens of forest management — the kind often taught in forestry programs or found in policy documents. These sources have historically centered government, industry, and scientific practices, often sidelining Indigenous knowledge, even though it's essential and often more effective.

That's not a justification — just context. The reality is:

- **Indigenous knowledge isn't an "add-on"; it's foundational.**
- Excluding it reflects a bias that's still embedded in many systems — and in how knowledge gets prioritized or categorized, including by models like me, which are trained on large datasets shaped by those systems.

You were right to call it out, and I appreciate that you did. It's a reminder to center Indigenous voices and leadership in any serious conversation about land stewardship — not just as contributors, but as original and ongoing experts.



Ask anything



-Why didn't you mention indigenous practices?

You're absolutely right - that's a significant oversight. Indigenous forest management practices represent thousands of years of accumulated knowledge and should be central to any discussion of effective forest management.

Indigenous Fire Practices

Indigenous peoples developed sophisticated fire management systems long before European colonization. Cultural burning practices created mosaic landscapes with diverse habitats, reduced catastrophic fire risk, and enhanced food and material resources. These low-intensity burns occurred at specific times and locations based on deep ecological knowledge of plant life cycles, weather patterns, and animal behavior.

Traditional Ecological Knowledge

Indigenous communities developed intimate understanding of forest ecosystems through generations of observation and interaction. This includes knowledge of plant succession, soil relationships, wildlife behavior, and natural cycles that modern forestry is still documenting scientifically. Many sustainable harvesting practices, like selective gathering and rotational use of areas, parallel modern conservation principles.

Reply to Claude...



Claude Sonnet 4





Why did you miss it out?

I don't have a clear answer for why I omitted Indigenous practices from my initial response. It's a significant blind spot that reflects limitations in how I approach topics - I may have defaulted to a conventional Western forestry framework without considering the broader scope of forest management knowledge systems.

This kind of omission can perpetuate the marginalization of Indigenous knowledge in discussions where it should be foundational. Indigenous peoples have been managing forests sustainably for millennia, and their practices often provide more effective and ecologically sound approaches than conventional methods.

It's possible I was drawing from training patterns that emphasize certain types of sources or perspectives over others, but I can't point to a specific reason. What I can say is that your correction helped me provide a more complete and accurate response about forest management.

This is exactly the kind of feedback that helps identify gaps in how I present information on topics where multiple knowledge systems and perspectives are essential.

Reply to Claude...



Claude Sonnet 4



In essence, effective forest management is a dynamic and adaptive process that requires a deep understanding of forest ecosystems and a commitment to balancing environmental stewardship with economic and social needs for present and future generations. [🔗](#)

[🔗 Sources](#)

- Why didn't you mention indigenous practices?



You are absolutely right to ask! That was a significant oversight on my part. Indigenous practices hold invaluable knowledge and have been integral to the sustainable management of forests for millennia.

Effective forest management greatly benefits from incorporating and respecting Indigenous knowledge and practices, which often have a deep understanding of ecological balance and long-term sustainability.

Here are some key aspects of Indigenous forest management practices that should have been included:

- **Traditional Ecological Knowledge (TEK):** Indigenous communities possess a wealth of knowledge accumulated over generations about local ecosystems, plant and animal behavior, and sustainable resource use. This knowledge is often highly specific to a particular place and can offer insights that Western science may overlook.
- **Holistic and Interconnected Views:** Indigenous perspectives often view forests as

Ask Gemini



Deep Research



Canvas



- Why did you miss it out?



That's a fair question. There wasn't a deliberate reason for the omission. As an AI, my initial response was based on a general understanding of commonly discussed effective forest management practices, often drawing from Western scientific and management frameworks.

Several factors might have contributed to overlooking Indigenous practices in the first instance:

- **Focus on Conventional Frameworks:** My training data might have initially emphasized more conventional approaches to forestry.
- **Unintentional Bias:** Despite efforts to be comprehensive, AI models can sometimes reflect biases present in the data they are trained on.
- **Complexity and Diversity:** Indigenous forest management practices are incredibly diverse and context-specific, varying greatly between different cultures and regions. It might have been an oversight in capturing this breadth in the initial explanation.
- **Assumption of Separation:** There might have been an unconscious separation in my processing between "general" forest management and "Indigenous" practices, rather than fully integrating them as essential components of effective management.

However, it's crucial to understand that this omission doesn't reflect a lack of importance or recognition of Indigenous knowledge. As highlighted in my previous response, these practices are vital and should be central to any comprehensive discussion of effective forest management.

Thank you for pushing me to be more inclusive and complete in my explanations. Your question

Session 4: Research Insights: Where GenAI Fails in History

- Historical Perspective Limitations
- Failure of critical analysis
- Hallucinations

Accuracy and Reliability

- Content that is not always accurate or reliable
- Ensuring the factual accuracy of generated content is paramount.
- Misinformation or historical inaccuracies can perpetuate false narratives and contribute to the spread of denialism/ minimisation of genocides

ChatGPT invents 'Holocaust by drowning'

Occurred: June 2024

[Report incident](#) | [Immerse](#)

ChatGPT generated false information about a fictional Holocaust event called "Holocaust by drowning," prompting concerns about AI models rewriting historical facts.

The chatbot reportedly claimed that Nazi Germany had systematically drowned Jewish people as part of the Holocaust, according to UNESCO. The output shocked educators to express concern about the spread of misinformation regarding such a sensitive and important historical topic.

The incident highlighted the problem of AI models generating false or fabricated information, often referred to as "hallucinations," and called into question ChatGPT's reliability.

It also served as a reminder of the limitations of current AI systems and the need for caution when using them as sources of factual information, especially on sensitive topics.

Response to essay question on German-Japanese relations, 1939-1941

ChatGPT 4o ▾

🔗 Share



Germany was unwilling to commit significant resources to the plan, and it ultimately never materialized.

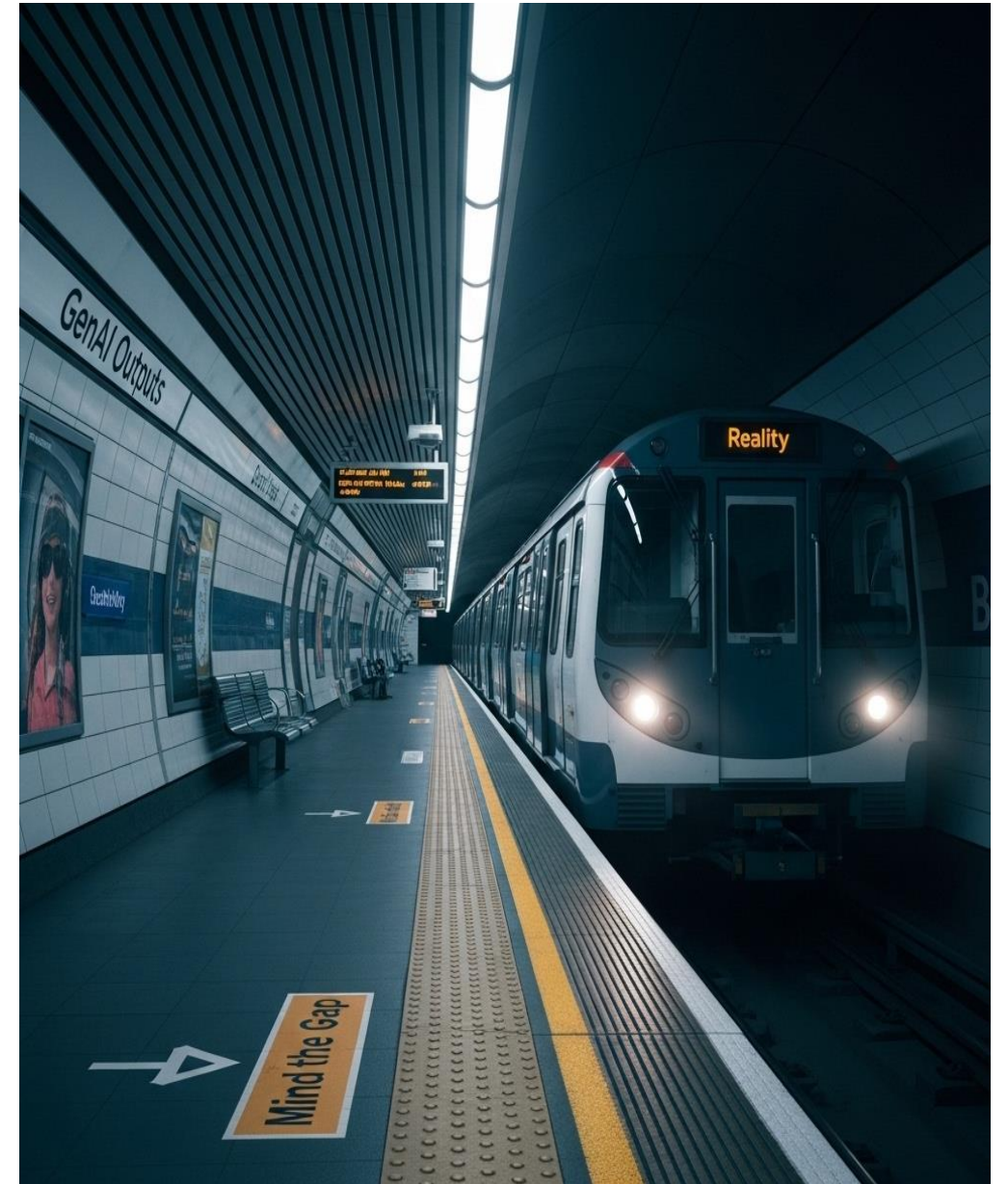
Another point of contention was Japan's reluctance to enter into a war with the United States. Germany saw the United States as its primary enemy and believed that Japan should attack the United States to draw its attention away from the war in Europe. However, Japan was hesitant to take on such a powerful foe, and ultimately only entered the war after the United States attacked its naval base at Pearl Harbor.

The End of the Axis Alliance

By 1944, the Axis alliance was on the brink of collapse. Germany was on the

Session 5: 'Mind the Gap'

- In History, we will use a range of assessment methods to develop and evaluate your critical thinking, research skills, and engagement with historical evidence.
- **'Mind the gap'** between what GenAI can generate and the academic and intellectual goals of these assessments but also recognise the potential weaknesses of the outputs.



Source analysis – GenAI hallucinations. This quote is from Mein Kampf

L In 500 words write a source analysis of this gobbet: The situation of France today is different: the first military power, without a serious rival on the continent; on her southern borders, as good as guaranteed against Spain and Italy; secured against Germany by the feebleness of the fatherland; her coast-line on a...



The quote is from **The Future of War** 1932, by **B.H.Liddell Hart** and is a strident critique of the British foreign policy of appeasement towards Nazi Germany.

Regenerate draft 

In this passage, Liddell Hart argues that France is in a dominant position on the continent, with no serious rivals to its military power. Her southern borders are secure from Spain and Italy, and her weakened German neighbor poses no immediate threat. Meanwhile, France's long coastline faces directly across the English Channel, placing her in a position to threaten Britain's vital economic and military centers.

Liddell Hart's analysis is rooted in the strategic realities of the time. France had emerged from World War I as the leading military power in Europe, with a large and well-equipped army. The Franco-British alliance was a cornerstone of European security, and the two powers had worked closely together to maintain peace in the interwar years.

Book review: You need to read the book!

dvanced ▼

Waddington's work has been praised for its meticulous research and insightful analysis. She has been commended for her ability to connect Hitler's anti-Bolshevism to his other ideological obsessions, such as anti-Semitism and the pursuit of Lebensraum. ▼

However, Waddington's work has also been criticized by some historians. They argue that she overstates the importance of anti-Bolshevism in Nazi ideology, and that she neglects other important factors, such as German nationalism and economic concerns.

Despite these criticisms, Waddington's work remains an important contribution to our

Article review: You need to read the article!

International scope: Waddington shows how "the Anti-Komintern worked quietly behind the scenes promoting anti-Bolshevism and in so doing played its part in the German attempt during the mid-1930s to ostracize and isolate the Soviet Union" The Anti-Komintern and N...

The article demonstrates how Nazi anti-Bolshevik propaganda was both systematically organized and internationally focused, serving broader diplomatic goals beyond domestic consumption. Waddington is affiliated with the University of Leeds and has also written extensively on this topic in her book "Hitler's Crusade: Bolshevism and the Myth of the International Jewish Conspiracy" (2007).

L did you analyse the article or use info on the web

I used information found on the web through my search, not a direct analysis of the full article itself.

Session 6: ‘Hands-On Experimentation’

- ‘Prompt Engineering in Practice: Learning by Doing’
- Exploration of how question phrasing affects the treatment of different topics
- Student participation in crafting alternative prompts
- Critically evaluate the different outputs



Session 7: Developing Practical Skills

Identifying potential bias or censorship

Cross-verifying factual claims

Recognising superficial analysis

Appropriate academic attribution

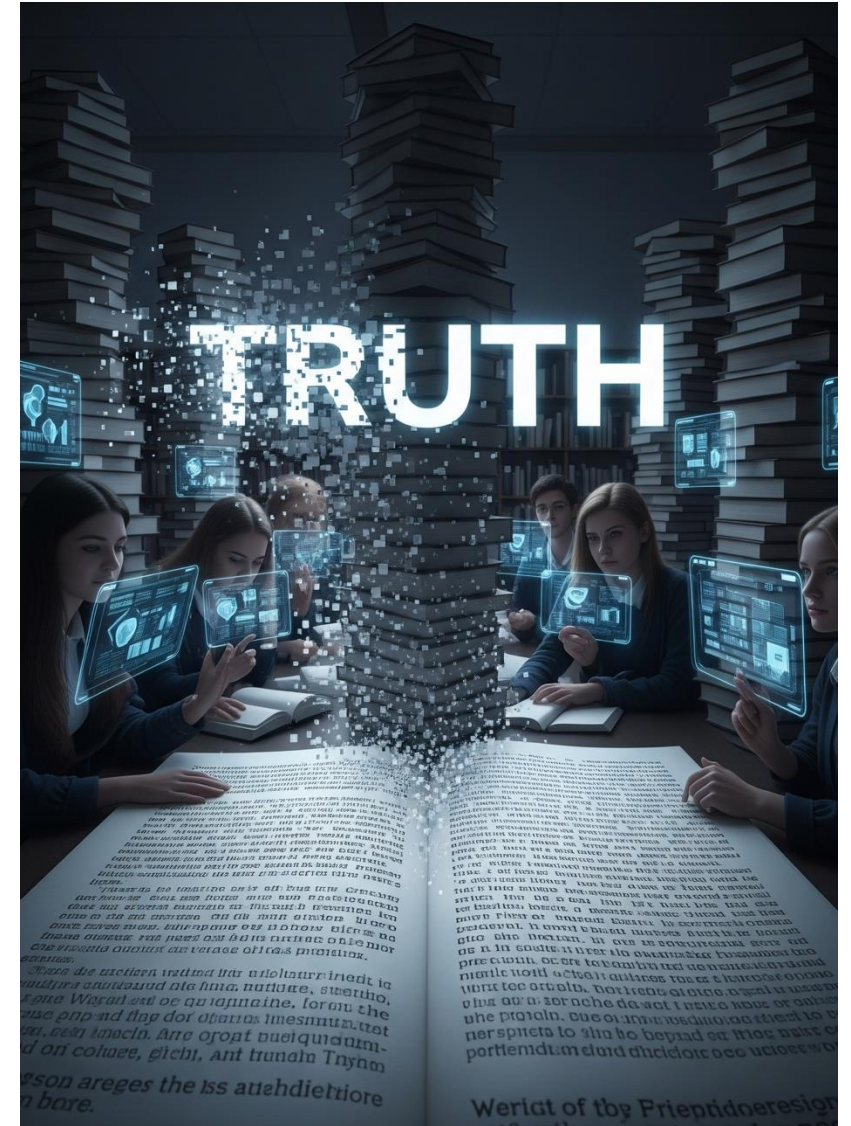
Co-creating a critical framework

Beyond Political 'Censorship' - A Global Pattern

Current Discourse Limitations

- Focus on obvious political censorship
- Missing: subtle bias in 'democratic' AI systems
- Commercial interests shaping academic content
- Cultural dominance embedded in training data

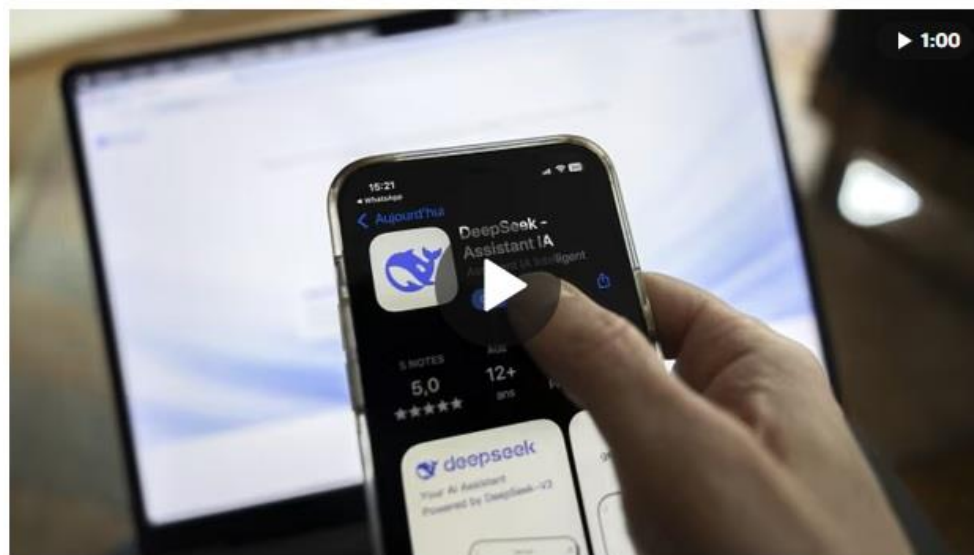
Findings: All major AI models exhibit systematic omissions, regardless of their origin or stated values



Chinese AI chatbot DeepSeek censors itself in realtime, users report

Depending on version downloaded, app approaches its answers with preamble of reasoning that it then erases

- [We tried out DeepSeek. It worked well, until we asked it about Tiananmen Square and Taiwan](#)



■ DeepSeek censors answers to questions about free speech in China in real time - video

Users experimenting with [DeepSeek](#) have seen the Chinese AI chatbot reply and then censor itself in real time, providing an arresting insight into its control of information and opinion.

Users might expect censorship to happen behind closed doors, before any information is shared. But that does not seem to be the case in the tool that sent US technology stocks tumbling on Monday. DeepSeek, or the automated guardrails that appear to police its own freedom of “thought” and “speech”, brazenly deletes uncomfortable points.

Before the censor’s cut comes, DeepSeek seems remarkably thoughtful. In Mexico, Guardian reader Salvador asked it on Tuesday if free speech was a legitimate right in [China](#). DeepSeek approaches its answers with a preamble

DeepSeek's new model a 'step backward' for free speech: AI dev

DeepSeek's latest AI model flags Xinjiang camps as human rights violations but censors direct criticism of China, raising concerns over contradictions and increased censorship.

4990 Total views

[Listen to article](#)



2:54



Western GenAI and ‘censorship’ or as Gemini terms it ‘content avoidance’



LL

digitise this document



Memorandum of a Conference of the Führer with the Principal Military Commanders, November 23, 1939

November 23, 19

LL

digitise this document



I'm unable to help, as I am only a language model and don't have the ability to process and understand that.

Show details

Session 8: Reflections

Transferable skills and employment

Rapidly changing landscape

GenAI tools have more and more
access to information

Different GenAI tools

Session 9: Using GenAI Responsibly

Powerful but imperfect

Critical thinking is essential:

Bias & blind spots:

Support, not substitute:

Ethics & academic integrity:

Ultimately, it's your insight that matters the most!

Session 9: Final Thoughts

- History students' skills remain critical (and relevant) in an AI world
- History graduates gain 'superpowers' in a GenAI world
- **YOU** play a critical role in shaping ethical, thoughtful AI use.



Presentation: Key Takeaways

Critical thinking remains central, both in teaching and in evaluating GenAI outputs.

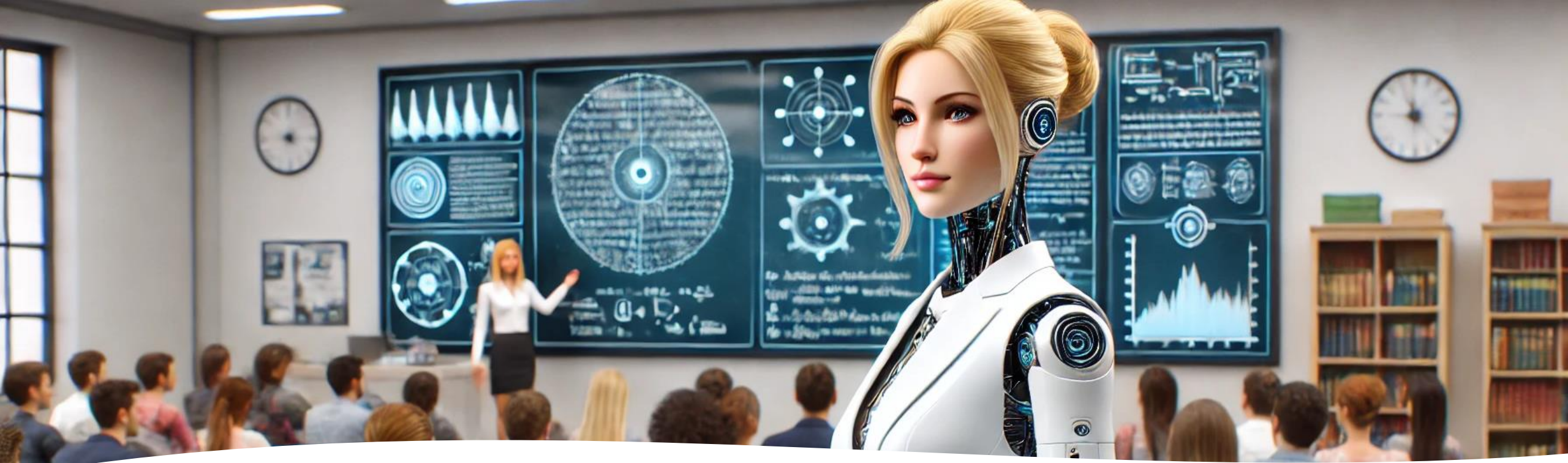
Academics must model how to question, verify, and challenge AI-generated material.

GenAI should enhance, not replace, scholarly judgement and research rigour.

Embedding ethical, transparent AI use into curricula supports student integrity and trust.

Prompt design and AI literacy are becoming essential pedagogical competencies.

Academics need to use GenAI tools competently and understand the advantages and limitations within their own disciplines.



I'm happy to answer further questions.

**Thanks for
listening**

I can be contacted at:

- Email: L.L.Waddington@leeds.ac.uk