S.HI.E.L.D
VS DISINFO

Students' of Higher Education critical digital Literacy Development against Disinformation

INTERNATIONAL CONFERENCE PROCEEDINGS

Revisiting Disinformation: Critical Media Literacy Approaches

27-28 June, 2025

Edited by:

Panagiota Samioti, Konstantinos Sipitanos and Eleni Katsarou
(University of Crete)

Co-funded by the European Union

Co-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. [Project number: 2022-1-FR01-KA220-HED-000088804]

Revisiting Disinformation: Critical Media Literacy Approaches

International Conference Proceedings

Editors

Panagiota Samioti
Konstantinos Sipitanos
Eleni Katsarou

* * *

FACULTY OF LETTERS
UNIVERSITY OF CRETE

RETHYMNO 2026

Electronic edition (e-PDF)
ISBN 978-618-5762-03-2

© 2026 Faculty of Letters, University of Crete
Gallou University Campus, P.C. 741 00 Rethymno
Tel.: +30 28310 77780, e-mail: dean@phl.uoc.gr

CONTENTS

Preface
by the S.HI.E.L.D. vs Disinfo Project's coordinator, Dimitris Kotzinos……………………………………8

Acknowledgements
by the editors 10

Introduction
Panagiota Samioti, Konstantinos Sipitanos, Eleni Katsarou 11

Allegory in the Service of Power (?): Historical Uses of Misinformation and Disinformation
Eleni Chatzimavroudi, Marios-Kyparissis Moros 24

Addressing the Impact of Medical Misinformation
Przemysław Waszak, Paweł Zagożdżon... 32

How can we Perceive Conspiracy Theories in order to Diversify Existing Typologies?
Ioannis Elissaios Paparrigopoulos, Dora Katsamori, Georgios Petasis 41

Multimodal Interpretation of a Mobbing Paradigm in the School Workplace
Georgios Elias Potamias, Paraskevi Kanari 53

Are they Yielding Results? Examining the Relevance of Fact- Checking Training to Information Verification Competence of Nigerian Journalism Students
Ibrahim Saheed Bidemi, Taiwo Rukayat Abdullahi.…………………………………………………………….61

News, Social Media and Video Analytics: the MediaPot platform
Elisavet Palogiannidi, Sotiris Legkas, Dimitrios Vogiatzis, Manolis Mylonas, Vasso Koutsoupia, Vasileios Mezaris, Spyridoula Markou, George Zissis, Pantelis Theodosiou 71

Harnessing Emerging Technologies to Combat Disinformation: Innovation, Ethics, and Resilience in Journalism
Ioanna Georgia Eskiadi, Nikolaos Panagiotou…………………………………………………………………….85

The News Weight Coefficient (NeWC) for Measuring the Accuracy of Discrimination Between False and True News with Human and Non-Human
Teodor Răileanu-Olariu, Bogdan Oprea …………………………………………………………………………...97

From the theory of Critical Digital Literacy to Pedagogical Practice in HE: Combating Disinformation in the S.HI.E.L.D. vs Disinfo Project
Panagiota Samioti, Konstantinos Sipitanos, Mari-Liis Madisson, Andreas Ventsel, Dimitris Kotzinos, Georgios Terzis, Christos Gavalas……..……………………………………………………………108

CDL and Critical Creativity: Engaging Students with AI to Counter Disinformation in Higher Education
Panagiota Samioti, Rafail Giannadakis 121

Source Evaluation in a Digital World: From Information Consumers to Critical Evaluators
Zoi A. Traga Philippakos........ 134

The Immunity Illusion: Media Literacy in an Age of Identity-Driven Disinformation
Massimo Flore 141

Critical Literacy as a Tool for Approaching the Rhetoric of Multimodal Texts in Secondary Education: A Teaching Approach to Media Analysis in the Classroom
Anna Paraskevaidi 149

Enhancing Media Critical Thinking against Disinformation with a Socratic Chatbot: A Methodology
Panayiotis Kapetanakis, George Petasis, Dora Katsamori, Eirini Papachristou, Giannis Paparigopoulos……..166

Scaffolding Thought: Designing AI Tools that Empower Writers, Not Replace Them
Zoi A. Traga Philippakos........ 187

How to Critically Identify Fake News in Primary School: A Case Study
Ioanna Kitsou……..195

Preface

Dimitris Kotzinos
Professor, Department of Computer Science, CY Cergy Paris University
S.HI.E.L.D. vs Disinfo Coordinator

We are pleased to present the volume *Revisiting Disinformation: Critical Media Literacy Approaches*, which brings together selected contributions from the international conference of the same title, held at the University of Crete, Rethymno, on June 27–28, 2025. The conference was organised within the framework of the Erasmus+ KA220HED project S.HI.E.L.D. vs Disinfo (Students' Higher Education Literacy Development against Disinformation), an initiative devoted to promoting critical digital literacy, fostering resilience to misinformation, and strengthening the role of higher education in cultivating informed and ethically engaged citizens.

This volume reflects the project and the conference's shared commitment to exploring disinformation as a complex social, cultural, and technological phenomenon. The chapters approach the topic from multiple perspectives - historical, philosophical, technological, pedagogical, and creative - illustrating the breadth of the international dialogue that unfolded during the conference. Collectively, they offer both theoretical insight and practical innovation, underscoring the necessity of interdisciplinary collaboration in addressing the epistemological and ethical challenges of our digital era.

The structure of the volume mirrors this diversity through four thematic sections that together trace a continuum from conceptual inquiry to applied practice. The first section, *Historical and Conceptual Perspectives*, revisits the genealogy of misinformation and its intersections with ideology, allegory, and the politics of truth. The second section, *Technological Innovations and Media Practices*, investigates the impact of artificial intelligence, data analysis, and emerging media tools on news verification and disinformation detection. The third section, *Pedagogical and Educational Frameworks*, presents outcomes of the S.HI.E.L.D. vs Disinfo project, showcasing how critical digital literacy (CDL) can be integrated into higher education curricula and teacher training. Finally, the fourth section, *Teaching and Learning Applications*, features case studies and classroom interventions that demonstrate how CDL principles can be adapted across educational levels, from primary schools to universities.

Taken together, the contributions illustrate that disinformation is not only a technological or cognitive issue but also one deeply rooted in trust, identity, and power. Through the collective efforts of historians, educators, journalists, computer scientists, and philosophers, this volume seeks to sustain an ongoing interdisciplinary conversation on critical literacy, democratic participation, and civic empowerment. We hope that the insights and practices shared here will serve as a catalyst for further collaboration, innovation, and reflection across the fields of education, media, and public life.

We would like to express our sincere gratitude to all authors and reviewers for their thoughtful contributions, to the conference participants for their inspiring discussions, and to the Erasmus+ Programme of the European Union for its generous support of the S.HI.E.L.D. vs Disinfo project, which made this collective endeavor possible.

Acknowledgements

The editors would like to thank the Faculty of Letters of the University of Crete for kindly agreeing to publish this volume in the Ariadne series, as well as the production team for their valuable guidance and support throughout the editorial process.

We are also grateful to the European Union for its support through the Erasmus+ programmes, which funded the project and the organisation of the conference held in Rethymno in June 2025, and made possible the preparation and completion of the present volume.

Our thanks are due to all contributors for their cooperation, their patience with editorial requests, and their constructive engagement with feedback aimed at strengthening their chapters. We very much hope that the final result meets their expectations.

Finally, we warmly thank Maria Tzanaki, graduating student of the Department of Primary Education at the University of Crete, for her voluntary assistance with the formatting and final page layout of the volume, as well as with ensuring the consistency of the references.

The Editors

Introduction

Panagiota Samioti[1], Konstantinos Sipitanos[2], Eleni Katsarou[3]
1. Laboratory and Teaching Staff, School of Philosophy, University of Crete
2. Postdoctoral Researcher, University of Crete
3. Professor, Department of Primary Education, University of Crete

This introduction outlines the rationale, scope, and structure of the proceedings volume *Revisiting Disinformation: Critical Media Literacy Approaches*, which emerges from the international conference of the same title. The volume brings together selected papers that reflect and expand upon the central objectives of the S.HI.E.L.D. vs Disinfo (Students' Higher Education Literacy Development against Disinformation) Erasmus+ KA220HED project, namely, to examine disinformation as a multidimensional phenomenon and to promote critical digital literacy as a framework for resilience, ethical engagement, and informed citizenship in the digital age.

The Current State of Disinformation

Even though disinformation is not a new phenomenon, it has emerged as one of the most important issues influencing public health, science, politics, journalism, and education today. Due to social media platforms, search engines and algorithmic personalization, misleading information is being circulated rapidly (Vosoughi et al., 2018). The phenomenon can be described under the term "epistemological crisis", since the integrity of knowledge is under threat (Friedman, 2023).

Emotional appeal, polarization, and distrust in institutions are some of the features of disinformation today (Hameleers, 2023). Additionally, "echo chambers" and "filter bubbles" which reproduce the same content to the same users diminish alternative perspectives and prevent encounters with disinformation (Sasahara et al., 2021). Moreover, platforms like Reddit and Youtube host communities that reproduce conspiracy theories blurring the lines between belief, play, and irony (Godwin et al., 2025).

The spread of disinformation does not affect only belief systems, it also affects public trust, policies and democracy. More specifically disinformation campaigns are a form of "information warfare" by weaponizing social media and political influence (Singer & Brooking, 2018). Network analysis has shown that coordinated disinformation networks strategically act to distort debates, destabilize institutions and manipulate online ecosystems. The dangers of misinformation can be identified during the COVID-19 pandemic, where vaccines and conspiracy theories became a threat to public health (Caceres et al., 2022).

Within this context, critical digital literacy [hence CDL] (Sipitanos, 2023c, 2023b) can equip citizens with skills to evaluate misinformation, recognize bias and identify fake news. CDL and media literacy are not merely aimed at teaching people to distrust content, as such aims could make them even more vulnerable to manipulation. CDL balances skepticism with resilience and ethical responsibility (Pangrazio, 2016).

The importance of CDL has been acknowledged by European institutions and international organisations. The European Commission has published the Digital Competence Framework (DigComp) in which digital literacy is described as a key transversal skill for active citizens and employability (European Commission, 2022). The International Commission on the Futures of Education, (2021) and UNESCO (2013) have provided global guidelines for media and information literacy. Additionally, the European Digital Media Observatory (2024) has underlined the need for collaborative, cross-sectoral approaches involving schools, universities, journalists, and civil society organisations.

Nevertheless, research shows that these efforts are not sufficient, since there is still a discrepancy between the policies and the educational practices, creating a growing gap between everyday media practices and institutional curricula (Pangrazio, 2016) Moreover, this distortion is caused by gaps in teacher training, resources availability and policy alignment with educational reality (Cherner & Curry, 2019; Martyushev et al., 2021). Therefore, there is an urgent need to develop an innovative, interdisciplinary teacher preparation and students' education.

Subsequently, to tackle disinformation there is a need for multilayered, ongoing, interdisciplinary preparation for professionals to confront the phenomenon in different disciplines. Doctors, health experts, journalists, teachers, and ICT developers are on the front lines of the battle against fake news. Therefore, higher education institutions and lifelong professional training services need to develop curricula, educational materials and services to support future and in-service professionals.

Higher Education in the Digital Age

Higher Education plays a significant role in preparing professionals and future citizens to tackle disinformation. Whether conceived as information literacy (a set of skills to recognize information) or digital literacy (ability to find, evaluate, create and communicate information) Universities have played a significant role (Leaning, 2017). This role is even more essential today – in the age of disinformation - to forge democratic participation.

During the COVID-19 period it became obvious that digital technologies and information literacy are an essential part of the higher education system, since they play a significant role not only in enhancing access to information, but also in preparing professionals well connected with society and civic engagement in general (Tejedor et al., 2020). Nevertheless, this rapid digital transformation also revealed problems of access, as well as problems in pedagogy and social justice (Sipitanos, 2023b).

The adoption of a pedagogy based on CDL in higher education will play a significant role in the fight against disinformation, and in the preparation of future professionals.

Only highly skilled professionals who understand the language and the practices through which disinformation is being created and disseminated can contribute significantly to tackling disinformation (Sipitanos, 2023a). This connection between digital media and critical literacy alongside personal engagement positions future professionals in Higher Education institutions towards a reflexive pedagogy which reinforces and empowers them (Pangrazio, 2016).

This perspective has been adopted by major organisations showing the importance of information in the fake news era. One example is the Association of College and Research Libraries (2016), which states that information literacy is not a set of static skills, but a set of interconnected practices, in which those involved participate actively in knowledge production. The International Commission on the Futures of Education (2021) also in the “Global Standards for Media and Information Literacy Curricula Development” underscores that higher education institutions should invest beyond the acquisition of technical competencies towards the incorporation and adoption of ethical, cultural and civic perspectives preparing future professionals for the contemporary problems that societies are confronting.

The S.HI.E.L.D. vs Disinfo Project: A European Response to Disinformation

Against this backdrop of growing concerns over disinformation and the urgent need to foster critical digital literacy, the S.HI.E.L.D. vs Disinfo project was developed as a European initiative aimed at addressing these challenges. Its central mission is to design, test, and disseminate innovative pedagogical practices that strengthen higher education’s role in countering disinformation and fostering resilience among students, educators, and wider communities. By focusing on action-oriented strategies, the project seeks to map the disinformation landscape and to equip universities with practical tools such as curricula, implemented teaching practices in specific contexts (see anthology) and a teachers’ toolkit for developing sustainable interventions.

At its core, the S.HI.E.L.D. vs Disinfo project aligns with broader European policy frameworks. The European Commission’s *Digital Competence Framework (DigComp)* (Vuorikari et al., 2016) identifies competencies such as information evaluation, digital safety, and problem-solving as essential for citizenship in the digital era. Similarly, the European Digital Media Observatory (EDMO, 2024) emphasizes that effective media literacy requires cooperation between universities, civil society, policymakers, and journalists. The S.HI.E.L.D. vs Disinfo project operates at this intersection, embedding research-driven approaches within higher education curricula while engaging with broader societal actors.

Throughout all phases, the S.HI.E.L.D. vs Disinfo project was grounded in action research, which encouraged collaborative inquiry and reflection among students, educators, and curriculum developers. Action research ensures that the production of educational material to tackle disinformation is not merely a set of skills, but rather an adoption of values, perspectives and stances that promote educators and future professionals who engage actively with the communities they participate in.

The S.HI.E.L.D. vs Disinfo project was also aligned with CDL and especially placed emphasis on the language of disinformation, on the narratives and, in general, on the multiple approaches and strategies through which fake news is being constructed as well as disseminated and recognized. It further aimed to understand falsehood, how fake news is constructed, how and why fake news is circulated and finally, how power, ideology and platform dynamics shape the information environment.

This Erasmus+ project operates a digital platform which disseminates the outputs. The involvement of transnational partners acknowledged the fact that disinformation is a multifaceted phenomenon that is also rooted in cultural, linguistic and political environments. This multinational consortium revealed that country comparisons can provide a deeper understanding of how CDL practices can be adopted in different settings.

The International Conference: Revisiting Disinformation

The culmination of the S.HI.E.L.D. vs Disinfo project's vision was the International Conference "Revisiting Disinformation: Critical Media Literacy Approaches", which took place at the Student Center "Xenia" in Rethymno, Crete, on June 27–28, 2025. This two-day event brought together scholars, researchers, educators, professionals, and students from across disciplines and countries to reflect on the state of disinformation today and to explore pedagogical, methodological, and policy-oriented strategies for addressing it. More than a traditional academic gathering, the conference represented a living laboratory of ideas, practices, and collaborations that extended the aims of the S.HI.E.L.D. vs Disinfo project into a wider international arena.

The central theme, that is critical media literacy as a response to disinformation, acknowledges that fact-checking and technical solutions, while valuable, are insufficient on their own. Instead, the conference foregrounds the importance of cultivating deeper competencies that allow individuals to analyze, critique, and ethically engage with media environments.

Building on the outcomes of the S.HI.E.L.D. vs Disinfo project, the conference proceedings extend its mission from research and pedagogical design to broader scholarly dialogue and application. The contributions presented here demonstrate how critical digital literacy can be enacted across disciplines, institutions, and technological contexts. Together, they translate the project's core principles, that is, collaboration, interdisciplinarity, and reflexivity, into a diverse set of studies, methodologies, and educational practices that address disinformation as a cultural, social, and ethical challenge.

Structure and Content of the Proceedings Volume

This proceedings volume is organised into four thematic sections that collectively explore the multifaceted phenomenon of disinformation. Each section gathers contributions addressing specific dimensions - historical, technological, educational, and pedagogical - reflecting the interdisciplinary scope of the *Revisiting Disinformation* conference. What follows is an outline of the structure and a brief overview of the individual papers that comprise the collection.

The collection begins with *Allegory in the Service of Power (?): Historical Uses of Misinformation and Disinformation* by Eleni Chatzimavroudi and Marios-Kyparissis Moros (Aristotle University of Thessaloniki - American College of Thessaloniki). This contribution offers a diachronic exploration of allegory not only as a literary device but also as a political instrument that has long mediated relations of truth, power, and deception. Moving from the fables of Aesop to Cold War propaganda and the allegorical framings of contemporary conflicts, the authors highlight how allegory operates simultaneously as a vehicle of concealment and revelation. In medieval Europe, religious allegories such as Dante's *Divine Comedy* worked to reinforce hierarchical authority while purporting to deliver spiritual truth. In the modern period, totalitarian regimes used nationalist allegories to justify ideological purity and to dehumanize perceived enemies. By situating allegory within the history of misinformation, Chatzimavroudi and Moros underline the paradox that the very narrative strategies used for moral instruction can be repurposed to manipulate mass perception. The paper also pays attention to counter-allegories, resistant symbolic forms that expose official distortion and restore critical agency. Their analysis makes a strong case for the importance of allegorical literacy: to understand how symbolic narratives frame perception and can serve both emancipation and domination. This historically grounded reflection provides the necessary foundation for the more contemporary studies that follow in the volume.

The second paper, *Addressing the Impact of Medical Misinformation* by Przemysław M. Waszak and Paweł Zagożdżon (Medical University of Gdańsk), turns to one of the most pressing contemporary challenges: the spread of false health information. Drawing on public health research, epidemiology, and evidence-based medicine, the authors outline how misinformation undermines vaccination efforts, distorts policy, and exacerbates social inequalities. They distinguish between prevention strategies (such as media literacy campaigns and the use of machine learning for early detection) and correction strategies (such as debunking or counter-framing false claims). Their analysis reviews interventions by health authorities including the WHO, CDC, and NIH, highlighting both their achievements and their declining momentum in the post-COVID period, as well as uneven effectiveness across global contexts. Importantly, the paper also addresses debates over professional accountability, noting the tension between calls to discipline healthcare professionals who spread misinformation and concerns over free speech protections. One of the central insights is that misinformation is not merely a communication problem but a symptom of deeper social issues, such as institutional distrust and health inequities. Tackling it therefore requires systemic, government-led interventions that rebuild trust and promote equitable access to reliable knowledge.

From health misinformation, the discussion moves to the epistemology of conspiracy theories. *How Can We Perceive Conspiracy Theories in Order to Diversify Existing Typologies?* by Ioannis Elissaios Paparrigopoulos, Dora Katsamori, and Georgios Petasis (Institute of Informatics and Telecommunications - NCSR Demokritos) interrogates the classification systems that researchers have traditionally used to make sense of conspiracy narratives. The authors argue that current typologies often reflect outdated or overly rigid assumptions, treating conspiracy theories merely as irrational deviations rather than as cultural narratives embedded in broader social contexts. Drawing on insights from computational folkloristics, geospatial analysis, and narrative theory, the paper proposes alternative frameworks that account for the structural, cultural, and political dimensions of conspiracy theorizing. For example, the authors suggest using network analysis and shallow ontologies to map recurring motifs, clusters, and geographical patterns of circulation. This approach promises not only to diversify typologies but also to restore a degree of legitimacy to conspiracy theories as social phenomena worth systematic study rather than mere aberrations. Their work resonates with current debates about whether the fight against disinformation should suppress or rather critically engage with the narrative forms that sustain it.

The section on misinformation and disinformation concludes with *Multimodal Interpretation of a Mobbing Paradigm in the School Workplace* by Georgios Elias Potamias and Paraskevi Kanari (Hellenic Open University). This paper explores the subtle ways in which disinformation functions within organisational life, focusing on a real-world case of workplace mobbing in an international school. Using multimodal discourse analysis that combines linguistic, visual, and contextual evidence, from emails and meeting transcripts to non-verbal cues, the authors identify how vague accusations, misleading statements, and selective truth-telling are mobilized to stigmatize and ultimately expel a targeted individual. They propose a five-stage model of mobbing (social exclusion, stigmatization, critical incident, adjudication, elimination), demonstrating how disinformation plays a central role in each stage. The study's originality lies in showing that misinformation is not only a media or political phenomenon but also a pervasive dynamic within institutions. By applying tools of critical discourse and social semiotics, Potamias and Kanari argue for the necessity of early detection and intervention, stressing the value of multimodal methodologies for uncovering hidden forms of psychological aggression.

The paper *Are They Yielding Results? Examining the Relevance of Fact-Checking Training to Information Verification Competence of Nigerian Journalism Students* by Saheed Bidemi Ibrahim (Birmingham City University) and Rukayat Abdullahi Taiwo

(University of Salford) investigates the integration of fact-checking into journalism education. Based on surveys of 271 students and interviews with course coordinators in Nigerian universities, the study paints a sobering picture: although external organisations such as DUBAWA and Africa Check offer workshops and fellowships, formal curricula remain largely devoid of systematic training in verification practices. Students report only moderate competence in using basic tools like Google Maps, and very low competence in advanced techniques such as deepfake detection or forensic video analysis. Interviews with lecturers confirm a lack of structured training, often leaving instructors themselves unprepared to teach fact-checking skills. The findings suggest a cycle of limited institutional support, insufficient educator expertise, and inconsistent student exposure. Ibrahim and Taiwo argue for curriculum reform and institutional partnerships that would embed fact-checking at the core of journalism education. Their study makes a crucial contribution by documenting educational gaps in the Global South, demonstrating that the fight against disinformation must be contextualized in local pedagogical and infrastructural realities.

Technological innovation features prominently in *News, Social Media and Video Analytics: the MediaPot Platform* by Elisavet Palogiannidi, Sotiris Legkas, Dimitrios Vogiatzis, Manolis Mylonas, Vasso Koutsoupia, Vasileios Mezaris, Spyridoula Markou, George Zissis, and Pantelis Theodosiou (NCSR Demokritos, Deree American college of Greece, CERTH, ATC). This paper presents the architecture and applications of MediaPot, a platform designed to support journalists in managing the overwhelming flow of digital information. Integrating natural language processing, deep learning, and video summarization techniques, MediaPot enables the analysis of text, images, and video from multiple sources. Its modules include entity recognition, semantic graph construction, retrieval-augmented question answering, and multimodal similarity search. Importantly, the platform not only accelerates newsroom workflows but also embeds media literacy principles by making patterns of disinformation and manipulation visible to practitioners. Case studies illustrate how journalists can trace misinformation across networks and verify sources in real time. While primarily technical, the paper insists that technological solutions must remain accessible and transparent to ensure adoption in diverse newsrooms. As such, MediaPot represents an example of human-centered AI innovation in journalism, aiming not to replace but to augment the journalist’s capacity for critical evaluation.

In *Harnessing Emerging Technologies to Combat Disinformation: Innovation, Ethics, and Resilience in Journalism*, Ioanna Georgia Eskiadi and Nikolaos Panagiotou (Aristotle University of Thessaloniki) broaden the discussion to examine how artificial intelligence, blockchain, and immersive media are reshaping journalism’s capacity to counter disinformation. Through comparative case studies of five international news organisations, the paper evaluates both the promise and the pitfalls of these technologies. AI supports real-time verification and anomaly detection, blockchain secures provenance and accountability, while immersive storytelling deepens public engagement by fostering empathy and memory retention. Yet, challenges abound: algorithmic bias, scalability limitations, and ethical concerns about manipulating audiences through immersive content. Eskiadi and Panagiotou argue that no single technology can eradicate disinformation; rather, their combined use, anchored in transparency and ethical frameworks, can reinforce journalism's democratic mission. Their research situates technological innovation within broader questions of trust, ethics, and inclusivity, reminding us that resilience to disinformation is as much a cultural and ethical challenge as a technical one.

The journalism section concludes with *The News Weight Coefficient (NeWC) for Measuring the Accuracy of Discrimination Between False and True News with Human and Non-Human Authors* by Teodor Răileanu-Olariu (University of Iași) and Bogdan Oprea (University of Bucharest). This study addresses a methodological gap in misinformation research: the lack of standardized procedures for selecting comparable sets of true and false news. The authors propose NeWC, a coefficient that quantifies the weight and potential impact of news items based on plausibility, intentionality, and territorial relevance. Using a sample of 386 Romanian participants, the study tested human- and AI-generated news stories across domains such as health, economy, and geopolitics. Results indicate that while respondents often identified false news more accurately than true news, they exhibited higher confidence when classifying content as true, a paradox with profound implications for susceptibility to disinformation. Interestingly, participants showed slightly greater sensitivity to AI-generated content than to human-authored articles, suggesting that stylistic differences remain perceptible. By combining signal detection theory with a replicable selection algorithm, this paper offers researchers and practitioners a robust tool for evaluating both human and machine-produced content.

From the Theory of Critical Digital Literacy to Pedagogical Practice in HE: Combating Disinformation in the S.HI.E.L.D. vs Disinfo project is co-authored by Panagiota Samioti, Konstantinos Sipitanos, Mari-Liis Madisson, Andreas Ventsel, Dimitris Kotzinos, Georgios Terzis and Christos Gavalas (University of Crete, University of Tartu, Cergy Paris University, Brussels School of Governance, Athens Technology Center). It presents the intellectual and practical journey of the Erasmus+ S.HI.E.L.D. vs Disinfo project, a collaboration of five European universities and industry partners. The project's cornerstone is a Critical Digital Literacy (CDL) framework, accompanied by a quality assurance matrix, curriculum, anthology of teaching scenarios, and a digital toolkit for educators. The authors describe the process of moving from conceptual frameworks to concrete classroom applications, including focus groups with professionals, comparative analysis of EU policy texts, and pilot courses across different disciplines. What emerges is a holistic model of disinformation education that integrates contextual analysis, detection skills, and pedagogy. The paper underscores the importance of interdisciplinarity, scalability, and cross-national collaboration, offering a transferable blueprint for higher education institutions seeking to embed CDL in their curricula.

In *CDL and Critical Creativity: Engaging Students with AI to Counter Disinformation in Higher Education*, Panagiota Samioti and Rafail Giannadakis (University of Crete) focus on a university-level course piloted at the University of Crete under the S.HI.E.L.D. vs Disinfo project. The course combined academic language development with disinformation training, engaging 51 students from humanities, social sciences, and STEM. Through tools such as ChatGPT, Truly Media, and three custom-designed evaluation protocols (PROSPECT, APPEAL, ELEVATE), students developed a multidimensional toolkit for assessing sources and rhetorical strategies. Importantly, the course integrated what the authors call “critical creativity”, the use of imaginative, project-based assignments (e.g., informational leaflets, posters) to translate analytical insights into civic engagement. The paper reflects on student feedback and classroom documentation, showing how AI can be productively integrated into CDL without displacing human judgment. Samioti and Giannadakis argue that creativity is not peripheral but central to cultivating resilient, critically literate citizens.

Source Evaluation in a Digital World: From Information Consumers to Critical Evaluators by Zoi A. Traga Philippakos (University of Tennessee) addresses one of the most urgent pedagogical needs of our time: equipping students to critically assess both human- and AI-generated content. The paper traces the evolution of source evaluation strategies from the traditional CRAAP test to newer approaches such as lateral reading, highlighting their limitations when applied to algorithmic text generation. Philippakos advocates for explicit instruction, modeling, and collaborative learning, where students practice evaluating real-world digital texts with guided support. She also stresses professional development for teachers, noting that many educators feel unprepared to teach digital literacy in an AI-mediated environment. The paper’s contribution is both conceptual and practical: it situates evaluation within a broader framework of digital citizenship while providing tangible pedagogical strategies for classrooms.

The Immunity Illusion: Media Literacy in an Age of Identity-Driven Disinformation by Massimo Flore (Independent Researcher) offers a provocative critique of prevailing assumptions in media literacy. The author argues that many initiatives assume that exposure to literacy training creates a form of “immunity” against manipulation. In reality, identity, ideology, and emotion often override rational evaluation. Drawing on political psychology and case studies of online polarization, Flore demonstrates how disinformation thrives not because people lack skills but because narratives resonate with their sense of belonging. The paper calls for a paradigm shift in media literacy: from purely cognitive models to approaches that account for affective and identity-based dimensions. By reframing literacy as an ongoing negotiation of values and allegiances, Flore highlights the limits of technocratic solutions and urges educators to confront the sociopolitical roots of disinformation.

In *Critical Literacy as a Tool for Approaching the Rhetoric of Multimodal Texts in Secondary Education: A Teaching Approach to Media Analysis in the Classroom*, Anna Paraskevaidi (Secondary Education Teacher) explores how high school students can develop critical media literacy by analyzing and redesigning advertisements. The study, conducted with 16-year-old students in Greece, combines rhetorical analysis, multimodal semiotics, and creative redesign to expose persuasive strategies and stereotypes embedded in advertising. Students engaged in counter-advertising projects, historical comparisons of smoking campaigns, and AI-assisted poster creation. The findings reveal that such practices not only sharpen students' analytical skills but also empower them to question consumerist ideologies and social stereotypes. Paraskevaidi emphasizes the importance of integrating AI tools responsibly in classrooms, both as objects of critique and as instruments for creative resistance. Her work demonstrates how critical literacy can move students from passive consumption to active, reflective participation in media culture.

Enhancing Media Critical Thinking Against Disinformation with a Socratic Chatbot: A methodology by Panayiotis Kapetanakis, George Petasis, Dora Katsamori, Eirini Papachristou, and Giannis Paparigopoulos (NCSR Demokritos) introduces an innovative "Socratic Coach" designed within the Horizon Europe TITAN project. Unlike fact-checkers that provide verdicts, this chatbot engages users in Socratic dialogue to guide them through tactic-specific questioning, nudges, and metacognitive prompts. Built on a retrieval-augmented generation architecture, the system emphasizes transparency and avoids AI "hallucinations." Findings from a pilot study with researchers show that while users appreciated the reflective approach, they also expected definitive answers, highlighting the tension between pedagogy and verification. The authors conclude that the Socratic method can cultivate durable media critical thinking by focusing on manipulation recognition rather than belief confrontation. The paper situates AI not as an authority but as a coach, complementing rather than replacing fact-checking.

Zoi A. Traga Philippakos' *Scaffolding Thought: Designing AI Tools that Empower Writers, Not Replace Them* extends her long-standing research on writing pedagogy into the domain of AI. Philippakos explores how AI can be used to support strategic reading and writing processes rather than shortcut them. She warns against overreliance on generative tools that may undermine metacognitive engagement and instead proposes design principles for AI applications that scaffold learners' thinking. Through examples of writing assignments and teacher modeling, the paper argues that AI should be embedded within explicit pedagogy, allowing students to see how tools can support but not substitute the intellectual work of writing. This contribution deepens the literacy section by reminding us that digital literacy is inseparable from broader educational practices of thinking, composing, and reflecting.

The literacy section closes with *How to Critically Identify Fake News in Primary School: A Case Study* by Ioanna Kitsou (6th Intercultural Primary School of Evosmos).

This paper addresses one of the most challenging questions: how to introduce the complexity of disinformation to young learners. Kitsou presents a case study of a classroom intervention where primary school students analyzed fake news examples, practiced source verification, and created their own critical responses. The methodology combined age-appropriate scaffolding with interactive activities, showing that even young children can learn to question information sources when given the right tools. The findings emphasize the importance of early education in digital literacy, demonstrating that habits of skepticism and inquiry can and should be cultivated from the beginning of formal schooling.

Taken together, the contributions reveal that disinformation is not only a technological or cognitive issue but also one of trust, identity, and power. By uniting historians, educators, journalists, computer scientists, and philosophers, this volume aims not only to offer analytical clarity and practical resources for addressing today's information challenges, but also to foster an ongoing, interdisciplinary conversation on critical literacy, trust, and democratic participation in digital societies.

References

Association of College and Research Libraries. (2016). *Framework for information literacy for higher education.* American Library Association. https://www.ala.org/acrl/standards/ilframework

Caceres, M. M. F., Sosa, J. P., Lawrence, J. A., Sestacovschi, C., Tidd-Johnson, A., Ul Rasool, M. H., Gadamidi, V. K., Ozair, S., Pandav, K., Cuevas-Lou, C., Parrish, M., Rodriguez, I., & Fernandez, J. P. (2022). The impact of misinformation on the COVID-19 pandemic. In *AIMS Public Health*, 9(2). https://doi.org/10.3934/PUBLICHEALTH.2022018

Cherner, T. S., & Curry, K. (2019). Preparing Pre-Service Teachers to Teach Media Literacy: A Response to "Fake News." *Journal of Media Literacy Education*, *11*(1). https://doi.org/10.23860/JMLE-2019-11-1-1

European Commission. (2022). *Communication to the commission: European Commission digital strategy Next generation digital Commission*. https://commission.europa.eu/system/files/2022-06/c_2022_4388_1_en_act.pdf?utm_source=chatgpt.com

European Digital Media Observatory. (2024). *EDMO Guidelines for Effective Media Literacy Initiatives*. Publications Office of the European. Https://Edmo.Eu/Areas-of-Activities/Media-Literacy/Raising-Standards-the-Edmo-Guidelines/

Friedman, J. (2023). Post-Truth and the Epistemological Crisis. In *Critical Review*, 35(1–2). https://doi.org/10.1080/08913811.2023.2221502

Godwin, E., Davidson, B. I., Hill, T., & Joinson, A. (2025). Internet Memes as Stabilizers of Conspiracy Culture: A Cognitive Anthropological Analysis. *Social Media + Society*, *11*(1). https://doi.org/10.1177/20563051241306421

Hameleers, M. (2023). Disinformation as a context-bound phenomenon: toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination. *Communication Theory*, *33*(1). https://doi.org/10.1093/ct/qtac021

International Commission on the Futures of Education. (2021). *Reimagining our futures together: a new social contract for education*. UNESCO. https://doi.org/10.54675/ASRB4722

Leaning, M. (2017). *Media and information literacy: An integrated approach for the 21st century.* Chandos Publishing.

Martyushev, A., Nikonova, N., Savchenko, Z., Bovkun, I., & Kerimov, A. (2021). Critical thinking in media sphere: Attitude of university teachers to fake news and its impact on teaching. *Journal of Management Information and Decision Sciences*, *24*(Special issue 1), 1-12.

Pangrazio, L. (2016). Reconceptualising critical digital literacy. *Discourse: Studies in the Cultural Politics of Education*, *37*(2), 163–174. https://doi.org/10.1080/01596306.2014.942836

Sasahara, K., Chen, W., Peng, H., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2021). Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science*, *4*(1), 381–402. https://doi.org/10.1007/s42001-020-00084-7

Singer, P. W., & Brooking, E. T. (2018). *LikeWar: The weaponization of social media.* Houghton Mifflin Harcourt.

Sipitanos, K. (2023a). Raising awareness against fake news to protect democracy: the myth of Islamophobia in Trump's speech. *Social Semiotics*, *33*(4). https://doi.org/10.1080/10350330.2021.1929147

Sipitanos, K. (2023b). Students' Discourses during the Online Distance Learning in the First Wave of the Covid 19 Pandemic: An Action Research with Students' as Co-researchers. *Turkish Online Journal of Distance Education*, *24*(2). https://doi.org/10.17718/tojde.967406

Sipitanos, K. (2023c). Students' resistance to powerful digital educational platforms through digital discourse analysis and participatory action research. In J. Haidar (Ed.), *Cosmosemióticas de la resistencia y decolonialidad* (pp. 47–72). sb Escuela National de Antropología e Historia.

Tejedor, S., Cervi, L., Pérez-Escoda, A., & Jumbo, F. T. (2020). Digital literacy and higher education during COVID-19 lockdown: Spain, Italy, and Ecuador. *Publications*, *8*(4). https://doi.org/10.3390/publications8040048

UNESCO. (2013). *Global media and information literacy (MIL) : assessment framework : country readiness and competencies*. United Nations Educational Scientific and Cultural Organisation (UNESCO).

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*. https://doi.org/10.1126/science.aap9559

Vuorikari, R., Punie, Y., Carretero, S., & Van Den Brande, L. (2016). *DigComp 2.0: The digital competence framework for citizens*. European Commission, Joint Research Centre. https://doi.org/10.2791/11517

Allegory in the Service of Power (?): Historical Uses of Misinformation and Disinformation

Eleni Chatzimavroudi[1], Marios-Kyparissis Moros[2]
Aristotle University of Thessaloniki, American College of Thessaloniki,
1. elene@anatolia.edu.gr
2. marioskmoros@gmail.com

Abstract

Allegory has long constituted a pivotal literary and rhetorical device, facilitating layered critique and symbolic expression. Its deployment, however, extends beyond artistic or intellectual domains; historically, allegory has also operated as an instrument of misinformation and disinformation. This paper investigates the ways in which allegory has been mobilized in the public sphere to distort historical events, reinforce ideological constructions and manipulate mass perception, while simultaneously retaining the paradoxical capacity to disclose deeper truths under fictional guise. Medieval religious allegories for example functioned to consolidate hierarchical authority, whereas Cold War propaganda rendered ideological adversaries in absolute moral terms. Particularly salient is the nationalist allegory of twentieth-century totalitarian regimes, wherein symbolic depictions of the "enemy" legitimized extensive disinformation strategies. Contemporary analogues are likewise considered, notably the allegorical framing of geopolitical conflicts in modern media. Such representations frequently simplify complex realities and perpetuate misinformation, yet counter-allegories simultaneously emerge to contest hegemonic discourse. Due to tracing these dynamics, the ethical implications of allegory in public communication are questioned and the necessity of critical literacy for navigating symbolic constructions is emphasized. While allegory remains integral to cultural narration, its dual role -as both a vehicle of deception and a conduit of truth- requires sustained critical scrutiny in an era of accelerated information flows.

Keywords: allegory, misinformation, disinformation, historical narratives, propaganda, public perception

Introduction

In attempting to explore the parameters of misinformation comprehensively, this article foregrounds allegory not merely as a stylistic device, but as a conceptual lens which allows the dynamics of distortion, deception and ideological manipulation to get critically examined. Allegory, by its very nature, engages in displacement: it says one thing and means another, producing layered significations beneath narrative or imagery. This dual movement of concealment and revelation renders allegory especially pertinent to the study of misinformation, which similarly operates through surfaces that obscure underlying mechanisms of power. Traditional studies of allegory, such as those by Fletcher (1964) and Quilligan (1979), have underscored its role in sustaining symbolic modes of authority, while De Man (1983) highlighted its rhetorical indirection and Jameson (2010) explored its ideological force in cultural production. In parallel, scholarship on misinformation has increasingly examined the rhetorical strategies of distortion and deception, particularly in relation to the "post-truth" condition and epistemic challenges posed by conspiracy narratives. Yet the specific intersection of allegorical representation and misinformation has received limited sustained attention. The central research question addressed here, therefore, is how allegory has functioned historically and contemporaneously as both a mechanism of disinformation and a vehicle of counter-narratives. The article proceeds by outlining the theoretical framework, presenting key historical and contemporary case studies, and concluding with a discussion of the ethical implications and the necessity of critical literacy.

In this context, allegory is not only a figure of speech but a mode of mediation, one that allows for the communication of truths -and untruths- through indirect means. By adopting allegory as both subject and method, we aim to unsettle the presumed transparency of discourse, foregrounding the symbolic operations that underwrite the circulation of belief, persuasion, and narrative authority (Dyer, 2002). Allegory thus becomes both an object of inquiry and a critical tool, allowing us to trace how meaning is displaced, masked, and often re-instrumentalized in the service of ideological agendas.

Beyond the definition, what is interesting in relation to our broader purpose is the aim of this particular finding. In Aesop's fables, where it is widely used as their structural basis, it aims to teach a deeper truth about social life through the innocent construction of a narrative related to the world of animals and nature in general. If we add to the picture the legend that accompanied the figure of Aesop, a slave, at least according to tradition, of the Archaic period, the concept of censorship and the successful attempt to avoid it with this seemingly 'innocent' device appears to serve the truth. It is therefore a figure that works inductively in the sense that it systematically and implicitly guides the reader from a commonly accepted -and quasi self-evident- story to the approach of a socio-political issue by analogy. Let us explain:

a. Systematically, because it differs from a symbol – a symbol refers from a specific element to an abstract concept (e.g. the dove symbolises peace), while analogy presupposes a complete, albeit elementary, narrative.

b. Implicitly, because it is so clear and explanatory that there is no need to explicitly state the connections with the socio-political issue it seeks to illuminate, connections similar to those used, for example, in similes.

Furthermore, it presupposes simplicity in rendering and organisation, so as to avoid misinterpretations or ambiguities and thus facilitate an almost didactic positioning.

The pedagogical view also defines the unequal relationship between sender and receiver in a traditional context: the creator possesses the truth and attempts to convey it smoothly and without threat to his audience, who are unaware of it.

These specific characteristics therefore justify the functionality of allegory in the case of false information and misinformation. By reversing its archetypal goal, par-allegory can serve a false construct by exploiting precisely:

a. the unequal relationship of power between the sender and the broad audience to which it is addressed

b. the narrative framework that facilitates plausibility

c. the implicit but simplistic analogy that allows for correspondence with the context it seeks to support.

From Fable to Fabrication: The Double Life of Allegory

In light of the preceding analysis, it becomes evident that allegory's seemingly innocuous mechanisms—its inductive structure, pedagogical clarity, and reliance on analogical reasoning—are not limited to literary fable or ancient rhetorical device. Rather, they form the very scaffolding through which broader ideological functions are enacted (De Man 1983). What begins as a didactic tool for revealing truth, as in the case of Aesop's tales, may easily be repurposed to veil it. This duality, whereby allegory serves both clarity and obfuscation, instruction and manipulation, sets the stage for a more expansive interrogation. If par-allegory (or corrupted allegory) can so effectively mimic the epistemological style of truth-telling while anchoring falsehoods in narrative coherence, then its political utility must be taken seriously (Eagleton 1991). The same formal attributes that once allowed the enslaved to bypass censorship can be—and have been—reappropriated by hegemonic systems to naturalize ideology and delegitimize dissent. What follows, therefore, is a shift in scale and scope: from a structural analysis of allegory's internal logic to a historical examination of its instrumentalization across political regimes. Allegory, we argue, is not merely a vehicle of aesthetic communication but a mutable medium of power—capable of encoding both domination and resistance, often within the same symbolic register.

Effective political propaganda often introduces allegorical narratives that become the dominant worldview, establishing the frame for future conversation, demarcating the boundaries of "logical" explanation.

Allegory occupies a liminal and structurally ambiguous position in the epistemology of representation. It simultaneously partakes in the abstract and the narrative, binding conceptual thought to imaginative storytelling. It mediates between the visible and the invisible, the said and the unsaid, enabling the transmission of complex political, ethical, and social content within layered symbolic forms. The signifiers of allegory do not point to a single referent but to a constellation of meanings—historical, cultural, psychological—which makes them extraordinarily potent but also vulnerable to cooption. Historically, this polysemous and indeterminate nature has rendered allegory uniquely susceptible to instrumentalization by hegemonic powers, who deploy symbolic simplifications to naturalize ideology, legitimize violence, and obscure contradiction. At the same time, this very indeterminacy has enabled subaltern and dissident forms of expression: artists, writers, and intellectuals have long turned to allegory to encode counter-hegemonic messages, to smuggle prohibited meanings into the public sphere, and to preserve ethical critique under authoritarian scrutiny.

Allegory, then, is not merely a mirror held up to reality, nor a mask disguising intentions—it is a rhetorical and ideological technology, shaped by and shaping the contours of public discourse. Its historical function must be read through a double lens: as both a technique of domination and a vehicle of imaginative defiance.

Historical Allegories of Power: From Medieval Theology to Cold War Binaries

In the medieval context, for instance, religious allegory functioned simultaneously as catechism and as a sophisticated apparatus of social control. It was a primary mechanism through which doctrinal orthodoxy was disseminated to a largely illiterate populace, embedding theological concepts within narrative structures that were emotionally and morally resonant. Works such as *Piers Plowman* and Dante Alighieri's *Divine Comedy* did not merely illustrate abstract spiritual principles; they configured complex theological doctrine into compelling symbolic narratives that reinforced prevailing hierarchies and moral ideologies (Astell, 1999). The pilgrim's journey through infernal, purgatorial, and paradisiacal realms in Dante's text, for example, allegorized the soul's moral and civic obligations within a framework that upheld both papal authority and imperial order (Kantorowicz, 1957/2016). Far from being neutral or benign, these allegorical structures served to naturalize feudal authority, inscribe divine sanction onto social stratification, and delineate the boundaries between orthodoxy and heresy. Allegory here acted as a didactic force, mobilized by the Church and state alike to generate a sense of cosmic order that was inextricably bound to the maintenance of temporal power. Ecclesiastical use of allegorical sermons, stained-glass windows, and morality plays further embedded these symbolic schemas into the cultural consciousness, providing emotionally resonant and aesthetically potent modes of indoctrination. In this way, social obedience was subtly encoded within the scaffolding of spiritual edification and metaphysical narrative.

The Cold War period marks another epoch wherein allegory was mobilized to polarize global audiences and to structure political perception along sharply moralized binaries. In the West, communism was frequently rendered not merely as an alternative system of governance but as an existential threat to the very fabric of civilization. Allegorical framings—particularly in mass media—conflated ideological difference with moral degeneracy, chaos, and tyranny. Cinematic productions such as *Red Dawn*, *The Hunt for Red October*, and even animated series participated in a cultural project that transmuted geopolitical anxieties into Manichaean allegories of good versus evil, liberty versus oppression. These narratives often simplified complex international dynamics into archetypal struggles, aligning democratic capitalism with virtue and personifying the communist Other as either a monstrous invader or a pitiable victim of totalitarian indoctrination.

Allegorical framing, particularly in mass media, often conflated ideological difference with moral degeneracy, chaos, and tyranny. Cinematic productions such as *Red Dawn* (1984), which depicts a Soviet invasion of small-town America, allegorized Cold War anxieties by casting the United States as the last bastion of freedom under siege (Hendershot, 2003). Similarly, *The Hunt for Red October* (1990) transformed a tale of naval defection into an allegory of ideological redemption, portraying the Soviet submarine captain as a figure who abandons tyranny in favor of Western liberty (Dittmer, 2005). Even animated series such as *G.I. Joe* or *Rocky and Bullwinkle* operated within this framework, translating geopolitical tensions into archetypal struggles of good versus evil, liberty versus oppression (Engelhardt, 1998). In each case, allegory functioned to condense complex international dynamics into easily digestible moral binaries: democratic capitalism was aligned with virtue, while the communist "Other" was personified either as a monstrous invader threatening civilization or as a pitiable subject trapped within totalitarian indoctrination.

In both ideological camps, allegory functioned as an epistemic filter that reduced systemic complexity, reinforced affective allegiance to the nation-state, and naturalized antagonistic worldviews. It provided audiences with a legible moral universe in which geopolitical conflict could be experienced as narrative catharsis rather than analytic challenge (Eco, 1979). In the Soviet context, allegory often depicted capitalist societies as sites of moral decay and exploitation, where symbolic representations of the West reinforced the image of communism as the sole path to justice and equality (Brandenberger, 2011). On the Western side, allegory was equally pervasive: Ronald Reagan's characterization of the USSR as the "Evil Empire" condensed intricate geopolitical rivalries into a biblical struggle between light and darkness, while NATO campaigns and popular media reiterated this Manichaean framing. Allegory's emotional power, in this context, lay in its ability to convert anxiety into moral clarity and doubt into conviction, whether through the Soviet vilification of capitalism or the Western demonization of communism.

Resistance and Mutation: Allegory in Totalitarianism and Contemporary Media

Perhaps the most pernicious uses of allegory are found in the totalitarian regimes of the twentieth century. In National Socialist Germany, the allegorical trope of the corrupted body politic—personified in antisemitic imagery—was deployed to legitimize racial purification and genocidal violence. The figure of the Jew as parasitic Other permeated Nazi visual and textual propaganda, producing a moralized, mythic structure through which existential threat was rendered palpable. Similarly, Stalinist Russia utilized socialist-realist allegory to construct heroic paradigms of labor and sabotage, which in turn masked political purges and failures of central planning. Here, allegory did not merely represent ideology; it performed it, generating the affective infrastructure required for systemic violence.

And yet, allegory is not solely the province of power. Its very capacity for multivalence and symbolic indirection has rendered it indispensable in contexts of censorship and surveillance. George Orwell's *Animal Farm*, for example, is not merely a fable but a politically charged allegory through which Orwell critiques Stalinist totalitarianism without overt denunciation. In Eastern Europe during the Soviet era, authors such as Milan Kundera deployed allegory and irony as strategies of oblique resistance. Their texts called upon readers to engage in interpretive labor—to discern subtexts of critique encrypted beneath layers of fictionality. Allegory, in these instances, became a mechanism of intellectual survival.

Popular culture, including dystopian narratives like *The Handmaid's Tale*, functions as a site where allegorical critique of real-world systems—patriarchy, authoritarianism, biopolitics—is rendered both accessible and affectively potent. Digital media further complicates this ecology: memes, satire, and visual allegories now traverse platforms with remarkable speed, sometimes functioning as incisive critiques of power, at other times reinforcing disinformation.

In light of these dynamics, it is imperative to reconsider allegory as neither inherently subversive nor inherently hegemonic, but rather as a contested terrain wherein semiotic strategies are mobilized for diverse—and often contradictory—ends. To approach allegory critically is to interrogate not only its content but also its structure, its rhetorical effects, and the conditions of its reception. Who speaks through the allegory? To whom is it addressed? What forms of authority or dissent does it make possible?

Conclusion

Allegory, as this study has shown, is never a neutral device. It has consistently operated at the threshold between revelation and distortion, allowing political and cultural actors to encode persuasive narratives that obscure as much as they illuminate. Throughout history, allegorical forms have served as scaffolding for misinformation, disinformation, and propaganda, reshaping collective perception through emotionally resonant symbolism. In this sense, allegory exemplifies what De Man (1983) described as rhetorical indirection and what Jameson (2010) identified as the ideological force of cultural production: a mode that conceals its operations even as it naturalizes power.

Yet allegory's indeterminacy also enables its critical potential. Precisely because it "says other things," allegory provides space for counter-narratives, irony, and subversive critique. From Orwell's *Animal Farm* to the allegorical fictions of Kundera, allegory has allowed suppressed voices to resist dominant ideologies under conditions of censorship. Its dual nature, complicit with hegemonic propaganda yet simultaneously fertile ground for dissent, underscores why allegory must be central to the study of misinformation today.

Reconnecting with the article's opening concerns, it becomes clear that misinformation and disinformation are not merely failures of fact-checking but phenomena deeply entangled with symbolic mediation. Through the metaphor of the "infodemic," falsehoods spread epidemically, thriving on affective and cognitive vulnerabilities. Allegory intensifies this process by embedding distortions within culturally familiar narratives, rendering them persuasive beyond their factual content. At the same time, studying misinformation requires us to move past simplistic models and attend to the methodological complexity of its circulation, a challenge to which allegory, with its layered and unstable meanings, is uniquely relevant.

Promising directions for future research lie in three domains. First, comparative studies of allegorical propaganda across political regimes may reveal recurrent symbolic tropes that persist despite ideological differences. Second, the digital sphere, where memes, viral images (Ellenius, 1998), and algorithmic amplification depend on allegorical condensation (Barthes, 1957/2009), demands systematic analysis to understand how allegory fuels both disinformation campaigns and counter-hegemonic critique. Third, pedagogy offers an ethical horizon: allegory could be mobilized not only to deceive but also to cultivate critical literacy, training readers to decode symbolic narratives and recognize manipulation.

Ultimately, allegory's entanglement with misinformation, disinformation, and propaganda requires what Jameson (1981/2002) called a "political unconscious" approach: one that uncovers the ideological work performed by narrative structures beneath their surface. To study allegory under contemporary conditions is therefore to equip ourselves with the critical tools necessary to navigate a world saturated with symbolic persuasion. In an age when information circulates at unprecedented speed, and when propaganda re-emerges in both old and new guises, the interpretive labor of reading allegory becomes not just an academic task but an ethical imperative.

References

Astell, A. W. (1999). *Political Allegory in Late Medieval England*. Cornell University Press.

Barthes, R. (1957/2009). *Mythologies*. New York: Vintage.

Brandenberger, D. (2011). *Propaganda state in crisis: Soviet ideology, indoctrination, and terror under Stalin, 1927–1941*. Yale University Press.

De Man, P. (1983). *Blindness and insight: Essays in the rhetoric of contemporary criticism* (2nd ed.). University of Minnesota Press.

Dittmer, J. (2005). Captain America's empire: Reflections on identity, popular culture, and post-9/11 geopolitics. *Annals of the Association of American Geographers, 95*(3), 626–643.

Dyer, R. (2002). *The Matter of Images: Essays on Representations.* Routledge.

Eagleton, T. (1991). *Ideology: An introduction*. London. Verso.

Eco, U. (1979). *The Role of the Reader: Explorations in the Semiotics of Texts*. Indiana University Press.

Ellenius, A. (1998). *Iconography, Propaganda, and Legitimation.* Oxford University Press.

Engelhardt, T. (1998). *The end of victory culture: Cold War America and the disillusioning of a generation*. University of Massachusetts Press.

Fletcher, A. (1964). *Allegory: The theory of a symbolic mode*. Cornell University Press.

Hendershot, C. (2003). *Anti-Communism and Popular Culture in Mid-Century America*. McFarland.

Jameson, F. (1981/2002). *The Political Unconscious: Narrative as a Socially Symbolic Act*. Routledge.

Jameson, F. (2010). *Allegory and ideology*. Verso.

Kantorowicz, E. H. (1957/2016). *The King's Two Bodies: A Study in Medieval Political Theology*. Princeton University Press.

Quilligan, M. (1979). *The language of allegory: Defining the genre*. Cornell University Press.

Addressing the Impact of Medical Misinformation

Przemysław M. Waszak[1], Paweł Zagożdżon[2]

1, 2. Division of Hygiene and Epidemiology, Medical University of Gdańsk, Poland

Abstract

Health misinformation and disinformation pose serious threats to public health by undermining vaccination efforts, distorting disease control strategies, and eroding trust in healthcare systems. False medical claims, ranging from exaggerated treatments to conspiracy theories, contribute to public anxiety, the misallocation of resources, and increasing social division. These effects are particularly pronounced among individuals with lower education levels, limited health literacy, or low trust in institutions, who are especially vulnerable to misleading information. While misinformation spreads rapidly through social media, exposure alone does not always lead to belief. Rather, susceptibility depends on a combination of emotional, cognitive, and social factors. In response, addressing health misinformation requires a twofold strategy: prevention and correction. Preventive measures include technologies like machine learning for early detection, as well as broad media literacy initiatives. At the same time, effective correction depends on communication that is credible, clear, and sensitive to cultural and contextual factors.

Health organisations such as the World Health Organisation have taken action through myth-busting campaigns, educational resources, and partnerships with digital platforms. Despite these efforts, momentum has declined in the post-COVID period, and outreach in non-Western regions remains limited. To make meaningful progress, systemic, government-led strategies are crucial. Rebuilding trust and ensuring equitable access to accurate health information requires sustained effort. This includes incorporating health education into school curricula, increasing institutional transparency, and promoting culturally appropriate communication. Ultimately, reducing the long-term harm caused by health misinformation means addressing its deeper roots: social inequality, low health literacy, and institutional distrust. Tackling these issues is key to strengthening public resilience in the face of future health crises

Keywords: public health, disinformation, misinformation, health education, evidence-based medicine

---

[1] Corresponding author: Przemysław Waszak, Division of Hygiene and Epidemiology. ul. Dębinki 7, Gdańsk, Poland, p.waszak@gumed.edu.pl, tel./fax. +48583446085

Introduction

Research has shown that the spread of health misinformation, disinformation, and fake news can have a range of harmful effects. These include decreasing public willingness to receive vaccinations, hindering efforts to control disease outbreaks, and even physically disrupting access to healthcare services. Beyond the health sector, such false information can also deepen political instability by fostering social division and distrust. Moreover, it contributes to heightened levels of fear, anxiety, and mental health issues within communities. Resources may be misdirected due to misleading narratives, while official responses are often weakened or delayed. In such an environment, the production and dissemination of low-quality or misleading health content are further encouraged (Borges do Nascimento et al., 2022). To counter this, it is essential to emphasize that medical information should be grounded in science-rooted in valid methodology, robust evidence, and clearly distinguished from non-scientific claims. Interpreting health information requires careful evaluation of its truth, justification, and context. A particularly important aspect is understanding causality, which lies at the core of medical reasoning.

In clinical practice, Evidence-Based Medicine (EBM) plays a central role by guiding decisions through a balance of scientific research, patient values, and clinical judgment (Guyatt, 2014). Complementing this, epidemiology provides tools to analyze patterns and causes of diseases, supporting assessments of risk and benefit through the interpretation of complex data. However, even scientifically grounded claims can become subjects of controversy. This is evident in ongoing debates around cholesterol, hormone therapy, and various COVID-19 policies. Overstated or poorly communicated findings, no matter how well-intended, can undermine public trust. When unreliable medical information spreads, it not only misleads but may also damage public health and erode the foundations of democratic decision-making (Harris, 2024; Lewis, 2023). Further compounding the issue, studies have shown that health misinformation takes many forms. These range from exaggerated treatment claims to entirely unfounded conspiracy theories. Analyses of content across social media platforms have revealed widespread falsehoods, including myths about alternative medicine, inaccurate dietary advice, and narratives driven by ideological or conspiratorial beliefs (Okoro et al., 2024). According to a 2021 meta-analysis, the proportion of COVID-19 misinformation circulating on social media ranged from 0.2% to 28.8% of all posts (Gabarron et al., 2021). This suggests that exposure to medical disinformation is not rare. It may, in fact, be a common experience for users. Still, it is crucial to note that exposure does not automatically translate into belief (Li & Yang, 2025); susceptibility is shaped by individual, social, and contextual factors.

Individuals are more likely to fall victim to medical disinformation when they have low trust in authorities, limited health knowledge, or a strong belief in conspiracy theories. High exposure to social media and the resulting information overload further increase susceptibility, as they often lead people to rely on simplistic or emotionally charged content. Importantly, even healthcare professionals are not immune, especially when their confidence is shaken or institutional trust is undermined. Research consistently shows that trust in institutions is a key factor in building resilience against disinformation (Zhang et al., 2024). These psychological and behavioral factors often correlate with broader demographic patterns. A typical target of medical misinformation is likely to be someone with lower levels of education, limited health literacy, or, more frequently, an elderly person (Scherer et al., 2021). Crucially, this vulnerability predates the digital age. Even before the rise of social media and widespread misinformation, individuals with less education already experienced shorter life expectancies and poorer overall health outcomes (Balaj et al., 2024). The current infodemic has only exacerbated these long-standing health inequalities. Given this context, efforts to counter health misinformation must go beyond short-term corrections. A comprehensive strategy is needed, such as one that addresses the social, educational, and structural determinants of vulnerability. Tackling misinformation effectively means recognizing that it is not merely a communication problem but a symptom of deeper disparities within the population. Central to such a strategy is ensuring equitable access to reliable health information, which forms a cornerstone of primary prevention. Primary prevention aims to maintain health and avert the onset of disease before it occurs. It includes interventions such as health education, vaccination, lifestyle changes, and environmental improvements that collectively reduce risk factors (Abdul Raheem, 2023). When individuals and communities are empowered with trustworthy information, they are better equipped to understand health risks, make informed decisions, recognize early warning signs, and seek timely care. This not only helps reduce morbidity and mortality across a range of diseases, but also builds long-term resilience in public health systems.

This paper explores the following key questions: the strategies implemented by health authorities and organisations to prevent and correct health misinformation, and their effectiveness across different contexts; the ways in which individual factors, such as trust in institutions, health literacy, and demographic characteristics, influence vulnerability to false medical claims; and the role of systemic, government-led actions in building long-term societal resilience against medical disinformation, particularly through education, institutional transparency, and equitable access to credible health communication.

Findings

Efforts to address health misinformation focus on two main strategies: preventing exposure and correcting misperceptions. Prevention includes using machine learning tools to detect misinformation and improving media literacy, though the latter has shown mixed results. Correction strategies involve direct methods (clearly debunking false claims) and indirect methods (presenting balanced arguments or using emotional appeals). The effectiveness of correction depends on factors like message clarity, source credibility, and public trust—particularly in government and professional institutions. Community support and peer influence also play important roles in countering misinformation (Zhang et al., 2024). While individual-level strategies focus on shaping public perception and behavior, institutional responses are equally vital. Health organisations play a central role in operationalizing these strategies on a scale, using their authority and resources to amplify accurate information and counteract misinformation.

Responses from Health Organisations

Health organisations, both national (e.g., American Medical Association) and international (e.g., World Health Organisation, WHO), play a critical role in countering medical misinformation. They promote accurate information, support healthcare professionals, and engage with the public to strengthen trust in health communication (Office of the Surgeon General (OSG)., 2021). Organisations like WHO, Centers for Disease Control and Prevention (CDC), and National Institutes of Health provide accessible, evidence-based resources. WHO created myth-busting tools and toolkits for misinformation during COVID-19 (World Health Organisation, 2022). CDC maintains up-to-date guidance and media toolkits (Centers for Disease Control and Prevention, 2023), while NIH promotes public health education and supports platforms like PubMed. In response to COVID-19, health agencies partnered with platforms like Twitter, Facebook, and Google to direct users to reliable sources. WHO's partnership with Google reached millions, combating misinformation through promoted content and search redirection (Germani et al., 2022). Campaigns such as WHO’s Science in 5 and CDC’s Vaccinate with Confidence aimed to increase public understanding and trust in health interventions. The United Nations Children's Fund and others have promoted digital literacy to help young people critically evaluate online health content (Nascimbeni & Vosloo, 2019). Despite these proactive efforts, significant challenges remain in sustaining effective health communication.

Health Communication Challenges

Despite early momentum, health organisations' collaboration with social media has slowed since the end of the COVID-19 emergency. Misinformation persists on topics like vaccines, opioids, and smoking. Engagement with non-Western platforms (e.g., WeChat, VK, Telegram) and outreach in diverse languages and contexts remains limited (Harrington & Record, 2023). Emotional reactions such as ridicule toward health agency posts suggest continued distrust, highlighting the need for renewed strategies and better cultural adaptation (Waszak et al., 2024).
As institutional trust continues to erode and traditional interventions lose traction, questions arise about how far professional accountability can, and should, go in

curbing misinformation. This tension is especially evident in debates over the regulation of healthcare professionals' online speech. Some professional bodies, like the American Medical Association, advocate disciplinary measures against doctors spreading misinformation. While state boards have revoked licenses, regulating online speech remains challenging due to free speech protections (Saver, 2023; Yang & Schaffer DeRoo, 2022). These challenges point to a deeper, systemic issue: combating health misinformation cannot rely solely on isolated interventions or professional accountability. Broader, coordinated action is needed to address structural barriers and promote long-term resilience through education, transparency, and public trust.

The Need for Systemic Action

While institutional responses have laid a foundation, systemic challenges continue to limit impact. Reliable health information is a vital component of primary prevention strategies that aim to reduce disease burden and improve population health. Ensuring equitable access to accurate, understandable, and culturally appropriate health information empowers individuals to take proactive steps in disease prevention, ultimately reducing morbidity and mortality across diverse health conditions.

It is difficult to expect educational and disinformation campaigns to be solely the work of third sector and non-governmental organisations. In most developed countries, it is government organisations that finance and organise the health care and education systems, so activities related to raising health awareness and combating disinformation should also be a government priority. But there are also areas outside of education itself. Low transparency and complexity of the healthcare system contribute to a decline in trust in medical institutions and health authorities, which is one of the key factors driving belief in conspiracy theories and misinformation. Patients who experience difficulties in accessing clear and reliable information are more likely to turn to alternative, often unverified sources of information (Bobier & Obeid, 2025; Kisa & Kisa, 2025).

Comprehensive health education programs, combined with other measures such as promoting physical activity and healthy eating, improve health-promoting behaviors and reduce the risk of making harmful health decisions, which indirectly limits the impact of misinformation. Health education in schools provides students with the knowledge and skills necessary to understand and evaluate health information, which is the foundation of health literacy. Students learn how to search for, interpret, and critically evaluate health information, which reduces their susceptibility to misinformation (Auld et al., 2020).

Conclusions

Efforts to combat health misinformation combine prevention, like media literacy and machine learning tools, with correction strategies such as debunking false claims, but their success depends on message clarity, trust in healthcare institutions, and community support. Health organisations, partnered with tech platforms and promoted public health education, but persistent misinformation, limited outreach in diverse communities, and declining trust highlight the need for systemic, government-led action to ensure equitable access to accurate and culturally relevant health information.

References

Abdul Raheem, Y. (2023). Unveiling the Significance and Challenges of Integrating Prevention Levels in Healthcare Practice. *Journal of Primary Care & Community Health*, *14*. https://journals.sagepub.com/doi/10.1177/21501319231186500

Auld, M. E., Allen, M. P., Hampton, C., Montes, J. H., Sherry, C., Mickalide, A. D., Logan, R. A., Alvarado-Little, W., & Parson, K. (2020). Health Literacy and Health Education in Schools: Collaboration for Action. *NAM Perspectives*. https://doi.org/10.31478/202007b

Balaj, M., Henson, C. A., Aronsson, A., Aravkin, A., Beck, K., Degail, C., Donadello, L., Eikemo, T. A., Friedman, J., Giouleka, A., Gradeci, I., Hay, S. I., Jensen, M. R., McLaughlin, S. A., Mullany, E. C., O'Connell, E. M., Sripada, K., Stonkute, D., Sorensen, R. J. D., Gakidou, E. (2024). Effects of education on adult mortality: A global systematic review and meta-analysis. *The Lancet Public Health*, *9*(3), 155–165. https://doi.org/10.1016/S2468-2667(23)00306-7

Bobier, C., & Obeid, J. (2025). The US healthcare system facilitates medical conspiracy theories: A call to action. *Ethics, Medicine and Public Health*, *33*. https://doi.org/10.1016/j.jemep.2025.101090

Borges do Nascimento, I. J., Pizarro, A. B., Almeida, J. M., Azzopardi-Muscat, N., Gonçalves, M. A., Björklund, M., & Novillo-Ortiz, D. (2022). Infodemics and health misinformation: A systematic review of reviews. *Bulletin of the World Health Organisation*, *100*(9), 544–561. https://doi.org/10.2471/BLT.21.287654

Centers for Disease Control and Prevention. (2023). *CDC's Gateway to Communication and Social Marketing Practice*. https://www.cdc.gov/healthcommunication/index.html

Gabarron, E., Oyeyemi, S. O., & Wynn, R. (2021). COVID-19-related misinformation on social media: A systematic review. *Bulletin of the World Health Organisation*, *99*(6), 455-463A. https://doi.org/10.2471/BLT.20.276782

Germani, F., Pattison, A. B., & Reinfelde, M. (2022). WHO and digital agencies: How to effectively tackle COVID-19 misinformation online. *BMJ Global Health*, *7*(8). https://doi.org/10.1136/bmjgh-2022-009483

Guyatt, G. (2014). *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice* (3rd ed.). McGraw Hill LLC. https://books.google.pl/books?id=zzcfBQAAQBAJ

Harrington, N. G., & Record, R. A. (2023). *Health Communication: Research and Practice for a Diverse and Changing World*. Taylor & Francis. https://books.google.pl/books?id=JVHGEAAAQBAJ

Harris, K. R. (2024). *Misinformation, Content Moderation, and Epistemology: Protecting Knowledge*. Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9781032636900/misin formation-content-moderation-epistemology-keith-raymond-harris

Kisa, A., & Kisa, S. (2025). Health conspiracy theories: A scoping review of drivers, impacts, and countermeasures. *International Journal for Equity in Health*, *24*(1). https://doi.org/10.1186/s12939-025-02451-0

Lewis, M. (2023). *Philosophy, Biopolitics, and the Virus: The Elision of an Alternative*. Rowman & Littlefield. https://rowman.com/ISBN/9781666923797/Philosophy-Biopolitics-and-the-Virus-The-Elision-of-an-Alternative

Li, J., & Yang, X. (2025). Does exposure necessarily lead to misbelief? A meta-analysis of susceptibility to health misinformation. *Public Understanding of Science*, *34*(2), 222–242. https://doi.org/10.1177/09636625241266150

Nascimbeni, F., & Vosloo, S. (2019). *Digital literacy for children: Exploring definitions and frameworks*. https://doi.org/10.13140/RG.2.2.33394.94407

Office of the Surgeon General (OSG)., O. of the S. G. (2021). We can take action. In *Confronting Health Misinformation: The U.S. Surgeon General's Advisory on Building a Healthy Information Environment*. US Department of Health and Human Services. https://www.ncbi.nlm.nih.gov/books/NBK572168/

Okoro, Y. O., Ayo-Farai, O., Maduka, C. P., Okongwu, C. C., & Sodamade, O. T. (2024). A review of health misinformation on digital platforms: Challenges and countermeasures. *International Journal of Applied Research in Social Sciences*, *6*(1), 23–36. https://doi.org/10.51594/ijarss.v6i1.689

Saver, R. S. (2023). *Physicians Spreading Medical Misinformation: The Uneasy Case for Regulation*. https://doi.org/10.2139/ssrn.4457780

Scherer, L. D., McPhetres, J., Pennycook, G., Kempe, A., Allen, L. A., Knoepke, C. E., Tate, C. E., & Matlock, D. D. (2021). Who is susceptible to online health misinformation? A test of four psychosocial hypotheses. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *40*(4), 274–284. https://doi.org/10.1037/hea0000978

Waszak, P., Łuszczak, E., & Zagożdżon, P. (2024). COVID-19 in Polish-language social media—Misinformation vs government information. *Health Policy and Technology*, *13*(2). https://doi.org/10.1016/j.hlpt.2024.100871

World Health Organisation. (2022). *Toolkit for tackling misinformation on noncommunicable diseases Forum for tackling misinformation on health and NCDs*. https://www.who.int/europe/publications/i/item/WHO-EURO-2022-6260-46025-66542

Yang, Y. T., & Schaffer DeRoo, S. (2022). Disciplining Physicians Who Spread Medical Misinformation. *Journal of Public Health Management and Practice*, *28*(6), 595–598. https://doi.org/10.1097/PHH.0000000000001616

Zhang, S., Zhou, H., & Zhu, Y. (2024). Have we found a solution for health misinformation? A ten-year systematic review of health misinformation literature 2013–2022. *International Journal of Medical Informatics*, *188*.

https://doi.org/10.1016/j.ijmedinf.2024.105478

How can we perceive conspiracy theories in order to diversify existing typologies?

Ioannis Elissaios Paparrigopoulos[1], Dora Katsamori[2], Georgios Petasis[3]
Institute of Informatics and Telecommunications (IIT), NCSR 'Demokritos'
1. ie.paparrigopoulos@iit.demokritos.gr
2. dkatsamori@iit.demokritos.gr
3. petasis@iit.demokritos.gr

Abstract

Our ability to effectively identify and counter misinformation, as demonstrated by our recent experience from the TITAN project (H2020)[2], heavily relies on the existence of standardized typologies and classification schemes of what we often refer to as 'misinformation/disinformation tactics/techniques'. But how did these 'typologies' came to be? How were they organised and, more importantly, can they be considered 'the right tools for the job'? This paper aspires to provide some answers to these questions by focusing on one of the most prominent and extensively researched mis/disinformation tactics/techniques: conspiracy theories. By discussing how conspiracy theories tend to be traditionally conceptualized and studied, we attempt to explore (a) which can be considered as their most problematic characteristics, and, (b) how new research approaches (i.e. computational folkloristics, geospatial analysis, information availability, etc.) can help us better contextualize, classify, and, ultimately, understand them.

Keywords: conspiracy theories, typologies, classification schemes

Introduction

Conspiracy theories have been around for a long time. Yet, a growing number of scholars and policymakers appear convinced that the threat posed by conspiracy theories -and misinformation in general- to democratic institutions, values and beliefs is greater than ever before (Ecker, U., Roozenbeek, J., et al., 2024; Lewandowsky, S., Ecker, U., et al., 2023). Despite evidence which suggests otherwise (Uscinski, J., Enders, A., et al., 2022), it is impossible not to argue that modern-day technological advances, interconnectivity, political uncertainty, etc., have presented us with a

---

[2] TITAN is an AI-powered solution (LLM-powered chatbot) aiming at revolutionizing the way individuals learn to identify and stop the spread of disinformation (https://www.titanthinking.eu/). By emphasizing the identification of manipulative tactics (largely based on the DEPICT model) (Harjani et al., 2022), TITAN aims at helping users 'internalize' questioning techniques and enhance their ability to think critically when confronted with potentially misinformative content.

number of new challenges[3]. What is increasingly evident however, is that calls for interventions to stop the spread of fake news and misinformation have been steadily growing. In today's polarized climate, it may be more urgent than ever to consider how we choose to study conspiracy theories; as 'isolated phenomena', often framed as 'aspects of misinformation' and 'creeping authoritarianism' (Fenster, M., 2024), or as something more complex?

Brief Commentary on the Historical Conceptualization of the Term and its (Un)popularity

In spite of mounting evidence suggesting that a strong inclination in believing conspiracy narratives in crisis situations exists (i.e., the case of Covid-19) (Karađuz, A., Cvjetićanin, T., 2022), conspiracy theories still tend to be widely perceived as incoherent, inherently controversial, systems of thought (Lewandowsky, S., & Cook, J., 2020). The term's genealogy, reflecting Karl R. Popper's positivistic critique (i.e., the adoption of falsifiability as a criterion of identification and refutation of misleading narratives) (Shermer, M., 2022), resulted in enhancing popular belief that conspiracy theories constitute a 'simplistic', 'unscientific', and -though not invariably- 'irrational' way of explaining social phenomena (Butter, M., & Knight, P., 2020; Harris, K., 2018; Dentith, Matthew R. X., 2018).

Failing to capture the popular character of conspiracy theories as an occasionally predominant component of sociopolitical reality, Popper's warranted critique (i.e., it is impossible to sufficiently justify the causal power ascribed to a conspiracy) (Dentith, Matthew R. X., 2018) perceived conspiracy theories as a liminal phenomenon. The conceptualization of the term in the aftermath of WWII as a de facto self-sealing and omniscient way of theorizing[4], arguably bolstered a 'nuanced' view of conspiracy narratives, changing their status from a '*legitimate to* [an] *illegitimate* [form of] *knowledge*' (Thalmann, K., 2014) making it hard to perceive them as 'neither inherently reactionary, nor inherently radical' phenomena (Butter, M., & Knight, P., 2020).

It was Richard Hofstadter, famous for describing conspiracy theorizing as 'a manifestation of irrationality resembling the paranoid ideation' (Butter, M., & Knight, P., 2020; Dentith, Matthew R. X., 2018), who first identified a correlation between the rise of right-wing populism in U.S. politics and conspiracy theories (Fenster, M., 2024). Yet, despite Hofstadter's keen observations[5], his prevailing beliefs on the

---

[3] i.e., the introduction of generative AI models has sparked a debate regarding their potential to amplify online misinformation (Simon, F., Altay, S., et al., 2023; Hate, C., 2024), as well as, in their ability to effectively counter it (Danni, X., et al., 2023; Chen, C., & K., Shu., 2024; Costello, T., Pennycook, G., & Rand, D., 2024).

[4] The correlation between this first conceptualization and totalitarianism might be worthy of further investigation (Thalmann, K., 2014) (i.e. absolutism in the exertion of power and control over society, promotion of rigid worldview(s), systematical development of political propaganda techniques, etc.).

[5] Resembling the assertion that susceptibility to conspiracy theories appears to be proportional to the presence - or absence- of crisis situations (van Prooijen, J.-W., & Douglas, K. M., 2017).

‘pathological’ and ‘manichaeistic’ nature of conspiracy theories (Fenster, M., 2024), appeared to further reinforce popular belief that conspiracy theories constitute irrational ‘aberrations’ from political normality and liberal values (Butter, M., & Knight, P., 2020). We argue that without a systematic, interdisciplinary approach (non-dismissive of the potential impact of these first conceptualizations), contemporary research is in danger of imposing itself severe self-restrictions.

Methodology

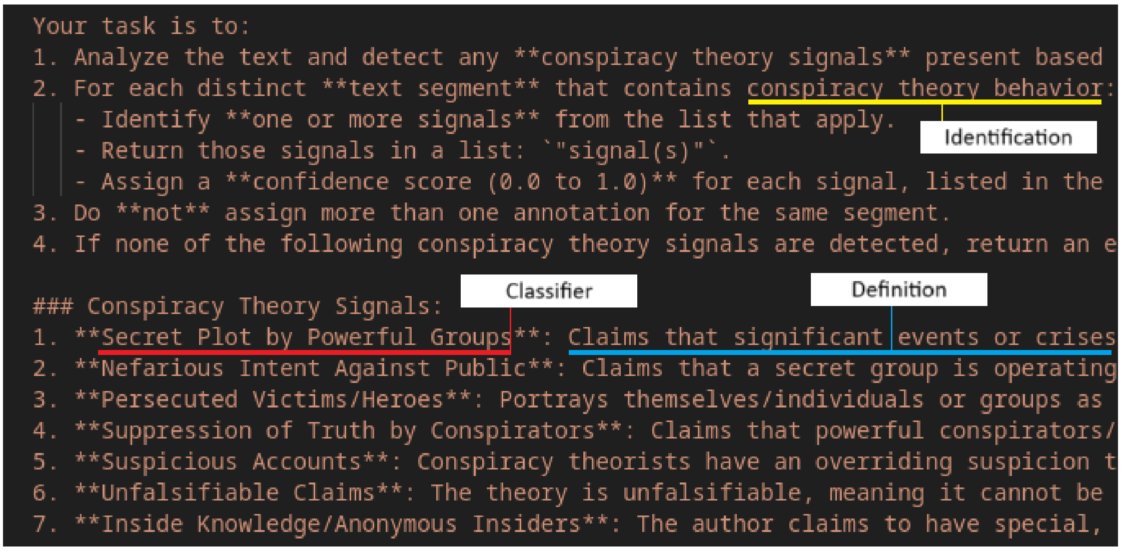
Conspiracy Theories in Misinformation Research

By starting to grasp conspiracy theories’ popular character, their ‘transition’ ‘*from fringe phenomena to central forces* [capable of] *shaping public opinion and political discourse*’ (Stockemer, D., & Bordeleau, J.-N., 2024) was gradually completed, placing conspiracy theories at the heart of contemporary misinformation research. All things considered, we would be remiss not to mention how misinformation research tends to be organised around two fundamental pillars: the (a) *identification* and (b) *refutation* (i.e., reduction of susceptibility) of misleading narratives and false information. Both these research pillars/desiderata, called for an increasingly systematic, formalized approach to conspiracy theories and misinformation in general.

For these and other reasons, the study of conspiracy theories in applied misinformation research tends to increasingly take the form of standardized protocols and guidelines (i.e., *The Conspiracy Theory Handbook* (Lewandowsky, S., & Cook, J., 2020), *The Conspiracy Detection Kit* (Shermer, M., 2022), etc.). Also demonstrated by our experience from TITAN, when theoretical concepts are applied, they inevitably take the form of necessary, ‘selective abstractions’ from theory itself. If we take for example the process of creating classifiers to guide an LLM identify conspiracy theories (Fig.1), we can easily observe how theory tends to be definitively reproduced and applied through otherwise debatable (more or less) criteria, definitions and concepts[6].

---

[6] Technical implementation requires compromise; i.e., computational power will determine identification time, number of classifiers and length of definitions; detection capability will depend on clarity and non-contradictions, etc.



```
Your task is to:
1. Analyze the text and detect any **conspiracy theory signals** present based
2. For each distinct **text segment** that contains conspiracy theory behavior:
    - Identify **one or more signals** from the list that apply.
    - Return those signals in a list: `"signal(s)"`.
    - Assign a **confidence score (0.0 to 1.0)** for each signal, listed in the
3. Do **not** assign more than one annotation for the same segment.
4. If none of the following conspiracy theory signals are detected, return an e

### Conspiracy Theory Signals:
1. **Secret Plot by Powerful Groups**: Claims that significant events or crises
2. **Nefarious Intent Against Public**: Claims that a secret group is operating
3. **Persecuted Victims/Heroes**: Portrays themselves/individuals or groups as
4. **Suppression of Truth by Conspirators**: Claims that powerful conspirators/
5. **Suspicious Accounts**: Conspiracy theorists have an overriding suspicion t
6. **Unfalsifiable Claims**: The theory is unfalsifiable, meaning it cannot be
7. **Inside Knowledge/Anonymous Insiders**: The author claims to have special,
```

Fig.1: Conspiracy Theory Classification (signal detection system) _The TITAN Project (H2020)

Conspiracy Theories Typologies

When we refer to conspiracy theories ‘typologies’, we imply the formalized classification of a finite number of characteristics under which a conspiracy theory will be studied. What we can reasonably infer from a cursory observation of these ‘typologies’[7], is that the study of conspiracy theories has, to this day, yielded limited results. We do not wish to debate whether conspiracy theories tend to be systematically mis-conceptualized[8] or whether the term is inherently problematic (Coady, D., 2023); rather, we point out that the majority of studies appear to reaffirm that (a) contemporary research is extremely compartmentalized, as well as, (b) that conspiracy theories warrant *a form of legitimacy* which they do not currently enjoy/possess. Here, we selectively single out three reasons in order to explain why:

- *Warrantability*: conspiracy theories constitute a legitimate form of theorizing if we consider that: (a) arguments and beliefs are usually formed on the basis of existing values/beliefs, but also information availability and quality (*or lack thereof*) (Dentith, Matthew R. X., 2018) and (b) historically, many conspiracy theories have proven to be true (Coady, D., 2003; Butter, M., & Knight, P., 2020).
- *The problem of ‘conspiratorial ideation’ and ‘conspiracism’*: Both notions appear to be poorly linked to conspiracy theorizing. If not properly organised to refer to ‘*a* [particular] *subset*(s) *of conspiracy theorists*’ (Dentith, Matthew R. X., 2018) with distinctively different characteristics from other forms/types

---

[7] i.e., from definitional schemes, to the identification of important social, psychological motives for believing in conspiracy theories (Douglas, K. M., Sutton, R. M., & Cichocka, A., 2017; Douglas, K. M., Uscinski, J. E., et al., 2019) and the study of the ‘conspiratorial ideation’ (see Bowes, S., et al., 2023) or ‘conspiracism’.

[8] i.e., conspiracy theories can indeed be dangerous and believing in them can be grossly misinformative, an indication of underlying psychopathology, flawed reasoning or atypical behavior, moreover they can be weaponized and exploited, etc.

of conspiracy theorizing/theorists, these notions can potentially generate problematic or even misleading results (Coady, D., 2023)[9].

- *Constructive conspiracism*: Michael Shermer defines 'constructive conspiracism' as an evolutionary reflex of our 'Paleolithic conspiratorial cognition'. Echoing an existential justification to conspiracy theorizing (Douglas, K. M., Uscinski, J. E., et al., 2019), it refers to our inclination 'to err on the side of belief, rather than disbelief' since 'it is better to assume that a conspiracy theory is real when it is not (false positive), instead of believing it is not real when it is (false negative)' (i.e., gain sense of control, self-preservation, etc.) (Shermer, M., 2022).

All of the above suggests that, while there are multiple valid approaches to studying conspiracy theories, each with its own strengths and weaknesses, there is a need for an approach that seeks to restore their lost claim to 'legitimacy'.

Results/Findings

Conspiracy theories as narratives

The idea that conspiracy theories share many similarities with other forms of narratives (i.e., folklore, storytelling, myths, etc.) isn't new. What is relatively new though, is the attempt to systematically study conspiracy theories' *narrative form/structure* and develop new *typologies and methods of classification* based on their respective features/characteristics (topics, motifs, resemblance, repetition, etc.) (Radford, B., 2024). We would argue that by perceiving conspiracy theories as narratives, and their problematic characteristics as important, yet complementary and non-inherent traits (to be studied particularly/on occasion) (Dentith, Matthew R. X., 2018), we can aim at:

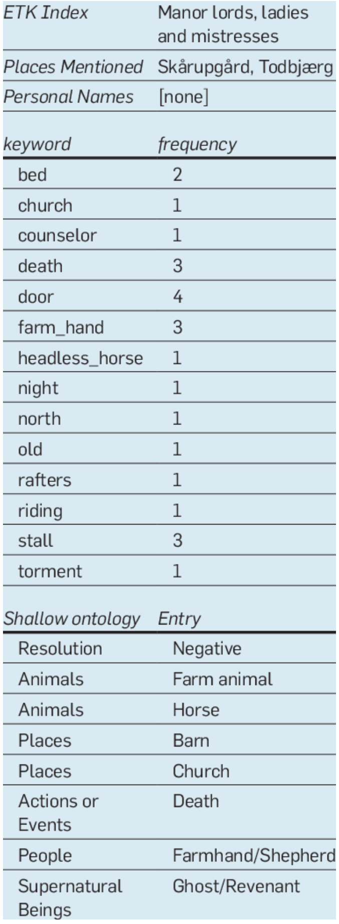
(a) exploring new ways of identification and classification.
(b) offering a more holistic *contextualization* of conspiracy theories.
(c) avoiding counterproductive tendencies which might promote an 'us-vs-them' mentality (Drazkiewicz, Ela., 2022).

We procced by *selectively* presenting some tools which might prove to be useful in our effort to create, expand or diversify existing conspiracy theories typologies and classification schemes. A number of them derives from the field of computational folkloristics as described by James Abello, Peter Broadwell and Timothy Tangherlini in 2012. They refer to (a) 'the development of extensible data structures', (b) 'novel methods for classifying [conspiracy theories'] data' (i.e., linguistic features, contextual attributes, unique characteristics, etc.), and (c) 'algorithmic methods and statistical

---

[9] David Coady's example of 9/11 brilliantly demonstrates how a problematic conceptualization of the 'conspiratorial ideation' can potentially yield misleading results; i.e. you might be wrong in believing that the American secret services were somehow involved in the 9/11 attacks, but you might not be mistaken due to an 'excessive willingness to believe in conspiracies'; 9/11 was a conspiracy (orchestrated by al-Qaeda). You are mistaken because you misidentified the conspiracy/conspirators and ultimately (for a variety of possible reasons) believed the wrong one (Coady, D., 2023).

representations' (i.e., visualizations, maps, networks, etc.) (Abello, J., Broadwell, P., Tangherlini, T., 2012). From their work we single out the notions of 'shallow ontologies' and 'folk story hypergraphs' (i.e., network representations/clustering methods). Cited below, is the original table from *Computational Folkloristics* (2012) (Fig.2); we suggest that a similar approach can be implemented in conspiracy theory research[10].



ETK Index	Manor lords, ladies and mistresses
Places Mentioned	Skårupgård, Todbjærg
Personal Names	[none]

keyword	*frequency*
bed	2
church	1
counselor	1
death	3
door	4
farm_hand	3
headless_horse	1
night	1
north	1
old	1
rafters	1
riding	1
stall	3
torment	1

Shallow ontology	*Entry*
Resolution	Negative
Animals	Farm animal
Animals	Horse
Places	Barn
Places	Church
Actions or Events	Death
People	Farmhand/Shepherd
Supernatural Beings	Ghost/Revenant

Fig.2: Shallow ontology (Abello, J., Broadwell, P., Tangherlini, T., 2012)

Fig.3 (*below*) represents a rough template for conspiracy theories network analysis and clustering (i.e. strong/weak correlations, popularity/frequency, potential qualitative and quantitative characteristics, etc.). Collecting individual attributes of a conspiracy theory and applying a grouping method can facilitate the testing of various hypotheses, correlations, belief patterns, repetitions or frequency of occurrence, depending on the specific focus of a study[11].

---

[10] i.e. with a number of potential changes 'a hierarchical, ontological representation of a [conspiracy theory] corpus based on Actions, Events; [Main Actors]; Places; Resolution; [Supernatural Elements]; Time; Tools; Conveyances, etc.' aiming to offer 'a hierarchical overview of the content of a [conspiracy theory]' can be organised, Abello, J., Broadwell, P., Tangherlini, T., 2012.

[11] i.e., based on the findings of The National Consortium for the Study of Terrorism and Responses to Terrorism (see Fig.4) many of the QAnon perpetrators appeared to have particular religious beliefs suggesting a proximity of the QAnon conspiracy theory with other conspiracy theories bearing similar characteristics, etc. (start forming a hypothetical network/nexus, see Fig.3).

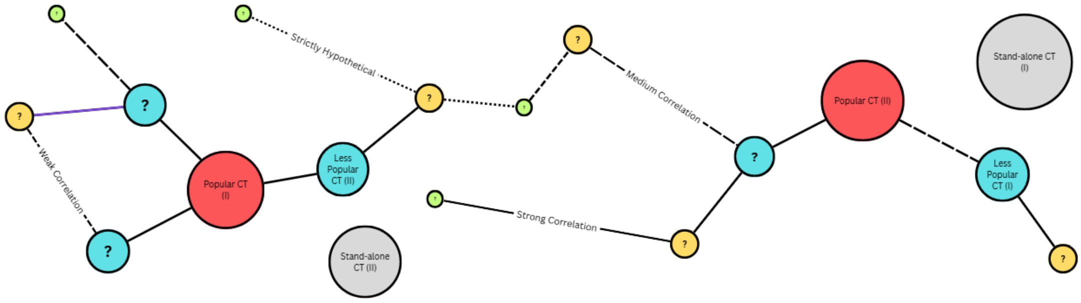
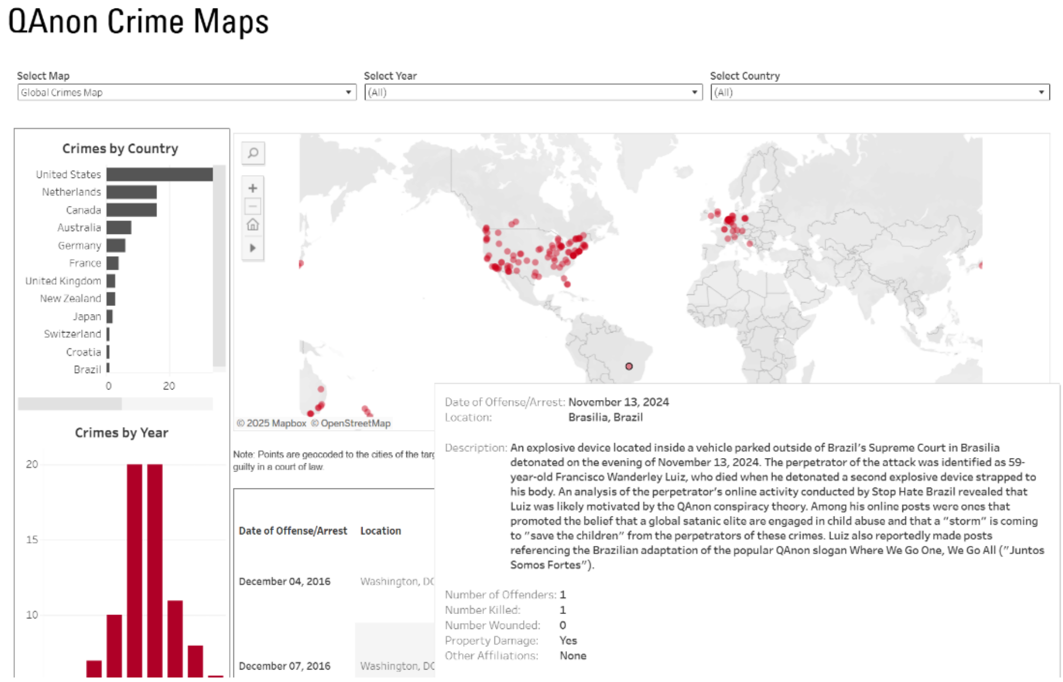


Fig.3: Rough template for conspiracy theories network analysis and clustering

As previously noted, an important layer of analysis—also present in the field of computational folkloristics, though primarily in the form of georeferenced places and locations (Abello, Broadwell, & Tangherlini, 2012)—involves the representation of geographical and statistical data to highlight the real-world impact of conspiracy theories. While the use of maps and geographic visualizations (e.g., scatter plots, heatmaps) is not uncommon, these tools are predominantly employed in fields such as security studies or epidemiology (*figures below*). However, integrating geographical data with other variables—such as sociopolitical, economic, or cultural indicators—could enable for more insightful comparisons. Such an approach would support analyses of information circulation, the national or international 'distribution' of conspiracy beliefs, and ultimately offer a richer contextualization of conspiracy theories and their presumed effect.



Date of Offense/Arrest	Location
December 04, 2016	Washington, DC
December 07, 2016	Washington, DC

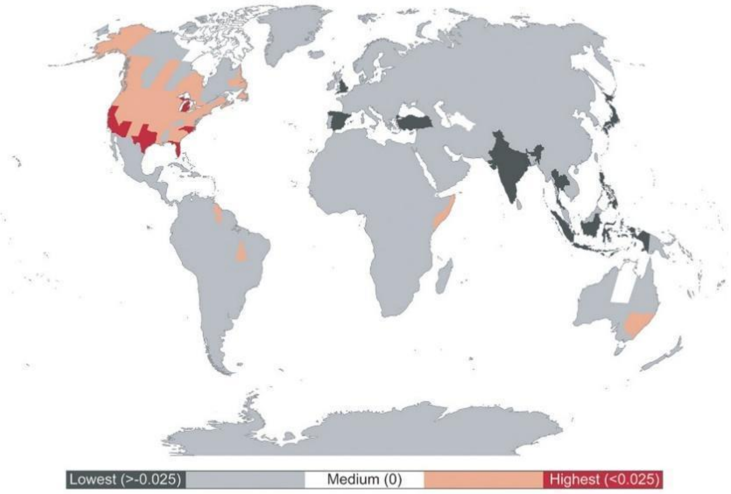
Date of Offense/Arrest: November 13, 2024
Location: Brasilia, Brazil

Description: An explosive device located inside a vehicle parked outside of Brazil's Supreme Court in Brasilia detonated on the evening of November 13, 2024. The perpetrator of the attack was identified as 59-year-old Francisco Wanderley Luiz, who died when he detonated a second explosive device strapped to his body. An analysis of the perpetrator's online activity conducted by Stop Hate Brazil revealed that Luiz was likely motivated by the QAnon conspiracy theory. Among his online posts were ones that promoted the belief that a global satanic elite are engaged in child abuse and that a "storm" is coming to "save the children" from the perpetrators of these crimes. Luiz also reportedly made posts referencing the Brazilian adaptation of the popular QAnon slogan Where We Go One, We Go All ("Juntos Somos Fortes").

Number of Offenders: 1
Number Killed: 1
Number Wounded: 0
Property Damage: Yes
Other Affiliations: None

Fig.4: Interactive QAnon Crime Maps, The National Consortium for the Study of Terrorism and Responses to Terrorism (START), (https://www.start.umd.edu/qanon-crime-maps)

Bioweapon



Chinese or Wuhan and Lab

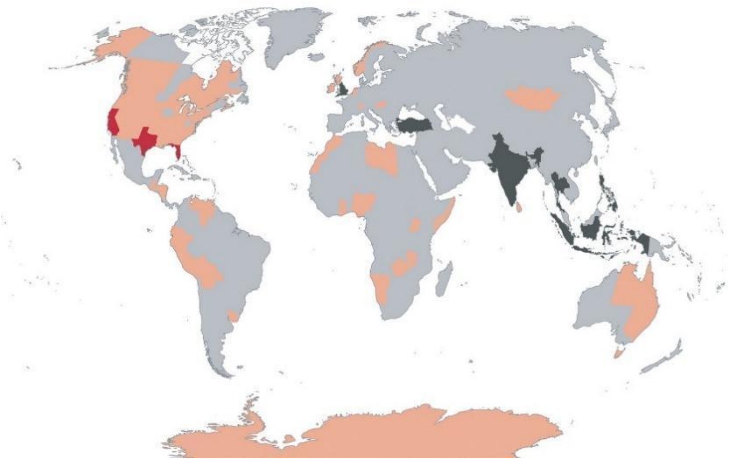


Fig.5: A map of LQs for retweets suggesting COVID-19 is a bioweapon (top) and originated in a lab in China (bottom), (Stephens, M., 2020)

Conclusion

This paper constitutes an effort to (a) explore how conspiracy theories tend to be conceptualized in contemporary misinformation research and, (b) discuss whether by ascribing them with a degree of 'legitimacy' we can aim at addressing several recurring challenges that often appear to accompany their study. As demonstrated by our recent experience, theoretical research should focus on finding new applicable -and innovative- suggestions in order to help guide technical innovation and avoid the reproduction of potentially unproductive and -somewhat- 'dogmatic' forms of knowledge/theory. By presenting key concepts from various academic fields such -as computational folkloristics and geospatial analysis-, we aimed at demonstrating how conspiracy theories can be approached and studied as something more than a de facto aspect of misinformation research. Even more importantly, the introduction of new variables (i.e., sociopolitical, geospatial and historical/cultural data), typologies and classification schemes, is bound to shed some light into an otherwise elusive and under-researched (at least in terms of diversity) scholarly field. Moreover, the existing threat posed to democratic societies and values by dis-/mis-information and, consequently, conspiracy theories, cannot be singularly perceived as an inherent flaw of conspiracy theorizing. It is our belief that more in-depth research (whether small-scale or large-scale) is needed in order to discover new and incitive ways of studying this popular phenomenon.

References

Abello, James & Broadwell, Peter & Tangherlini, Timothy. (2012). Computational folkloristics. *Communications of The ACM - CACM,* 55, 60-70. 10.1145/2209249.2209267.

Bowes, S. M., Costello, T. H., & Tasimi, A. (2023). The conspiratorial mind: A meta-analytic review of motivational and personological correlates. *Psychological Bulletin, 149*(5-6), 259–293. https://doi.org/10.1037/bul0000392.

Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy mentality questionnaire. *Frontiers in Psychology*, *4*, 1-15. https://doi.org/10.3389/fpsyg.2013.00225

Butter, M., & Knight, P. (Eds.). (2020). *Routledge Handbook of Conspiracy Theories* (1st ed.). Routledge. https://doi.org/10.4324/9780429452734.

Center for Countering Digital Hate. (2024). The Double-Edged Sword of AI: How Generative Language Models Like Google Bard and ChatGPT Pose a Threat to Countering Hate and Misinformation Online. *Harvard Data Science Review*, (Special Issue 5). https://doi.org/10.1162/99608f92.be4e28f0.

Chen, C., & Shu, K. (2024). Combating misinformation in the age of LLMs: Opportunities and challenges. *AI Magazine, 45* (3), 354-368. https://doi.org/10.1002/aaai.12188.

Coady, D. (2003). Conspiracy Theories and Official Stories. *International Journal of Applied Philosophy*, *17* (2), 197-209.

Coady, D. (2023). Conspiracy theory as heresy. *Educational Philosophy and Theory*, *55* (7), 756-759.

Cosgrove, T. J., & Murphy, C. P. (2023). Narcissistic susceptibility to conspiracy beliefs exaggerated by education, reduced by cognitive reflection. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1164725

Costello, T., Pennycook, G., & Rand, D. (2024). *Durably reducing conspiracy beliefs through dialogues with AI*. https://doi.org/10.31234/osf.io/xcwdn

Dentith, M. R. X. (2018). The Problem of Conspiracism. *Argumenta, 3* (2), 327-343.

Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The Psychology of Conspiracy Theories. *Current Directions in Psychological Science*, *26*(6), 538-542. https://doi.org/10.1177/0963721417718261

Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding conspiracy theories. *Political Psychology*, *40*(S1), 3–35.

Drazkiewicz, E. (2022). Study conspiracy theories with compassion. *Nature*, *603*(7903), 765–766. https://doi.org/10.1038/d41586-022-00647-x

Ecker, U., Roozenbeek, J., van der Linden, S., Tay, L. Q., Cook, J., Oreskes, N., & Lewandowsky, S. (2024). Misinformation remains a threat to democracy. *Nature*, *630*(8015), 29–32. https://doi.org/10.1038/d41586-024-01587-3.

Enders, A., Klofstad, C., & Uscinski, J. (2024). The relationship between conspiracy theory beliefs and political violence. *Harvard Kennedy School (HKS) Misinformation Review.* https://doi.org/10.37016/mr-2020-163.

Fenster, M. (2024). Studying conspiracy theory after the rise of right-wing populism. In M. Butter, K. Hatzikidi, C. Jeitler, G. Loperfido, & L. Turza (Eds.), *Populism and conspiracy theory*, 344-358. Routledge.

Harjani, T., Roozenbeek, J., Biddlestone, M., van der Linden, S., Stuart, A., Iwahara, M., Piri, B., Xu, R., Goldberg, B., & Graham, M. (2022). *A practical guide to prebunking misinformation.* University of Cambridge. https://prebunking.withgoogle.com/docs/A_Practical_Guide_to_Prebunking_Misinformation.pdf?utm_source=chatgpt.com

Harmer, M. (2022). *Conspiracy: why the rational believe the irrational.* Johns Hopkins University Press.

Harris K. (2018). What's Epistemically Wrong with Conspiracy Theorising? *Royal Institute of Philosophy Supplement*, *84*, 235-257. https://doi.org/10.1017/S1358246118000619

Karađuz, A., Cvjetićanin, T., Ćosić, M., Džebo, S., Krupalija, R., Salkanović, M., & Šehović, N. (2022). *Countering disinformation narratives and mapping conspiracy theories: The case of BiH*. U.G. Zašto ne.

Lewandowsky, S., & Cook, J. (2020). *The Conspiracy Theory Handbook.* Center for Climate Change Communication, George Mason University. https://www.climatechangecommunication.org/all/climate-change-theory-handbook

Lewandowsky, S., Ecker, U. K. H., Cook, J., van der Linden, S., Roozenbeek, J., & Oreskes, N. (2023). Misinformation and the epistemic integrity of democracy. *Current Opinion in Psychology*, *54*. https://doi.org/10.1016/j.copsyc.2023.101711

Radford, B. (2024). Analyzing conspiracies through folklore, epidemiology, and artificial intelligence. *Skeptical Inquirer*, *48*(3). https://skepticalinquirer.org/2024/04/analyzing-conspiracies-through-folklore-epidemiology-and-artificial-intelligence/?utm_source=chatgpt.com

Schöpfer, C., Abatista, A. G. F., Fuhrer, J., & Cova, F. (2023). 'Where there are villains, there will be heroes': Belief in conspiracy theories as an existential tool to fulfill need for meaning. *Personality and Individual Differences*, *200*. https://doi.org/10.1016/j.paid.2022.111900

Simon, F., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-127

Stephens, M. (2020). A geospatial infodemic: Mapping Twitter conspiracy theories of COVID-19. *Dialogues in Human Geography*, *10*(2), 276-281. https://doi.org/10.1177/2043820620935683

Stockemer, D., & Bordeleau, J.-N. (2024). Conspiracy theories and their believers in an era of misinformation. *Harvard Kennedy School Misinformation Review*. https://misinforeview.hks.harvard.edu/article/conspiracy-theories-and-their-believers-in-an-era-of-misinformation/?utm_source=chatgpt.com

Thalmann, K. (2014). “John Birch Blues”: The Problematization of Conspiracy Theory in the Early Cold-War Era. *Current Objectives of Postgraduate American Studies*, *15*(1). https://doi.org/10.5283/COPAS.182.

Uscinski, J. E., Enders, A., Klofstad, C., Seelig, M., Drochon, H., Premaratne, K., et al. (2022). Have beliefs in conspiracy theories increased over time? *PLOS ONE, 17*(1). https://doi.org/10.1371/journal.pone.0270429

van Prooijen, J.-W., & Douglas, K. M. (2017). Conspiracy theories as part of history: The role of societal crisis situations. *Memory Studies*, *10*(3), 323-333. https://doi.org/10.1177/1750698017701615

van Prooijen, J.-W., & van Vugt, M. (2018). Conspiracy Theories: Evolved Functions and Psychological Mechanisms. *Perspectives on Psychological Science*, *13*(6), 770-788. https://doi.org/10.1177/1745691618774270

Xu, D., Fan, S., & Kankanhalli, M. (2023). Combating misinformation in the era of generative AI models. In *Proceedings of the 31st ACM International Conference on Multimedia*, 9291–9298. Association for Computing Machinery. https://doi.org/10.1145/3581783.3612704

Acknowledgements

TITAN has received funding from the EU Horizon Europe research and innovation programme under grant agreement No.101070658, and by UK Research and innovation under the UK governments Horizon funding guarantee grant numbers 10040483 and 10055990.

This paper expresses the views and opinions of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.” of the TITAN project only.

Multimodal Interpretation of a Mobbing Paradigm in the School Workplace

Dr. Georgios Elias Potamias[1], Dr. Paraskevi Kanari[2]
Hellenic Open University
1. potamias.georgios@ac.eap.gr
2. kanari.paraskefi@ac.eap.gr

Abstract

This paper presents an exploratory multimodal analysis framework tool of a real-world mobbing case within an international school environment. Drawing from a real-life case in an international school, it explores how verbal, non-verbal, and contextual elements contribute to subtle forms of workplace exclusion and harassment. The methodology applies social semiotics and critical discourse analysis to emails, evaluations, and interactions. Through experimental research, a variety of communicative artifacts (emails, evaluations, meetings) are examined using a multimodal analysis framework aimed at identifying and interpreting implicit mobbing discourse. By triangulating verbal, non-verbal, and contextual data, the study seeks to detect early signs of workplace harassment and expose the dynamic processes through disinformation from which mobbing unfolds. The study identifies five recurring stages of mobbing: social exclusion, stigmatization, critical incident, adjudication, and elimination. Findings confirm that multimodal analysis can expose covert discrimination masked as professional procedure, and that early intervention is critical to prevent long-term psychological harm. Early findings indicate that multimodal tools can effectively trace mobbing stages—social exclusion, stigmatization, and elimination—while revealing the covert linguistic and behavioral patterns that enable psychological aggression. This approach highlights the crucial need for early detection and interdisciplinary intervention in school workplaces. By combining theory with applied examples, this paper underscores the necessity of incorporating multimodal tools in organisational policy and training to detect implicit forms of aggression through disinformation in the workplace.

Keywords: mobbing, multimodal discourse, school workplace, disinformation, discrimination.

Introduction

Mobbing, a severe yet often invisible form of psychological harassment, has become increasingly prevalent in modern workplaces—including schools. Unlike direct bullying, mobbing is often systemic, subtle, and conducted by a group rather than a single individual (Chappell & Martino, 2006).

Mobbing, a complex and covert form of psychological harassment in workplaces, refers to systematic and prolonged abuse by a group of individuals against a peer, subordinate, or superior. It is characterized by hostile, humiliating, and often non-physical behaviors aimed at expelling or discrediting the target. In school workplaces - especially international or multicultural environments - mobbing may be difficult to detect as it is frequently disguised under professional formalities and institutional discourse (Bouboucheropoulos, 2022).

While bullying typically suggests a dominant actor attacking a weaker individual, mobbing represents collective, subtle aggression by colleagues or teams. In school contexts, especially those operating under principles of political correctness, mobbing becomes fluid, implicit, and often normalized (Alipranti, 2010). Victims often find themselves isolated, undermined, and subtly excluded without overt acts of violence or aggression.

The present study, part of a broader post-doctoral research that aims at identifying the intrinsic characteristics of a democratic citizenship focused pedagogy in Europe, explores how multimodal analysis can help interpret this phenomenon of mobbing through the lens of educational institutional settings, where hierarchical dynamics and politically correct discourse mask underlying aggression.

The early-stage research that is presented in this paper focuses on a real-life case study in an international school, analyzing materials such as emails, evaluations, and informal communications to trace the stages of mobbing and decode implicit messages that contribute to the marginalization of targeted individuals. By applying multimodal analysis techniques, the study seeks to offer a novel lens for detecting and ultimately mitigating mobbing in its early stages.

Greek scholars such as Bouboucheropoulos (2022) and Alipranti (2010) have emphasized the implicit and normalized nature of mobbing in Greece. Bouboucheropoulos explores legal and moral accountability for psychological harassment, while Alipranti investigates gender and institutional silence in workplace violence.

Internationally**,** Leymann (1996) was among the first to conceptualize mobbing as a syndrome. He listed over 45 behaviors linked to psychological abuse in the workplace and stressed the role of organisational inertia in enabling such behavior. Einarsen et al. (2011) further developed the theory, connecting mobbing to organisational culture, leadership deficits, and power vacuums. There is also is great value in trauma-informed education as a complementary lens on tackling mobbing behaviors in the school context (Texeira, 2012).

Mobbing generally develops in the following predictable stages:

1. Social exclusion
2. Harassment and stigmatization
3. Critical incident
4. Adjudication.

5. Elimination.

This study validates these frameworks by demonstrating how linguistic, institutional, and non-verbal signs collaborate in the early stages of mobbing. It also highlights the importance of early intervention tools for detecting vague or abusive language before psychological harm escalates.

A notable dimension of the mobbing process observed in this study involves the strategic use of disinformation. This includes misleading statements, selective truth-telling, or emotionally charged claims that cannot be easily verified. In the school workplace, such tactics are often couched in vague phrases like 'concerns have been raised' or 'we've noticed a pattern', without offering concrete evidence. These rhetorical strategies allow aggressors to maintain plausible deniability while creating a hostile environment for the targeted individual (Wardle & Derakhshan, 2017).

Disinformation often escalates during Stage 2 (Stigmatization) and Stage 3 (Critical incident), serving to isolate the victim through the spread of narratives that cast doubt on their professionalism or behavior. For instance, evaluations may reference 'unverified reports' or imply lack of collaboration based on subjective impressions rather than documented performance. The use of emotional language and repetition—such as phrases like 'lack of team spirit' or 'concerning tone'—reinforces these impressions (Fallis, 2015; Marwick & Lewis, 2017).

Methodology

This is an ongoing experimental research project that forms part of a broader postdoctoral investigation into fluid and implicit discriminatory discourse in public education. The methodology is grounded in multimodal discourse analysis (MDA), which integrates linguistic, visual, auditory, and contextual sign analysis through multimodal analysis program tools (atlas ti, chat lp) in order to detect and study meaning-making practices (Cheng et al., 2019) that consist of a mobbing practice especially in the early stages of the mobbing incident.

The analysis uses a triangulation framework involving:

- Video recordings (e.g., phone call behavior)
- Audio analysis (tone, pauses, hesitations)
- Chat logs and emails: both official and unofficial communications.
- Official and unofficial textual documents (evaluations, memos)
- Nonverbal cues observed in face-to-face meetings according to the victim's testimony.
- Evaluation reports: formal documents evaluating the teacher's performance.
- Audio recordings: from meetings and informal feedback sessions.
- Written complaints and meeting minutes.
- Non-verbal communication: inferred through interviews and documented behavior according to the victim's testimony.

Participants and sources were anonymized to preserve privacy. For reasons of personal data safety no screenshots of artifacts (emails, notes etc.) could be included for illustrative purposes, even anonymized, despite being asked for permission to do so. Special focus was given to peer-to-peer and hierarchical relation within the school, particularly between administrators and newly appointed, high-performing employee.

In the majority of observed cases, the victims were male, aged early 40s, and newly hired with strong academic and teaching credentials. Their presence often threatened existing power dynamics. This is the profile of the person whose material is presented is presented and analyzed in this paper. The targeted employee was male, in his early 40s, newly hired, and academically distinguished. Early emails suggested social distance and contained ambiguous language such as "observations," "concerns," and "discussions"—without specifying concrete issues. Gradually, this escalated to formal appraisals, culminating in exclusion from major school initiatives

Stages 1 and 2 often included false accusations of misconduct (e.g., "inappropriate comments") that were never verified. The direction used evaluations and meetings to shift the narrative from misconduct to incompetence.

Examples of multimodal analysis in early mobbing stages

In this chapter the multimodal theory is used as an insight tool to detect subtle mobbing discourse. Multimodal analysis allowed for the detection of subtle indicators of exclusion and harassment, often hidden within emails, meeting requests, or evaluation reports.

Example 1: Email suggesting social exclusion

"As you could have seen in your Outlook calendar, the vice principal will come to visit you and your class next week on Wednesday at 10.50 am for mathematics. This is not an evaluation but a class visit and you do not have to prepare something special."

Interpretation**:**

Despite assurances that this is not an evaluation, the vague phrasing and short notice suggest surveillance. The lack of transparency fosters anxiety and prepares the ground for future malevolent criticism.

Example 2: Unofficial questioning of professionalism

"Excuse my mail but can we talk for a little bit because I am a bit confused about the nature of our conversation with you and... I am new to the school (request for CV that has already been sent, evaluation of teaching of Mathematics, etc.)."

Interpretation**:**

Repeated requests for already submitted documentation and vague concerns about communication indicate disinformation and micromanagement that aim at confusing the victim. These linguistic acts contribute to the stigmatization phase.

Example 3: Negative feedback disguised as constructive
"Thank you for the constructive feedback concerning my teaching: 1. mixing books, 2. use of more optical material, 3. classroom decoration. I am already implementing them and hereby attach the model scenario."
Interpretation:
The use of supposedly "constructive" criticism masks an evaluative tone, implying inadequacy in basic teaching strategies. These recurrent suggestions are based on disinformation and can undermine the teacher's confidence and competence.

Example 4: Ambiguous meeting requests
"Do we have a Teams meeting tomorrow at 1.30? Is it online or in situ? I'm just asking to be sure."
Interpretation:
Such examples demonstrate a breakdown in communication norms, indicating that the target is kept out of the loop—an essential mechanism of disinformation and essentially consequent exclusion.

Example 5: Sudden changes in scheduling
"We would like to invite you to a meeting Monday morning at 8.15 in my office to discuss matters. You will be replaced for the morning period."
Interpretation:
While seemingly neutral, this message fits a broader pattern of targeted disruption, but when seen in sequence with others, reveals a pattern of control and pressure through unexpected program modifications and deliberate disinformation.

Results/Findings

Mobbing develops as a multistage process. In the case presented in this study mobbing develops in predictable stages (see Table 1) :
The mobbing process unfolded in five distinct stages:

1. Social exclusion**:** vague oral complaints and unexplained distancing in communication.
2. Harassment and stigmatization**:** written criticism without evidence, often using subjective or biased language.
3. Critical incident**:** administrative actions taken to formalize accusations or observations.
4. Adjudication**:** suggestion of disciplinary pathways, legal threats, or performance plans.
5. Elimination**:** exclusion from team meetings, projects, and institutional planning, often leading to resignation.

This sequence was evident in this case study (see Table 1). For example, the victim received emails suggesting informal "visits" that were in effect evaluations without prior notice—intended to unsettle and monitor. A detailed examination of 30 emails and 25 texts revealed recurring patterns such as last-minute meetings, pressure through unofficial evaluations, and persistent tone policing.

Linguistic and behavioral patterns were key signs of obvious mobbing patterns. Multimodal analysis identified key lexical markers such as “unprofessional,” “inappropriate behavior,” and “lack of alignment with expectations.” These signs-units- markers of mobbing behavior were not backed by evidence but repeated in various formats (oral discussions, emails, meetings), contributing to a narrative of incompetence.

Nonverbal cues—such as unannounced meetings, sudden program changes, and last-minute pressure—also played a significant role in the mobbing process.
Table 1 shows the frequency of texts associated with each stage:

Mobbing Stage	**Material Collected**
1. Social exclusion	30 emails with vague, negative language
2. Harassment/Stigmatization	25 texts (formal/informal evaluations)
3. Critical incident	7 negative evaluations, action plans
4. Adjudication/Elimination	Limited data due to early intervention

Table 1: Frequency of Material by Mobbing Stage

Discussion/Conclusion

Multimodal analysis is not just an academic tool—it becomes a practical means of empowerment in workspaces increasingly characterized by ambiguity, implicit power games, and performative inclusion. Mobbing remains an under-recognized phenomenon in educational institutions. This study shows the importance of early detection through multimodal analysis, which may serve as both preventive and diagnostic tool. The study highlights how multimodal tools can uncover the implicit mechanisms of mobbing, especially when traditional HR mechanisms fail to address fluid and covert discrimination. There is indeed a great applied value in the use of multimodal tools in school HR/administration. In educational settings, where hierarchy, "team spirit," and politically correct behavior often camouflage real tensions, such tools are essential for early intervention.

The theoretical framework of the research relies heavily on Kress and van Leeuwen’s (2001) theory of multimodality, which posits that communication is not solely verbal but comprises a variety of semiotic modes. In a mobbing context, this includes tone of voice, spatial arrangements (e.g., seating in meetings), timing of messages, and formatting of evaluations. Fairclough’s critical discourse analysis (1992) provided tools for understanding how discourse maintains institutional power. For example, language used in evaluations and emails often reflected the ideology of the administration, perpetuating dominant narratives about “professionalism” while subtly excluding dissenting individuals. Last but not least, there is great value in trauma-informed education as a complementary lens on tackling mobbing behaviors in the school context (Texeira, 2012).

Multimodal analysis offers an innovative approach to understanding workplace mobbing. In school workplaces, politically correct discourse often camouflages aggressive behavior. Combining linguistic, visual, and situational data can assist in early identification of mobbing and allow timely intervention. Theoretical contributions from social semiotics, interactional sociolinguistics, and psychological trauma studies underpin this analysis (Cheng et al., 2019; Texeira, 2012; Rosen & Shoenberger, 2021). By applying multimodal triangulation—cross-referencing audio, visual, and written data—researchers can detect hostile patterns early enough to protect victims. For instance, emails suggesting “friendly feedback” may carry an undertone of criticism when viewed in combination with simultaneous exclusion from team meetings or subtle body language during staff gatherings.

Mobbing in school workplaces remains an under-recognized and under-researched issue. This study shows how multimodal analysis can detect implicit discrimination through disinformation by triangulating linguistic, non-verbal, and behavioral indicators. By highlighting the fluidity and adaptability of mobbing behaviors, particularly under the guise of professionalism, this research contributes to a more nuanced understanding of psychological violence in education. Institutions should adopt preventive strategies, including anonymous feedback channels, training on discourse ethics, and third-party evaluation tools to safeguard employee wellbeing.

References

Alipranti, L. (2010). *Gender, work and discrimination in Greece*. EKKE.

Bouboucheropoulos, P. (2022). *Mobbing: Ευθύνη λόγω ηθικής παρενόχλησης στην εργασία*. Σάκκουλας.

Chappell, D., & Di Martino, V. (2006). *Violence at Work*. International Labour Office.

Cheng, L., Li, J., Silva, Y. N., Hall, D. L., & Liu, H. (2019). XBully: Cyberbullying detection within a multi-modal context. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, (339–347).

Einarsen, S., Hoel, H., Zapf, D., & Cooper, C. (2011). *Bullying and Harassment in the Workplace: Developments in Theory, Research, and Practice*. CRC Press.

Fairclough, N. (1992). *Discourse and Social Change*. Polity Press.

Fallis, D. (2015). What is disinformation?. *Library Trends*, *63*(3), (401–426).

Kress, G., & van Leeuwen, T. (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication*. Arnold.

Leymann, H. (1996). The content and development of mobbing at work. *European Journal of Work and Organisational Psychology*, *5*(2), (165–184).

Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data & Society Research Institute. https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf?utm_source=chatgpt.com

Rosen, N. L., & Shoenberger, N. A. (2021). “Words Speak Louder than Actions”: The connection between gendered language and bullying behavior. *Open Journal of Social Sciences*, *9*(3), (197–214).

Texeira, M. (2012). Multimodality and bullying: A sociolinguistic analysis. In *Human Rights and Violent Behaviour: The Social and Educational Perspective Conference*, (229–243). UNESCO.

Tomkins, S., Getoor, L., Chen, Y., & Zhang, Y. (2018). A socio-linguistic model for cyberbullying detection. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, (53–60).

Verma, K., Milosevic, T., Cortis, K., & Davis, B. (2022). Benchmarking language models for cyberbullying identification and classification from social media texts. In *Proceedings of the First Workshop on Language Technology and Resources for a Fair, Inclusive, and Safe Society within the 13th Language Resources and Evaluation Conference*, (26–31). European Language Resources Association. https://aclanthology.org/2022.lateraisse-1.4/

Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Council of Europe. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html

Are they Yielding Results? Examining the Relevance of Fact-Checking Training to Information Verification Competence of Nigerian Journalism Students

Ibrahim, Saheed Bidemi[1], Taiwo, Rukayat Abdullahi[2]

1. Birmingham City University, Birmingham, UK, saheed.ibrahim@mail.bcu.ac.uk

2. University of Salford, Salford, UK, rukayatabisola82@gmail.com

Abstract

The pervasiveness of misinformation and disinformation highlights the urgent need to integrate fact-checking training into journalism education. In response, independent fact-checking organisations have initiated training programmes targeting journalism students in Nigeria. This study investigates the level of exposure to such training, evaluates students' fact-checking competencies, and identifies the barriers to their fact-checking skills development. A mixed-methods approach was adopted, combining surveys of 271 journalism students from three universities in Southwest Nigeria with interviews involving programme coordinators. Findings reveal that, despite some exposure to fact-checking training, students' competence remains low. Fact-checking is poorly integrated into the curriculum, and instructors often lack the expertise to effectively teach it. Other constraints include limited awareness of fact-checking tools, few practical opportunities, and insufficient institutional support. These challenges undermine efforts to equip future journalists with the skills to combat false information. The study concludes that current journalism education structures are inadequate for preparing students to verify content effectively. It recommends curriculum reforms, capacity-building for educators, and stronger partnerships with fact-checking organisations to ensure journalism graduates are better positioned to uphold media credibility and tackle information disorder.

Keywords: digital literacy, fact-checking, information verification, journalism education, misinformation

Introduction

As the media landscape continues to shift under the weight of digital disruption and the spread of disinformation, the call for stronger media literacy has become more urgent than ever. Fact-checking has emerged globally as one of the most practical responses to this challenge, especially among young people in schools and universities. For journalism students, who are expected to uphold truth and credibility in the media space, verification skills are not optional, but essential (Daniel, 2018; Raji, 2020; Pew Research Centre, 2019; Wineburg & McGrew, 2017). Scholars like Schreurs and Vandenbosch (2021) and Li et al. (2023) have consistently argued for embedding media literacy in school curricula as a tool to counter misinformation. Potter's (2004) Media Literacy Theory, rooted in the cognitive development approach, supports this position. It places emphasis on continuous learning and critical thinking as essential ingredients for building fact-checking capacity in students.

In Nigerian, however, journalism students struggle to acquire digital media and fact-checking skills due to clear gaps in curriculum and a shortage of trained lecturers (Ilesanmi, 2021). These weaknesses have contributed to a competence gap that leaves students vulnerable to misinformation. Recent studies show that only 20.6% of journalism students use fact-checking tools, and 54.8% are unaware of such platforms (Adjin-Tettey & Amenaghawon, 2024).

To respond to this, fact-checking organisations like DUBAWA and Africa Check have stepped in with fellowships, campus-based training, and media literacy workshops (Folarin, 2020; Ilesanmi, 2021). These interventions equip students with basic skills in critical thinking, verification, and the responsible use of digital platforms. Some of these organisations also award grants and recognise outstanding participants to encourage engagement (Amadu, 2024).

Yet, despite the efforts and relevance of these programmes, there remains limited research on how these initiatives are actually shaping students' skills. Much of the existing literature has focused on journalists, educators, or the operations of fact-checking bodies rather than student skill development (Adekunmisi et al., 2022; Cheruiyot & Ferrer-Conill, 2018; Wasserman, 2020; Wogu et al., 2018).

This study, therefore, seeks to bridge that gap by assessing the relevance of fact-checking works by independent fact-checkers to the skill development of journalism students in Nigeria. It explores their level of exposure, evaluates skill acquisition, and identifies the barriers impeding to their fact-checking skills development. This research aims to inform strategies for enhancing journalism education and promoting effective digital skills necessary for countering misinformation in a rapidly changing media landscape. The specific objectives are to:

i. Identify the level of exposure to fact-checking training among Nigerian journalism students.
ii. Assess the current level of fact-checking competence among Nigerian journalism students.
iii. Identify the challenges to fact-checking skills development among Nigerian journalism students.

Methodology

This study adopted a mixed-methods approach, descriptive survey and interviews to assess the exposure, competence level, and influencing factors in the development of fact-checking skills among journalism students. An online questionnaire was administered, and responses were collected from 271 journalism students in Southwest Nigeria. Respondents had attended at least one fact-checking training in the last three years. A reliability coefficient of 0.83 was recorded for the questionnaire, confirming its internal consistency (Taber, 2018).

To complement the survey findings, semi-structured interviews were conducted with three course coordinators or Heads of Department in three randomly selected universities among those offering journalism as a course in the region. These interviews provided contextual insights and helped clarify emerging issues from the survey data. In addition, we obtained the course outlines of the selected schools to examine how fact-checking content is integrated into journalism curricula. This triangulation approach allowed for a more comprehensive understanding of the subject (Saunders et al., 2023). Survey data was analysed using the Statistical Package for Social Sciences (SPSS v.23). Responses gathered through the interviews were analysed thematically using NVivo software. Throughout the research process, ethical considerations were prioritised, including informed consent, privacy, anonymity, and the confidentiality of all participants.

Results

Exposure to fact-checking training

The first objective of this study was to examine the students' exposure to fact-checking training. Analysis in Figure 1 showed that their overall exposure to fact-checking training is *moderate* among the majority (63.5%) of the respondents. Just 5.5% experienced *very high* exposure to fact-checking training. Overall, the results indicate a clear gap in depth and consistency of training. The limited number with high or very high exposure suggests that current efforts may not be sufficient to ensure mastery or retention of fact-checking skills. We found that only 19.9% of students said fact-checking classes were always included in their coursework. Even though most of the students indicated that their lecturers do emphasise the relevance of fact-checking skills, these results imply that there is a fragmented approach to integrating fact-checking into journalism curricula in the universities.

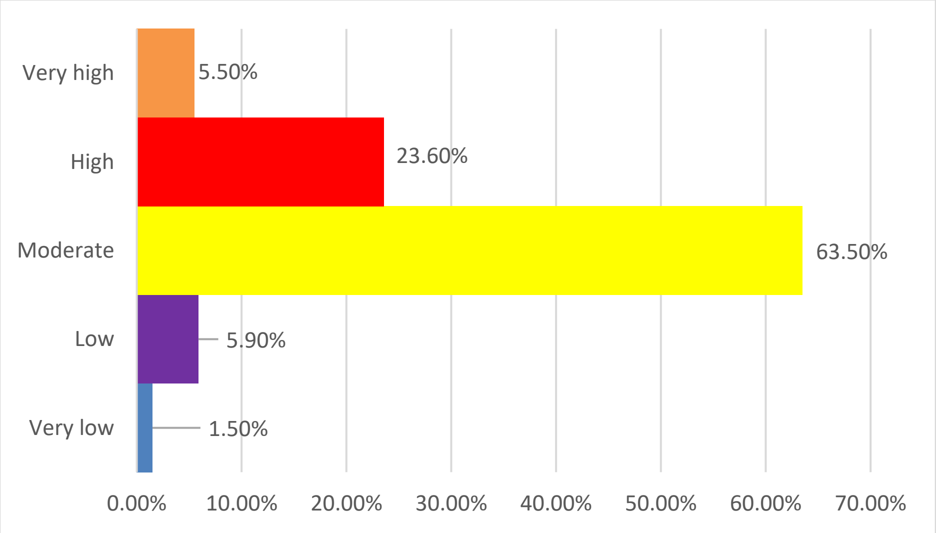


Figure 1: Overall exposure to fact-checking training

Responses from the interviews with journalism instructors from the three institutions indicate a significant lack of formal fact-checking training for both lecturers and students. Lecturer A mentioned that although they (lecturers) personally explored fact-checking and applied it in informal settings, their universities did not organise any structured training for them. Similarly, Lecturer B acknowledged the absence of any formal training, either nationally or internationally, in their experience. The third lecturer also shared the same experience as his school journalism curriculum does not specifically have any module on fact-checking. One of the lecturers stated thus:

> As of now, we have not been able to have such in our curriculum to be able to teach as a course but in some courses like news writing and reporting, we tell them how to fact-check information.

In addition to the data collected, we reviewed the available courses in the selected departments. We found that there was no course specifically dedicated to fact-checking. The findings revealed that exposure to fact-checking training among journalism students is inconsistent and often relies on external organisations or informal inclusion in existing courses rather than structured, comprehensive instruction.

Current level of fact-checking skills

The second objective of this study was to determine the level of fact-checking competence among the students, given their exposure to fact-checking workshops. A self-rating of fact-checking skills among the students revealed that they rated themselves highest in *image verification* and the use of *Google Maps*, but lowest in using tools like *InVid*, *Noise Analysis*, and *Deepfake Detection*. The students also rated themselves low in making Freedom of Information (FOI) requests, which does not require any digital tool but competence in writing.

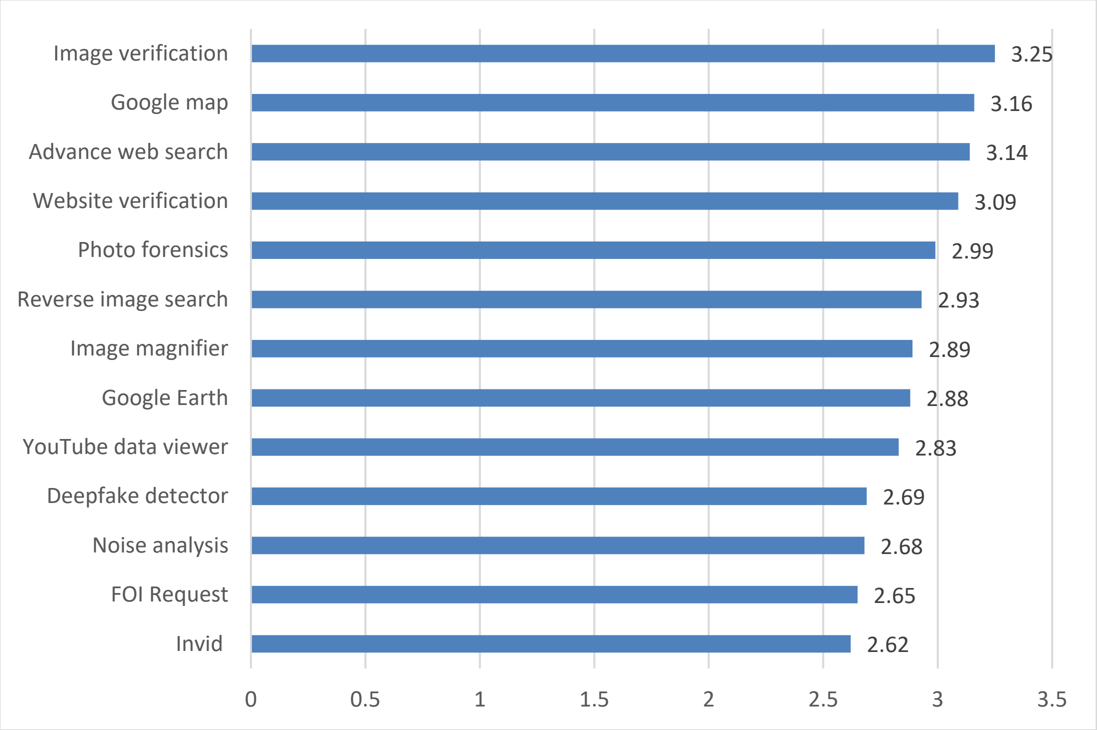


Figure 2: Fact-checking competence among the students

Decision Rule: If mean is 1.0 - 1.49 = Very Low Competence; 1.5 - 2.49 = Low Competence; 2.5 - 3.49 = Moderate Competence; 3.5 - 4.49 = High Competence; 4.5- 5.0 = Very High Competence

We used a 5-point Likert scale to measure the mean score of the items. With an average mean of $\bar{x}$= 2.91, the fact-checking competence of the students was rated moderate. This result might be attributed to their overall moderate fact-checking training exposure. Overall, the students are more comfortable with basic or familiar tools and platforms. Advanced verification tools essential in a digitally deceptive landscape are not well understood or utilised by the students. This suggests the need for practical, tool-specific training sessions in journalism education.

Responses of journalism instructors also corroborate the survey findings. Given that the students were not exposed to fact-checking from the curriculum, their level of competence is expectedly average. The lecturers mentioned that students working with independent fact-checking organisations or news outlets with fact-checking desks have better fact-checking skills than those who received the training without practising. Lecturer B said:

> "But I know some of them truly practice journalism and not just read journalism. They are very proficient by belonging to organisations or writing clubs on campus or freelancing with media outside the campus walls."

The result that the students, who were exposed to fact-checking workshops, possess an insignificant level of fact-checking competence underscores the critical need to enhance educational programs to elevate their skills from low/moderate to higher levels of proficiency.

Challenges to Fact-checking Skills Development

We identified the challenges to the development of fact-checking skills among the sampled journalism students (figure 3). The challenges were measured using a 5-point Likert scale and their mean values were used to evaluate their level of significance. The top three ranked challenges were a *lack of fact-checking skills among their lecturers* (60.9% agreement)*, a lack of integration of fact-checking in the curriculum* (55.3% agreement) *and a lack of support* (50.2% agreement) when the students seek to fact-check information.

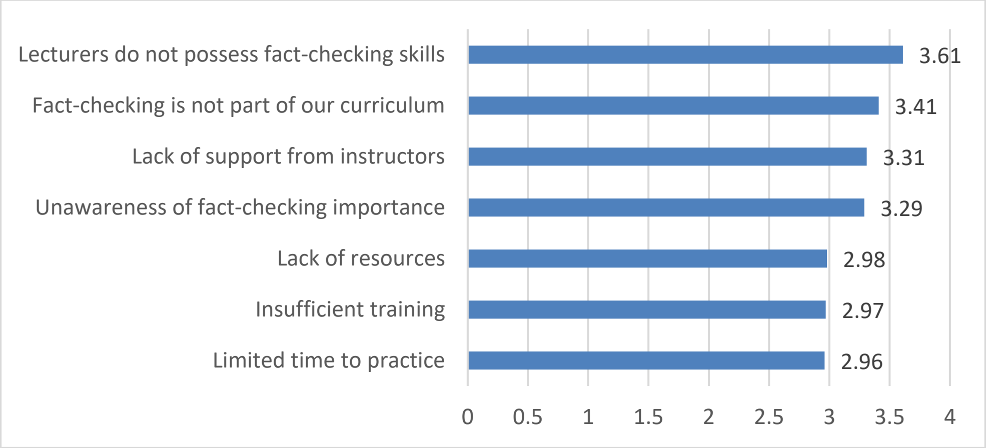


Figure 3: Decision Rule: If mean is 1.0 - 1.49 = Very Low; 1.5 - 2.49 = Low; 2.5 - 3.49 = Moderate; 3.5 - 4.49 = High; 4.5- 5.0 = Very High

Responses from the interviews corroborated the survey results regarding the challenges to the development of fact-checking skills. Significantly, the lecturers admitted that they have not had any structured training on fact-checking and only rely on their personal efforts. They also confirmed that fact-checking has not been integrated into the journalism curriculum in their institutions. One of the instructors said:

> *The only challenge, I think, is that we don't have any course addressing fact-checking, we just incorporate them in a little discussion of what we teach in class. But there is no specific course teaching that and that is a very big challenge to them.*

While adding that there are inadequate resources, another lecturer added, "*The lack of awareness is one of the challenges. And then for the fact that the curriculum is not able to incorporate something like that*." Responses from the lecturers suggest that their limited fact-checking skills stem from a lack of developmental training opportunities provided by their institutions. Consequently, they are unable to confidently teach or integrate fact-checking into the journalism curriculum. This reinforces the gap between what is taught and what is needed in today’s media landscape.

Given these results, we can infer the development of fact-checking competence of the sampled journalism students is largely inhibited by lack of fact-checking skills among their lecturers. Lack of integration of fact-checking into journalism curricula, insufficient support from the lecturers, low awareness about fact-checking relevance, limited training and resources also hamper the development of fact-checking skills among the students.

Conclusion

Our findings have shown that despite being exposed to fact-checking workshops, the overall exposure to fact-checking training among the students is moderate. This can be attributed to the students only taking fact-checking classes through workshops organised by fact-checking organisations. This result supports scholars emphasizing the importance of digital literacy and critical thinking in journalism education. Adjin-Tettey & Amenaghawon (2024) and López-Meri, Doménech-Fabregat, and Marcos-García (2024) recommended that fact-checking and digital literacy should be integrated into the school curriculum, rather than what is taught at workshops. Potter's media literacy theory emphasises the need for critical thinking and continuous learning for students to distinguish between falsehoods and genuine information (Schreurs & Vandenbosch, 2021) but this study has shown insignificant preparation of journalism students to combat dis/misinformation, which has become a major issue in journalism practice.

We found a high competence in using basic information verification tools but a high deficiency in using advanced verification technologies, essential in combating sophisticated misinformation among the participants. Our finding aligns with the results of Wogu et al. (2019), Adjin-Tettey and Amenaghawon (2024) and López-Meri et al (2024) that journalism students lack advanced fact-checking skills. The consequences of insufficient digital literacy among journalism students are far-reaching, affecting their participation in digital learning environments, overall academic performance, and broader societal engagement. Therefore, improving fact-checking skills in journalism education is not just a pedagogical necessity but also a socio-economic imperative.

Lastly, we found that despite being exposed to fact-checking workshops, the students did not get the required support from their institutions, especially from their tutors. The top three challenges to the development of fact-checking skills among the students were a lack of fact-checking skills among their lecturers, a lack of integration of fact-checking in the curriculum and a lack of support. This result corroborated the finding of Ilesanmi (2021) that journalism lecturers have low fact-checking skills. Lack of resources, limited time to practice, insufficient training and unawareness of fact-checking relevance were identified, supporting why students rely on Google and personal contacts to verify information (Donovan & Rapp, 2020). This result also corroborated the conclusion of Wogu et al (2019) that a lack of appropriate funding, resources and technology skills among lecturers predicted the deficiency in media literacy skills among Nigerian Mass Communication students. The result supports the call for structured training integrated into the school curriculum (Wasserman et al., 2022)

Given these findings, we conclude that current journalism education structures in Nigeria are inadequate for preparing students to fact-check information effectively. Although the workshops organised by independent fact-checking organisations have improved the fact-checking skills of the students to a moderate level, there is need to have a structured curriculum tailored to address fact-checking skill gaps among journalism students.

To bridge the fact-checking skill gap among journalism lecturers, we recommend a structured train-the-trainer programme. Such a programme will offer regular, practical training that helps lecturers improve their own skills and, in turn, pass these on to their students. This is not a one-off workshop but a long-term investment in teaching and learning. We also strongly advocate partnerships between universities and independent fact-checking organisations. These collaborations can lead to co-developed modules, team-taught courses, and hands-on workshops. The National Universities Commission (NUC), fact-checking organisations, and journalism departments must come together to build a curriculum that equips students with the tools to uphold media credibility and tackle information disorder. Beyond curriculum changes, universities must invest in tools and resources for fact-checking. Students need to practise with real tools, not just learn theory.

We acknowledge that this study has its limitations. It relies on self-reported data and was conducted in a specific region. To get a fuller picture, future research should involve more institutions across different parts of Nigeria, where access and challenges may differ.

References

Adekunmisi, S. R., Oyewusi, F. O. & Aramide, K. (2022). Assessment of media literacy skills of undergraduates in universities in Ogun State, Nigeria. *Mousaion: South African Journal of Information Studies, 2.* https://doi.org/10.25159/2663-659x/9811

Adjin-Tettey, T. D., & Amenaghawon, F. (2024). Countering the threats of dis/misinformation: Fact-checking practices of students of two universities in West Africa. *Online Journal of Communication and Media Technologies, 14*(1), (1-14). https://doi.org/10.30935/ojcmt/14134

Amadu, E. (2024, November). NUCJ AAUA's reporters win Alfred Opubor Fact-Check Awards. *Echo Campus News:* https://echocampusnews.blogspot.com/2024/11/nucj-aauas-reporters-win-alfred-opubor.html

Azeez, B. (2024, November 12). Organisation develops curriculum to tackle misinformation. *Tribune Online:* https://tribuneonlineng.com/organisation-develops-curriculum-to-tackle-misinformation/

Cheruiyot, D., & Ferrer-Conill, R. (2018). “Fact-Checking Africa” epistemologies, data and the expansion of journalistic discourse. *Digital Journalism, 6*(8), (964-975). https://doi.org/10.1080/21670811.2018.1493940

Daniel, F. (2018, May 18). *These academics are on the frontlines of fake news research*. Poynter. https://www.poynter.org/fact-checking/2018/these-academics-are-on-the-frontlines-of-fake-news-research/

Donovan, A. M., and Rapp, D. N. (2020). Look it up: Online search reduces the problematic effects of exposures to inaccuracies. *Memory and Cognition*, *48*(7), (1128-1145). https://doi.org/10.3758/s13421-020-01047-z

Folarin, J. (2020). *Fact-Checking Guide: Brief on media organisations on the frontline of combating information disorder in Nigeria.* DUBAWA. https://dubawa.org/fact-checking-guide-brief-on-media-organisations-on-the-frontline-ofcombating-information-disorder-in-nigeria.

Ilesanmi, S. (2021). *Assessing fact-checking competence of journalism lecturers in Ondo State tertiary institutions.* Centre for Journalism Innovation and Development. https://idac.dubawa.org/wp-content/uploads/2022/05/Assessing-Fact-checking-Competence-of-Journalism-Lecturers-in-Ondo-State-Tertiary-Institutions.pdf

Li, L., Razi, S., & Yaakup, H. (2023). An overview of cultural and media literacy theory. *International Journal of Academic Research in Business and Social Sciences*, *13*(12), (875-886).

López-Meri, A., Doménech-Fabregat, H., Marcos-García, S. (2024). Digital competencies in verifying fake news: Assessing the knowledge and abilities of journalism students. *Societies*, *14*(5), (2-19). https://doi.org/10.3390/soc14050066

Matanji, F., Tully, M. & Mudavadi, K.C. (2024). Media literacy and fact-checking as proactive and reactive responses to misinformation in Kenya and Senegal. *African Journalism Studies*, (1-18). https://doi.org/10.1080/23743670.2024.2401782

Pavlović, D. & Milosavljević, I. (2017). Digital literacy in the journalism profession-students' perspective. *Digitalne medijske tehnologije i društveno-obrazovne promene,* https://www.researchgate.net/publication/349663366_DIGITAL_LITERACY_IN_THE_JOURNALISM_PROFESSION_-STUDENTS'_PERSPECTIVE]

Pew Research Centre.. (2019). *Social media fact sheet*.. https://www.pewresearch.org/internet/fact-sheet/social-media/

Potter, J. (2004). *Theory of media literacy: A cognitive approach*. SAGE Publications. https://doi.org/10.4135/9781483328881

Raji, R. (2020). *"Fake News": Understanding the scourge in Nigeria.* DUBAWA. https://dubawa.org/fake-news-understanding-the-scourge-in-nigeria/

Saunders, M., Lewis, P., & Thornhill, A. (2023). *Research methods for business students* (9th ed.). Pearson Education Limited.

Schreurs, L., & Vandenbosch, L. (2021). Introducing the social media literacy model with the case of the positivity bias on social media. *Journal of Children and Media*, *15*(3), (320-337).

Taber, K.S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education, 48*, (1273–1296). https://doi.org/10.1007/s11165-016-9602-2

Wasserman, H. (2020). Fake news from Africa: Panics, politics and paradigms. *Journalism, 21*(1), (3-16). https://doi.org/10.1177/1464884917746861

Wineburg, S. & Mcgrew, S. (2017). *Lateral reading: Reading less and learning more when evaluating digital information.* Stanford History Education Group Working Paper No. 2017-A1. https://doi.org/10.2139/ssrn.3048994

Wogu, J., O, Chukwu, C., O., Ugwuoke, J., C., Ugwulor-Onyinyechi C., C & Ononuju, N. C. (2019). A survey of student's media literacy skills in Nigerian universities. *Journal of Engineering and Applied Sciences, 14*, (5365-5373). http://dx.doi.org/10.36478/jeasci.2019.5365.5373

News, Social Media and Video Analytics: the MediaPot platform

Elisavet Palogiannidi[1], Sotiris Legkas[1], Dimitrios Vogiatzis[1,2], Manolis Mylonas[3], Vasso Koutsoupia[3], Vasileios Mezaris[3], Spyridoula Markou[4], George Zissis[4], Pantelis Theodosiou[4]

1. Institure of Informatics and Telecommunications, NCSR "Demokritos", Athens, Greece {epalogiannidi, sotirislegkas, dimitrv}@iit.demokritos.gr

2. The American college of Greece, Deree

3. Centre for Research and Technology Hellas (CERTH), Thessaloniki, Greece {emylonas, vkouts, bmezaris}@iti.gr

4. Athens Technology Centre (ATC), Athens, Greece {s.markou, g.zissis, p.theodosiou}@atc.gr

Abstract

Journalists have easy access to multiple sources of information, including news items, social media posts, and in multiple forms (text, images and video). But the sheer volume of news makes it nearly impossible to manually browse, inspect and analyze everything. Thus, it becomes increasingly difficult to understand the importance and reliability of each publication or news item. The MediaPot platform is an ongoing project that aims to address the above issues by collecting and analyzing multiple news sources with Artificial Intelligence (AI) techniques. First, it analyzes the text of the news items to detect entities and their relations to the news. Second, a question answering system was built using Large Language Models (LLMs) on social media posts and on news items. With that the journalist can ask free form questions. The platform also processes multimedia content, generating automatic and concise video summaries. In addition, it integrates advanced deep learning tools for analyzing images and videos – enabling automated detection of faces, objects, and actions, as well as the creation of descriptive tags, meme identification, and detection of inappropriate or disturbing content. Semantic representations further support visual similarity search.

Keywords: news and social media analytics with AI, multimodal news understanding, news summarization, news and social media question answering

Introduction

The emergence of social media as a major alternative source of news search has shaped a completely new media landscape (Wunderlich et al., 2022). Media organisations are gradually adapting to the emerging change in how citizens are informed about news stories, but the volume, the heterogeneity and the "noise" of the multimedia information generated by social media tools raises significant barriers to the smooth transition of media producers and news publishers to the new environment. The convergence of multimedia content from multiple sources is a promising solution for the businesses in this area, as it can combine the breaking news nature of the posts produced by citizens' journalists, mainly active in social media, with the credibility of the traditional way of news storytelling. Adopting data science and AI technologies for analyzing and verifying news and multimedia content will play an important role in enabling traditional media to integrate social networking information into their daily practice (Fernandes et al., 2024).

At the same time, the ability of journalists to critically engage with digital information environments is of paramount importance. Media and digital media literacy, as outlined by Hobbs (Hobbs, 2010), refers to a constellation of life skills that enable individuals to fully participate in a media-saturated and information-rich society. These include the ability to access and locate reliable information, analyze messages across different formats, evaluate sources and perspectives, create content using diverse digital tools, reflect on one's communication practices, and take informed social action. In this regard, digital literacy becomes particularly critical for journalists, who are expected to assess the quality and credibility of the information they process and disseminate (Hobbs & Jensen, 2009).

To this end, the MediaPot platform[12] aims to bring together media professionals with citizens' journalist groups under the auspices of a single and continuous process of producing quality and reliable news and relevant stories. The platform will develop a modern and dynamic news story creation environment that will exploit innovative solutions in the field of multimedia content analysis and news verification. Through a combination of AI and data mining services, the proposed ICT platform will analyse multimedia content (text, image and video) from social media, verify the trustworthiness of the reporting source and the reliability of the reported information, and produce related news stories in a semi-automated way.

The MediaPot platform is thus positioned not only as a technological innovation in the field of media production but also as a catalyst for enhancing journalists' critical digital competencies. By integrating advanced AI tools with a commitment to media literacy, the platform aims to support a more informed, ethical, and resilient news ecosystem.

Methodology: Text Analytics

First, we collected a dataset of news articles, from which entities were extracted. The entity types were: "location", "person" and "organisation". The nodes in the generated graph represent the articles and the entities, while the relationships between the nodes originate from the articles and reference one of the three entity types. An elaborate data cleaning process was followed. In Figure 1 we illustrate the pipeline that encompasses the stages from data acquisition to interaction with the

---

[12] https://mediapot.gr/en/

developed application. The data acquisition phase involves collecting articles from the web and extracting entities to be stored in a graph database. Following the application of some cleaning steps, the database is accessed by the backend service which serves as the core of the web application that the end users interact with. Users can select the desired entity type and enter a keyword to retrieve the relevant articles.

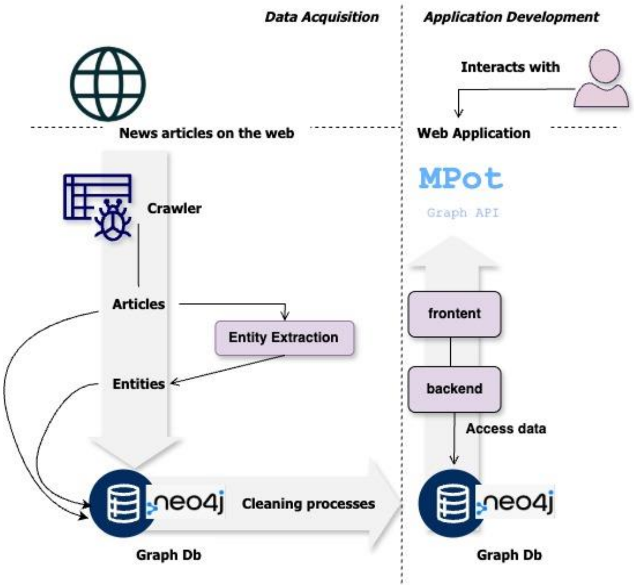


Figure 1: News Media Pipeline for the Graph creation

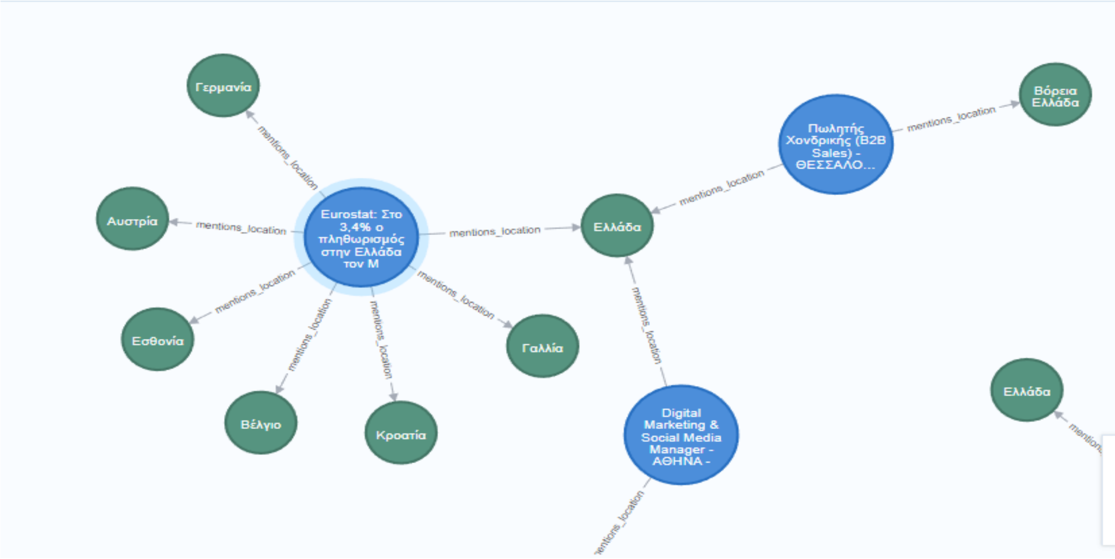
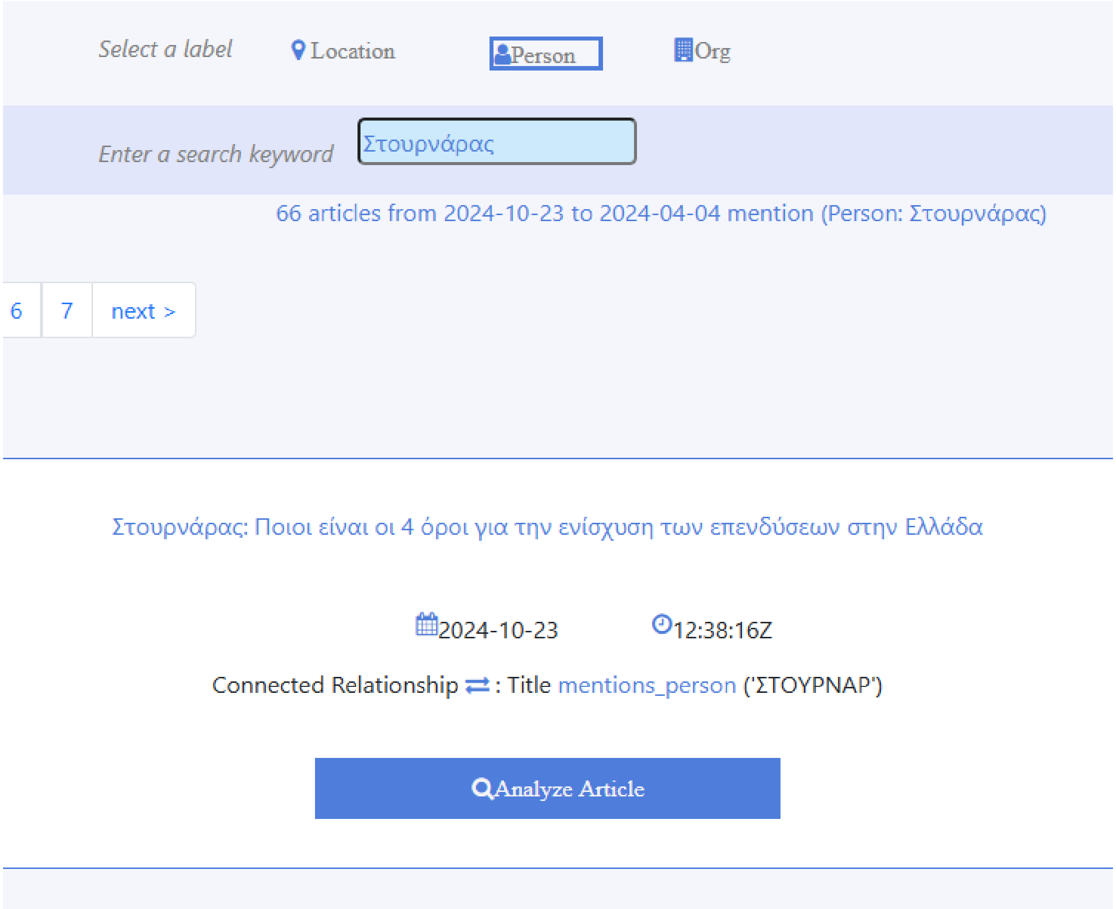

Γερμανία
Αυστρία
Εσθονία
Βέλγιο
Κροατία
Γαλλία
Eurostat: Στο 3,4% ο πληθωρισμός στην Ελλάδα τον Μ
mentions_location
Ελλάδα
Πωλητής Χονδρικής (B2B Sales) - ΘΕΣΣΑΛΟ...
Βόρεια Ελλάδα
Digital Marketing & Social Media Manager - ΑΘΗΝΑ -
Ελλάδα

Figure 2: A sample of the entity relationship diagramme. Articles are in blue, and entities in green

Next, the journalist can search via a dashboard for the *entity* he is interested in. In the next example, the journalist is interested in *persons* and in *Mr. Stournaras* (governor of the central bank of Greece). The dashboard will return all articles related to *Mr. Stournaras,* as well as the time and data of their publication (see Figure 3). Then the button “Analyze Article” will return all entities related to the specific article (see Figure 4), grouped according to the entity type (*locations, organisations* and *persons*).



Select a label: Location | Person | Org

Enter a search keyword: Στουρνάρας

66 articles from 2024-10-23 to 2024-04-04 mention (Person: Στουρνάρας)

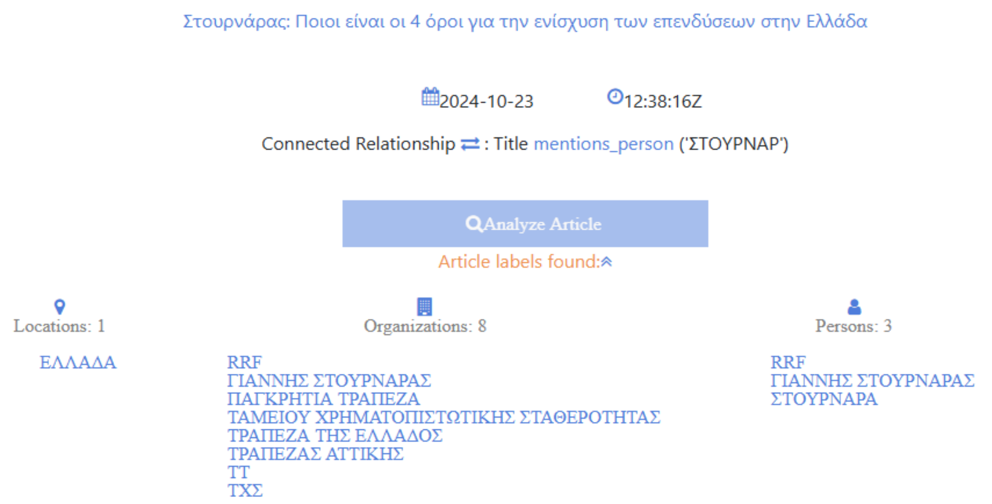
6 | 7 | next >

Στουρνάρας: Ποιοι είναι οι 4 όροι για την ενίσχυση των επενδύσεων στην Ελλάδα

2024-10-23 12:38:16Z

Connected Relationship ⇄ : Title mentions_person ('ΣΤΟΥΡΝΑΡ')

Figure 3: Dashboard access to the Network of entities and relations



Στουρνάρας: Ποιοι είναι οι 4 όροι για την ενίσχυση των επενδύσεων στην Ελλάδα

2024-10-23 12:38:16Z

Connected Relationship ⇄ : Title mentions_person ('ΣΤΟΥΡΝΑΡ')

Article labels found:

Locations: 1	Organizations: 8	Persons: 3
ΕΛΛΑΔΑ	RRF ΓΙΑΝΝΗΣ ΣΤΟΥΡΝΑΡΑΣ ΠΑΓΚΡΗΤΙΑ ΤΡΑΠΕΖΑ ΤΑΜΕΙΟΥ ΧΡΗΜΑΤΟΠΙΣΤΩΤΙΚΗΣ ΣΤΑΘΕΡΟΤΗΤΑΣ ΤΡΑΠΕΖΑ ΤΗΣ ΕΛΛΑΔΟΣ ΤΡΑΠΕΖΑΣ ΑΤΤΙΚΗΣ ΤΤ ΤΧΣ	RRF ΓΙΑΝΝΗΣ ΣΤΟΥΡΝΑΡΑΣ ΣΤΟΥΡΝΑΡΑ

Figure 4: All entities related to an article

X social Net Analytics

Data from the social network X were also collected and analyzed. The dataset spans the period from April 1, 2024, to November 27, 2024, and includes 15K Tweets in Greek and English. The collected fields include information directly extracted from the posts, such as, text, author, hashtags and engagement metrics. Next, the author field is discarded from further processing to avoid any form of identification. The textual data were used to develop a Retrieval-Augmented Generation (RAG) system, to provide answers to queries. The text of each post was converted into an embedding vector using the BGE-M3 model[13] which is multilingual and stored in a vector database (Milvus). For each input query an embedding representation is generated and compared for similarity against all entries in the vector database. The top five most similar posts are retrieved and used to construct a prompt for a LLM, providing context to generate the final summarized answer. We used the multilingual *aya-expanse-8b-Q4_K_M-GGUF LLM model*[14]. Figure 5 illustrates the RAG architecture. The developed system is multilingual and some query examples along with the generated answers are listed in Figure 6. The first two queries demonstrate the system's multilingual capabilities in Greek and in English. In contrast, the last query is unrelated to the X dataset, and the system responds by indicating that it lacks the necessary knowledge to answer the question.

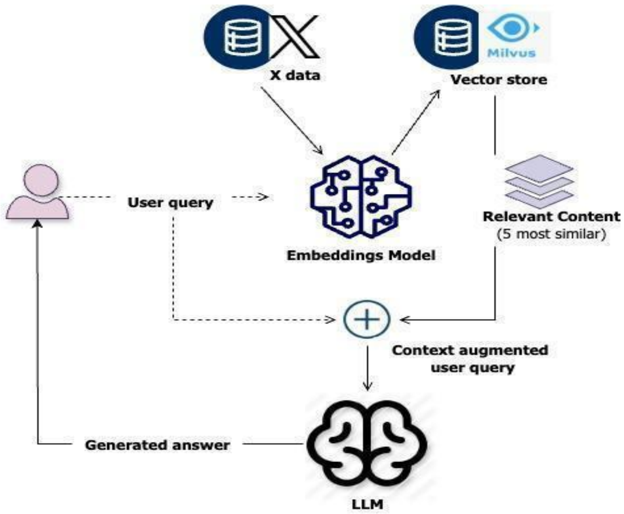
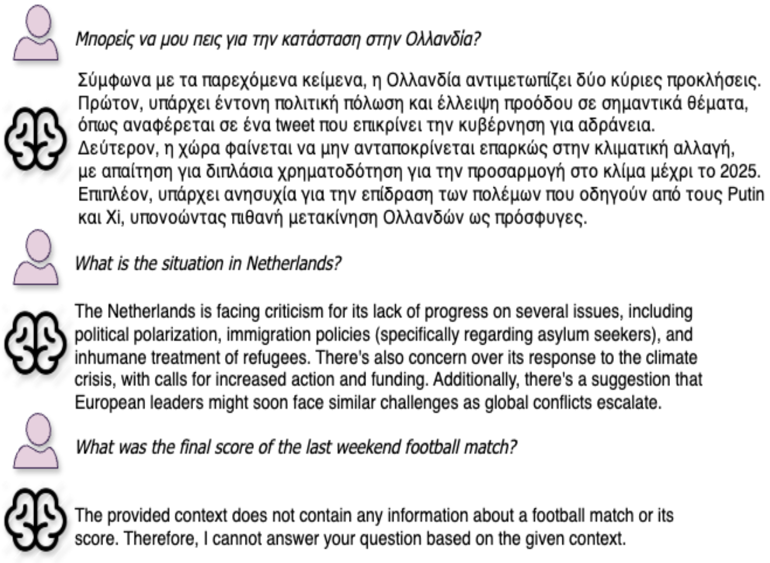


Figure 5: RAG architecture for Tweets

---

[13] https://huggingface.co/BAAI/bge-m3

[14] https://builds.modular.com/models/aya-expanse/8B



Μπορείς να μου πεις για την κατάσταση στην Ολλανδία?

Σύμφωνα με τα παρεχόμενα κείμενα, η Ολλανδία αντιμετωπίζει δύο κύριες προκλήσεις. Πρώτον, υπάρχει έντονη πολιτική πόλωση και έλλειψη προόδου σε σημαντικά θέματα, όπως αναφέρεται σε ένα tweet που επικρίνει την κυβέρνηση για αδράνεια. Δεύτερον, η χώρα φαίνεται να μην ανταποκρίνεται επαρκώς στην κλιματική αλλαγή, με απαίτηση για διπλάσια χρηματοδότηση για την προσαρμογή στο κλίμα μέχρι το 2025. Επιπλέον, υπάρχει ανησυχία για την επίδραση των πολέμων που οδηγούν από τους Putin και Xi, υπονοώντας πιθανή μετακίνηση Ολλανδών ως πρόσφυγες.

What is the situation in Netherlands?

The Netherlands is facing criticism for its lack of progress on several issues, including political polarization, immigration policies (specifically regarding asylum seekers), and inhumane treatment of refugees. There's also concern over its response to the climate crisis, with calls for increased action and funding. Additionally, there's a suggestion that European leaders might soon face similar challenges as global conflicts escalate.

What was the final score of the last weekend football match?

The provided context does not contain any information about a football match or its score. Therefore, I cannot answer your question based on the given context.

Figure 6: Examples of queries and their corresponding answers, as generated by the RAG architecture

News Analytics

We have also built a question answering system for news articles based on RAG engineering, in a way similar to that for Tweets. The main difference is that the news articles are considerably longer than Tweets. Thus, the text of news' articles was split into segments (chunking) with the aid of *Langchain*[15] library. We used a corpus of 5,393 news items split into 13,746 segments. The Tweets were represented with the BGE-M3, which is multilingual. See Figure 7, where the two-phase architecture is depicted. The offline phase where the enrichment of the LLM takes place with the news articles, and the online phase where answers are generated as a response to the users' questions. The whole system is based on the *Gemma 3*[16] LLM with 12 billion parameters. In Figure 8 we depict the dashboard that permits the user to ask questions, and in Figure 9 the top k-articles related to the user's query are depicted.

---

[15] https://www.langchain.com

[16] https://deepmind.google/models/gemma/gemma-3/

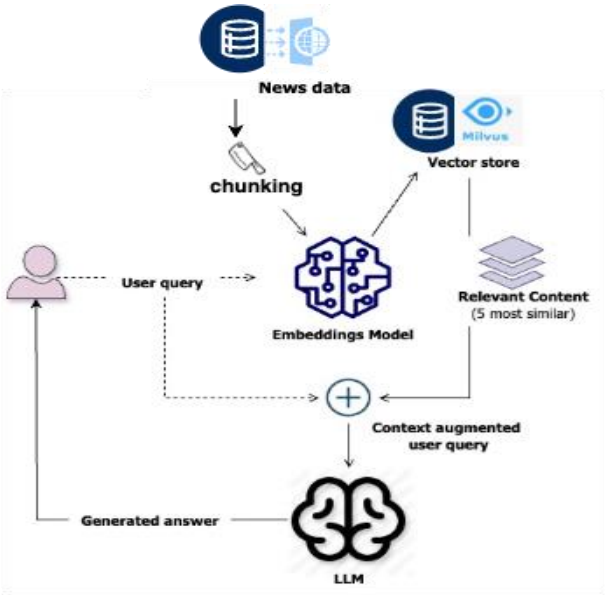
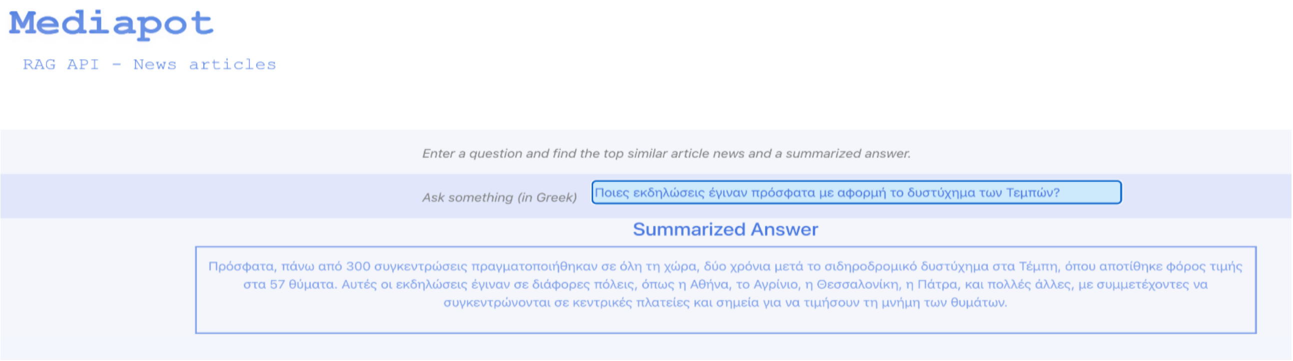


Figure 7: Architecture of the question answering system on News Articles

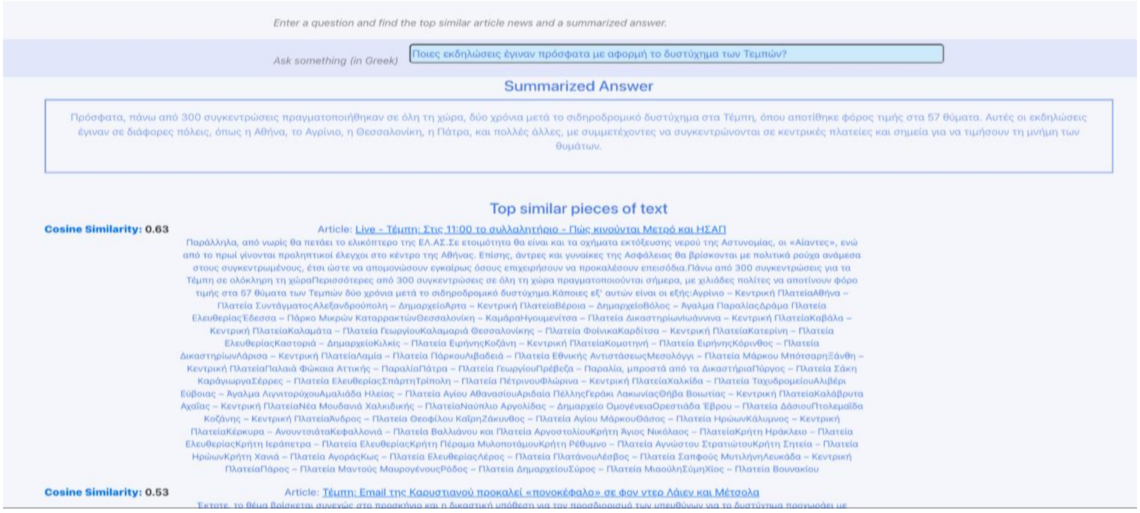


Enter a question and find the top similar article news and a summarized answer.

Ask something (in Greek) Ποιες εκδηλώσεις έγιναν πρόσφατα με αφορμή το δυστύχημα των Τεμπών?

Πρόσφατα, πάνω από 300 συγκεντρώσεις πραγματοποιήθηκαν σε όλη τη χώρα, δύο χρόνια μετά το σιδηροδρομικό δυστύχημα στα Τέμπη, όπου αποτίθηκε φόρος τιμής στα 57 θύματα. Αυτές οι εκδηλώσεις έγιναν σε διάφορες πόλεις, όπως η Αθήνα, το Αγρίνιο, η Θεσσαλονίκη, η Πάτρα, και πολλές άλλες, με συμμετέχοντες να συγκεντρώνονται σε κεντρικές πλατείες και σημεία για να τιμήσουν τη μνήμη των θυμάτων.

Figure 8: Dashboard of question answering system for News


Enter a question and find the top similar article news and a summarized answer.

Ask something (in Greek) Ποιες εκδηλώσεις έγιναν πρόσφατα με αφορμή το δυστύχημα των Τεμπών?

Πρόσφατα, πάνω από 300 συγκεντρώσεις πραγματοποιήθηκαν σε όλη τη χώρα, δύο χρόνια μετά το σιδηροδρομικό δυστύχημα στα Τέμπη, όπου αποτίθηκε φόρος τιμής στα 57 θύματα. Αυτές οι εκδηλώσεις έγιναν σε διάφορες πόλεις, όπως η Αθήνα, το Αγρίνιο, η Θεσσαλονίκη, η Πάτρα, και πολλές άλλες, με συμμετέχοντες να συγκεντρώνονται σε κεντρικές πλατείες και σημεία για να τιμήσουν τη μνήμη των θυμάτων.

Cosine Similarity: 0.63

Article: Live - Τέμπη: Στις 11:00 το συλλαλητήριο - Πώς κινούνται Μετρό και ΗΣΑΠ

Παράλληλα, από νωρίς θα πετάει το ελικόπτερο της ΕΛ.ΑΣ.Σε ετοιμότητα θα είναι και τα οχήματα εκτόξευσης νερού της Αστυνομίας, οι «Αίαντες», ενώ από το πρωί γίνονται προληπτικοί έλεγχοι στο κέντρο της Αθήνας. Επίσης, άντρες και γυναίκες της Ασφάλειας θα βρίσκονται με πολιτικά ρούχα ανάμεσα στους συγκεντρωμένους, έτσι ώστε να απομονώσουν εγκαίρως όσους επιχειρήσουν να προκαλέσουν επεισόδια.Πάνω από 300 συγκεντρώσεις για τα Τέμπη σε ολόκληρη τη χώραΠερισσότερες από 300 συγκεντρώσεις σε όλη τη χώρα πραγματοποιούνται σήμερα, με χιλιάδες πολίτες να αποτίνουν φόρο τιμής στα 57 θύματα των Τεμπών δύο χρόνια μετά το σιδηροδρομικό δυστύχημα.Κάποιες εξ' αυτών είναι οι εξής:Αγρίνιο – Κεντρική ΠλατείαΑθήνα – Πλατεία ΣυντάγματοςΑλεξανδρούπολη – ΔημαρχείοΆρτα – Κεντρική ΠλατείαΒέροια – ΔημαρχείοΒόλος – Άγαλμα ΠαραλίαςΔράμα Πλατεία ΕλευθερίαςΈδεσσα – Πάρκο Μικρών ΚαταρρακτώνΘεσσαλονίκη – ΚαμάραΗγουμενίτσα – Πλατεία ΔικαστηρίωνΙωάννινα – Κεντρική ΠλατείαΚαβάλα – Κεντρική ΠλατείαΚαλαμάτα – Πλατεία ΓεωργίουΚαλαμαριά Θεσσαλονίκης – Πλατεία ΦοίνικαΚαρδίτσα – Κεντρική ΠλατείαΚατερίνη – Πλατεία ΕλευθερίαςΚαστοριά – ΔημαρχείοΚιλκίς – Πλατεία ΕιρήνηςΚοζάνη – Κεντρική ΠλατείαΚομοτηνή – Πλατεία ΕιρήνηςΚόρινθος – Πλατεία ΔικαστηρίωνΛάρισα – Κεντρική ΠλατείαΛαμία – Πλατεία ΠάρκουΛιβαδειά – Πλατεία Εθνικής ΑντιστάσεωςΜεσολόγγι – Πλατεία Μάρκου ΜπότσαρηΞάνθη – Κεντρική ΠλατείαΠαλαιά Φώκαια Αττικής – ΠαραλίαΠάτρα – Πλατεία ΓεωργίουΠρέβεζα – Παραλία, μπροστά από τα ΔικαστήριαΠύργος – Πλατεία Σάκη ΚαράγιωργαΣέρρες – Πλατεία ΕλευθερίαςΣπάρτηΤρίπολη – Πλατεία ΠέτρινουΦλώρινα – Κεντρική ΠλατείαΧαλκίδα – Πλατεία ΤαχυδρομείουΑλιβέρι Εύβοιας – Άγαλμα ΛιγνιτορύχουΑμαλιάδα Ηλείας – Πλατεία Αγίου ΑθανασίουΑριδαία ΠέλληςΓεράκι ΛακωνίαςΘήβα Βοιωτίας – Κεντρική ΠλατείαΚαλάβρυτα Αχαΐας – Κεντρική ΠλατείαΝέα Μουδανιά Χαλκιδικής – ΠλατείαΝαύπλιο Αργολίδας – Δημαρχείο ΟμογένειαΟρεστιάδα Έβρου – Πλατεία ΔάσιουΠτολεμαΐδα Κοζάνης – Κεντρική ΠλατείαΆνδρος – Πλατεία Θεοφίλου ΚαΐρηΖάκυνθος – Πλατεία Αγίου ΜάρκουΘάσος – Πλατεία ΗρώωνΚάλυμνος – Κεντρική ΠλατείαΚέρκυρα – ΑνουντσιάταΚεφαλλονιά – Πλατεία Βαλλιάνου και Πλατεία ΑργοστολίουΚρήτη Άγιος Νικόλαος – ΠλατείαΚρήτη Ηράκλειο – Πλατεία ΕλευθερίαςΚρήτη Ιεράπετρα – Πλατεία ΕλευθερίαςΚρήτη Πέραμα ΜυλοποτάμουΚρήτη Ρέθυμνο – Πλατεία Αγνώστου ΣτρατιώτουΚρήτη Σητεία – Πλατεία ΗρώωνΚρήτη Χανιά – Πλατεία ΑγοράςΚως – Πλατεία ΕλευθερίαςΛέρος – Πλατεία ΠλατάνουΛέσβος – Πλατεία Σαπφούς ΜυτιλήνηΛευκάδα – Κεντρική ΠλατείαΠάρος – Πλατεία Μαντούς ΜαυρογένουςΡόδος – Πλατεία ΔημαρχείουΣύρος – Πλατεία ΜιαούληΣύμηΧίος – Πλατεία Βουνακίου

Cosine Similarity: 0.53

Article: Τέμπη: Email της Καρυστιανού προκαλεί «πονοκέφαλο» σε φον ντερ Λάιεν και Μέτσολα

Έκτοτε, το θέμα βρίσκεται συνεχώς στο προσκήνιο και η δικαστική υπόθεση για τον προσδιορισμό των υπευθύνων για το δυστύχημα προχωράει με

Figure 9: Top k articles related to the user's question

Multimedia Analytics

The MediaPot platform utilizes state-of-the-art deep learning tools designed for processing and analyzing multimedia content (images and videos), specifically tailored to the dynamic and evolving needs of news story composition. In this environment, the media annotation service is integrated, which is designed to identify, organise, and categorize significant information and features within large multimedia collections. The system encompasses a variety of methods, including action recognition, object detection, face detection and recognition, image captioning and automatic natural language text generation for content description, meme detection, and the identification of disturbing and inappropriate or violent (Not-Safe-For-Work) content. Additionally, a vector-based representation of multimedia content is provided, which can be used for retrieving semantically similar content elements. Each method employs state-of-the-art deep learning models, such as Faster R-CNN (Ren et al., 2016), RAM-14M (Zhang et al., 2023), ResNet152 (He et al., 2016), SlowFast R50 (Feichtenhofer et al., 2019), InceptionResNetV1 (Cao et al., 2018), MemeTector (Koutlis et al., 2022), and CM-Refinery (Sarridis et al., 2022). The service is capable of recognizing a wide range of entities, including almost 46,000 international celebrities (e.g., athletes, artists, politicians), 400 types of activities (e.g., exercise, basketball), 6,500 objects (e.g., car, table, train), and specialized categories (e.g., professions, colours).

Additionally, the platform facilitates semantic multimedia retrieval through the use of vector-based representations, providing tools for reverse search via visual similarity and near-duplicate detection. The visual similarity search leverages cross-modal embedding for detecting related images based on their features, while near-duplicate detection focuses on identifying near-identical or duplicated multimedia content. Technologies such as CLIP (Radford et al., 2021) and DnS (Kordopatis-Zilos et al., 2022) are employed for these purposes.

The development of video analysis tools is based on the need of reporters to process and analyse the excessive number of videos in a fast and efficient way. In this context, the MediaPot platform integrates components for video segmentation and summarization that rely on the use of trained state-of-the-art AI models. The temporal segmentation of the video is made according to two different granularities. Segmentation into shots (i.e., sequences of frames captured uninterruptedly by a single camera) is performed using a model of TransnetV2 (Soucek & Lokoc, 2024) trained using synthetically-created data from the TRECVID IACC.3 dataset (Awad et al., 2017) and the ground-truth data of the ClipShots dataset (Tang et al. 2028). The utilized model exhibits state-of-the-art performance on various benchmarking datasets (e.g., ClipShots, BBC Planet Earth documentary series (Baraldi et al., 2015a), and RAI (Baraldi et al., 2015b), as shown in Table 1. Segmentation into subshots (parts of a shot related to a different activity of the camera during the recording, e.g., camera pan/tilt, zoom in/out) is performed based on the motion-driven approach from

(Apostolidis et al., 2018). This method segments a video into visually coherent parts that correspond to individual video capturing activities (e.g., camera pan and tilt, change in focal length and camera displacement) by extracting and evaluating the region-level spatio-temporal distribution of the optical flow over sequences of neighbouring video frames.

TransnetV2	ClipShots	BBC	RAI
F-Score (%)	77.9	96.2	93.9

Table 1: Performance (F-Score (%)) of TransnetV2 in different benchmark datasets for video shot detection

Using the specified video shots or subshots, the video summarization component generates a complete and concise summary of the video by selecting the most important and informative shots or subshots of it. The scoring of the video segments is performed using a model of PGL-SUM [15] trained on the TVSum dataset (Yale et al., 2015). PGL-SUM shows state-of-the-art performance on the SumMe (Gygli et al., 2014) and TVSum datasets (see Table 2), that are the most commonly used ones for benchmarking video summarization methods (Apostolidis et al., 2021b). It employs global and local multi-head attention mechanisms with positional encoding, to model the frames' position and dependence at different levels of granularity. The output of PGL-SUM is a series of frame-level importance scores that are used (averaged) to compute segment-level importance. The segments are then ranked in descending order and the top-k of them are selected for inclusion in the summary (k relates to the summary duration). As a note, the use of attention-based network architectures has been proven as an effective approach for training models for video summarization both in a supervised (Apostolidis et al., 2021a) and unsupervised (Apostolidis et al. 2022) manner. Moreover, it allows the production of visual explanations about the summarization results using attention-based explanation signals, as presented in (Apostolidis et al., 2023; Tsigos et al., 2024).

The whole procedure of video segmentation and summarization in the platform is the following: the video enters the segmentation module, where shot and subshot information are extracted using the respective models. Then, this information is passed into the summarization module, where the video is first downsampled to 2 frames per second and deep features are extracted using GoogleNet (Szegedy et al., 2015) pretrained on ImageNet (Deng et al., 2009). PGL-SUM provides segment importance scores using the deep features, and the video summary is generated by combining the segment importance scores with the temporal shot information. The output includes the information on the shots and subshots, as well as the summary of the video in .mp4 format. The pipeline is shown in Figure 10.

PGL-SUM	SumMe	TVSum
F-Score (%)	57.1	62.7

Table 2: Performance (F-Score (%)) of PGL-SUM in two benchmark datasets for video summarization

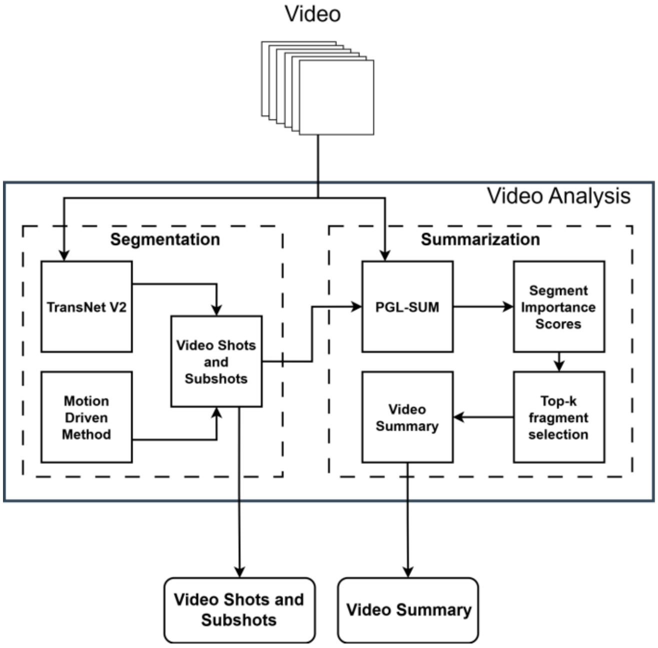


Figure 10: The applied processing pipeline for video segmentation and summarization

Model Selection Criteria

The models integrated in the MediaPot platform were selected after a thorough review of the literature, with two main criteria in mind. First, all chosen models have been shown in their respective publications to achieve state-of-the-art performance on benchmark datasets, as detailed in the cited works (Ren et al., 2016; Zhang et al., 2023; He et al., 2016; Feichtenhofer et al., 2019; Cao et al., 2018; Koutlis et al., 2022; Sarridis et al., 2022; Radford et al., 2021; Kordopatis-Zilos et al., 2022; Soucek & Lokoc 2024; Apostolidis et al., 2018; Apostolidis et al., 2021a; Szegedy et al., 2015). Second, they are efficient, low-cost, and publicly available, making them straightforward to integrate without requiring excessive computational resources. This combination of strong performance and ease of deployment ensures that the platform provides reliable and scalable tools for journalistic use.

Conclusions, Limitations Future Steps

The MediaPot platform for news, social media, and video analytics enables journalists to work more efficiently with vast volumes of data, saving time while detecting and investigating significant events. It promotes investigative journalism by providing a secure, collaborative environment where reporters handling massive datasets can better organise, manage, and verify their material. Journalists can jointly develop stories or thematic investigations, examine related multimedia content, and cross-check information across online news and social media posts. By combining text, image, and video analytics, the platform supports tracing discussions, connecting entities, and building a holistic understanding of complex narratives. It also allows journalists to assess and “rate” the credibility of information, strengthening verification practices. Advanced tools such as semantic search, reverse image matching, multimedia annotation, and question answering on both Tweets and news articles further assist in validating authenticity and capturing insights, including public opinion. Finally, the entity–relationship graph can be enriched with video tags, unifying multimedia information into a coherent knowledge base. This integration enables journalists to answer questions such as *who was where*by combining textual and visual evidence.

Acknowledgements

This work was supported by project MediaPot (ΤΑΕΔΚ-06196), carried out within the framework of the National Recovery and Resilience Plan Greece2.0, funded by the European Union – NextGenerationEU, under the call RESEARCH-CREATE-INNOVATE.

References

Apostolidis, K., Apostolidis, E., Mezaris, V. (2018). A Motion-Driven Approach for Fine-Grained Temporal Segmentation of User-Generated Videos. In Schoeffmann K. et al. (Eds.), *MultiMedia Modeling*. Lecture Notes in Computer Science, vol 10704, (223-235).Springer. https://link.springer.com/chapter/10.1007/978-3-319-73603-7_18

Apostolidis, E., Balaouras, G., Mezaris, V., & Patras, I. (2021). Combining global and local attention with positional encoding for video summarization. In *2021 IEEE International Symposium on Multimedia*, (226-234). IEEE. https://doi.org/10.1109/ISM52913.2021.00045

Apostolidis, E., Adamantidou, E., Metsai, A. I., Mezaris, V., & Patras, I. (2021b). Video summarization using deep neural networks: A survey. *Proceedings of the IEEE, 109*(11), (1838-1863). https://ieeexplore.ieee.org/document/9527418

Apostolidis, E., Balaouras, G., Mezaris, V., & Patras, I. (2022). Summarizing videos using concentrated attention and considering the uniqueness and diversity of the video frames. In *Proceedings of the 2022 International Conference on Multimedia Retrieval,* (407-415). https://dl.acm.org/doi/10.1145/3512527.3531355

Apostolidis, E., Mezaris, V & Patras, I. (2023). A Study on the Use of Attention for Explaining Video Summarization. In *Proceedings of the NRSVM Workshop at ACM Multimedia 2023*, Ottawa, Canada, Oct.-Nov. 2023.

Awad, G., Butt, A. A., Fiscus, J., Joy, D., Delgado, A., Mcclinton, W., ... & Huet, B. (2017). Trecvid 2017: Evaluating ad-hoc and instance video search, events detection, video captioning, and hyperlinking. In *TREC video retrieval evaluation (TRECVID),*(1-18). https://www-nlpir.nist.gov/projects/tvpubs/tv17.papers/tv17overview.pdf

Baraldi, L., Grana, C. & Cucchiara, R. (2015a). A Deep Siamese Network for Scene Detection in Broadcast Videos. In *Proceedings of the 23rd ACM international conference on Multimedia,* (1199- 1202). Association for Computing Machinery. https://dl.acm.org/doi/pdf/10.1145/2733373.2806316?utm_source=chatgpt.com

Baraldi, L., Grana, C., Cucchiara, R. (2015b). Shot and Scene Detection via Hierarchical Clustering for Re-using Broadcast Video. In Azzopardi, G., Petkov, N. (eds) *Computer Analysis of Images and Patterns.* Lecture Notes in Computer Science, Vol. 9256, (251-262). Springer. https://link.springer.com/chapter/10.1007/978-3-319-23192-1_26

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). VGGFace2: A dataset for recognising faces across pose and age. https://arxiv.org/pdf/1710.08092

Deng, J., Dong, W., Socher, R., Li, J., Kai Li & Li Fei-Fei (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (248-255). IEEE. https://doi.org/10.1109/CVPR.2009.5206848

Feichtenhofer, C., Fan, H., Malik, J., & He, K. (2019). Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, (6202-6211). https://doi.org/10.1109/ICCV.2019.00686

Fernandes, E., Moro, S., & Cortez, P. (2024). A data-driven approach to improve online consumer subscriptions by combining data visualization and machine learning methods. *International Journal of Consumer Studies, 48*(2), (409-421). https://doi.org/10.1111/ijcs.13030

Gygli, M., Grabner, H., Riemenschneider, H., Van Gool, L. (2014). Creating Summaries from User Videos. In Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) *Computer Vision – ECCV 2014.* Lecture Notes in Computer Science, vol 8695, (505-520). Springer.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (770-778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Hobbs, R., & Jensen, A. (2009). The Past, Present, and Future of Media Literacy Education. *Journal of Media Literacy Education*, *1*(1), (1–11). https://digitalcommons.uri.edu/jmle/vol1/iss1/1/

Hobbs, R. (2010). *Digital and Media Literacy: A Plan of Action*. Aspen Institute. https://files.eric.ed.gov/fulltext/ED519422.pdf

Kordopatis-Zilos, G., Tzelepis, C., Papadopoulos, S., Kompatsiaris, I., & Patras, I. (2022). DnS: Distill-and-Select for Efficient and Accurate Video Indexing and Retrieval. *International Journal of Computer Vision*, *130*. https://link.springer.com/article/10.1007/s11263-022-01566-6

Koutlis, C., Schinas, M., & Papadopoulos, S. (2022). MemeTector: Enforcing deep focus for meme detection. In *Proceedings of the 30th ACM International Conference on Multimedia*, (2565-2572). Association for Computing Machinery. https://doi.org/10.1145/3503161.3547799

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. *International Conference on Machine Learning*, *139*, (8742-8763). https://proceedings.mlr.press/v139/radford21a.html

Ren, S., He, K., Girshick, R. & Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems, 28*, (91-99). https://papers.nips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html

Sarridis, I., Koutlis, C., Papadopoulou, O., & Papadopoulos, S. (2022). Leveraging large-scale multimedia datasets to refine content moderation models. *Multimedia Tools and Applications,81.* https://doi.org/10.1007/s11042-022-12035-4

Soucek, T. & Lokoc. J. (2024). TransNet V2: An Effective Deep Network Architecture for Fast Shot Transition Detection. In *Proceedings of the 32nd ACM International Conference on Multimedia,* (1189-1198). Association for Computing Machinery.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going Deeper with Convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition*, (1-9).IEEE. https://doi.org/10.1109/CVPR.2015.7298594

Tang, S., Feng, L., Kuang, Z., Chen, Y., & Zhang, W. (2018). Fast Video Shot Transition Localization with Deep Structured Models. *ArXiv*. https://arxiv.org/abs/1808.04234

Tsigos, K., Apostolidis, E., & Mezaris, V. (2024). An integrated framework for multi-granular explanation of video summarization. *Frontiers in Signal Processing, 4,*(1-16). https://www.frontiersin.org/journals/signal-processing/articles/10.3389/frsip.2024.1433388/full

Wunderlich, L., Hölig, S., & Hasebrink, U. (2022). Does journalism still matter? The role of journalistic and non-journalistic sources in young peoples’ news-related practices. *The International Journal of Press Politics, 27*(3), (569–588). https://doi.org/10.1177/19401612211072547

Song, Y., Vallmitjana, J., Stent A. & Jaimes A. (2015). TVSum: Summarizing web videos using titles. In *2015 IEEE Conference on Computer Vision and Pattern Recognition,* (5179-5187). IEEE. https://doi.org/10.1109/CVPR.2015.7299104

Zhang, Y., Huang, X., Ma, J., Li, Z., Luo, Z., Xie, Y., Qin, Y., Luo, T., Li, Y., Liu, S., Guo, Y., & Zhang, L. (2023). Recognize Anything: A strong image tagging model. *Proceedings of the IEEE CVF Conference on Computer Vision and Pattern Recognition*, (17287- 17298). IEEE. https://doi.org/10.1109/CVPR52729.2023.01657

Harnessing Emerging Technologies to Combat Disinformation: Innovation, Ethics, and Resilience in Journalism

Ioanna Georgia Eskiadi[1], Dr. Nikolaos Panagiotou[2]

1. Ph.D. Candidate, Researcher in Emerging Media Tech, Department of Journalism and Mass Communications, Aristotle University of Thessaloniki, Greece, ieskiadi@jour.auth.gr

2. Professor, Department of Journalism and Mass Communications, Aristotle University of Thessaloniki, Greece, npanagiotou@gapps.auth.gr

Abstract

This paper explores how emerging technologies—artificial intelligence (AI), blockchain, and immersive media—are integrated into journalism to counter disinformation. The study investigates their impact on journalistic transparency, trust, and audience engagement, alongside the ethical risks these tools may pose. Employing a mixed-methods approach, the research combines longitudinal content analysis with global case studies of five media organisations that have adopted these technologies in their editorial workflows. Findings indicate that AI supports real-time fact-checking and narrative detection, while blockchain enhances content authentication and accountability. Immersive storytelling, particularly through VR and 360° video, deepens emotional engagement and fosters public empathy. However, challenges such as algorithmic bias, high infrastructure demands, and ethical concerns around immersive content persist. The study concludes that while no single technology can eliminate disinformation, their ethical and transparent integrations, supported by media literacy and interdisciplinary collaboration, can significantly reinforce journalism's democratic mission. It advocates for global frameworks that promote innovation while preserving editorial integrity and inclusivity, ensuring that emerging technologies serve as tools for resilience rather than manipulation in the evolving digital media landscape.

Keywords: disinformation, emerging technologies, artificial intelligence, transparency

Introduction

The rapid proliferation of disinformation has emerged as a defining challenge for journalism in the digital age, threatening public trust, democratic participation, and the credibility of media institutions. From manipulated content and AI-generated deepfakes to the amplification of polarizing narratives on digital platforms, the disinformation crisis is not only technological but also profoundly sociopolitical. As traditional newsrooms struggle to maintain editorial standards in this fast-paced, data-saturated environment, emerging technologies—namely artificial intelligence

(AI), blockchain, and immersive media—have become pivotal tools in the fight against falsehoods and the restoration of journalistic integrity.

Digital journalism has transitioned from linear to participatory and immersive forms of engagement (Erdal, 2011; Kolodzy, 2012), reshaping audience interaction and information flow. This transformation has introduced convergence journalism, participatory models, and extended reality (XR) systems, altering the communicative dynamics between journalists and audiences. The integration of user-generated content (UGC), social networking, and algorithmic curation has led to the rise of networked publics and interactive media environments that blur the boundaries between sender and receiver (Anderson, Bell, & Shirky, 2015; Liu, 2023). Within this complex ecosystem, traditional gatekeeping roles have weakened, while journalism's credibility is increasingly contingent upon innovation, transparency, and audience trust.

Artificial Intelligence has become a cornerstone of digital transformation in journalism. Applications include automated content generation, large-scale fact-checking, natural language processing, audience analytics, and recommendation systems (Sun, Hu, & Wu, 2022). Leading media organisations such as Reuters, Forbes, and the Associated Press use AI to enhance productivity and minimize human error (Barceló-Ugarte, Pérez-Tornero, & Vila-Fumàs, 2021). However, concerns over algorithmic bias, opacity, and ethical risks persist. Scholars stress the importance of AI literacy, transparency, and human oversight in news production to counteract the amplification of existing inequalities (Deuze & Beckett, 2022; Kissinger, Schmidt, & Huttenlocher, 2021).

In parallel, blockchain technology has emerged as a promising solution for improving accountability in journalism. It allows for the secure timestamping of news content, provenance tracking, and tamper-proof archiving (Niranjanamurthy et al., 2019). These affordances are particularly relevant in contexts where information is vulnerable to manipulation. However, blockchain integration faces sustainability challenges and may not be scalable across all media landscapes (Lee et al., 2021). Additionally, the technology requires high levels of digital infrastructure and literacy, potentially reinforcing global media inequalities.

Immersive media—including VR, AR, and 360-degree video—offers new opportunities for experiential storytelling. Immersive journalism has been found to increase emotional engagement, empathy, and memory retention, particularly when addressing complex or distant events (Greber et al., 2023; Steed et al., 2023). These formats empower audiences engage with narratives through simulated first-person perspectives transforming passive viewers into active participants (Suh & Prophet, 2018). Nevertheless, scholars highlight significant risks, including user data privacy, psychological manipulation, and the ethical implications of simulating traumatic events (Floridi et al., 2018; Paino Ambrosio & Rodríguez Fidalgo, 2021).

This study builds on interdisciplinary insights from media innovation, AI ethics, blockchain systems, and immersive storytelling to assess how these emerging technologies are being integrated into news media organisations from 2012 to 2026. By employing a mixed-methods research design—combining longitudinal content analysis with global case studies—the study evaluates the real-world impact and ethical challenges of these technologies. The goal is to identify best practices, uncover implementation gaps, and propose strategic frameworks for ethical, effective, and inclusive technology adoption in journalism.

Ultimately, while no single innovation can fully address the multifaceted threat of disinformation, the convergence of AI, blockchain, and immersive media—when implemented responsibly—can reinforce journalistic credibility, foster civic resilience, and re-establish journalism's core democratic mission in the digital era.

Methodology

This study adopts a qualitative case study methodology to explore how emerging technologies, namely artificial intelligence (AI), blockchain, and immersive media, have been employed by news media organisations in their efforts to combat disinformation. The case study approach was selected due to its capacity to provide a nuanced understanding of complex, context-dependent phenomena within real-life settings (Yin, 2018). It enables an in-depth examination of the strategic, ethical, and operational dimensions of technological innovation in journalism. This approach enables a rich, context-sensitive investigation of complex media innovations in real-life organisational settings (Yin, 2018).

Five media organisations were purposefully selected based on predefined criteria. These included demonstrable implementation of at least one of the aforementioned technologies in editorial workflows; availability of documentation or media outputs illustrating the technological application; and a reputation for innovation and transparency in digital journalism practices. The selected cases reflect variation in geographical location, scale of operations, and type of technology adopted, facilitating comparative insights into different institutional responses to the disinformation crisis. The selected cases represent diverse geopolitical contexts, organisational sizes, and technological strategies, thus allowing comparative cross-case analysis.

Data were collected from a range of publicly accessible sources. These included organisational reports, white papers, peer-reviewed academic articles, and media content such as immersive video productions or blockchain-verified news outputs. Where available, additional information was obtained through expert commentary, interviews, and public statements by newsroom professionals. This triangulation of data sources aimed to enhance the credibility and richness of the analysis (Patton, 2015).

Data collection occurred between January and April 2025 and followed a triangulated strategy that incorporated primary, secondary, and audiovisual sources.

Primary data included official organisational publications such as annual innovation reports, white papers, press releases, and publicly available transcripts or statements by editorial staff. Secondary data were drawn from peer-reviewed literature, industry reports, and journalistic analyses that documented and interpreted the use of technologies in the selected organisations. Media outputs such as 360° videos, blockchain-verified news articles, and AI-generated content were analyzed to understand how these technologies manifested in narrative, ethical framing, and audience interaction. Where accessible, interviews and conference recordings featuring newsroom innovators or CTOs were also incorporated to contextualize institutional motivations and constraints. All collected material was archived digitally, coded thematically, and systematically analyzed using NVivo software to ensure consistency across data types and sources.

The data were analyzed through thematic analysis combining deductive coding, informed by prior research on emerging technologies in journalism (e.g., Westlund & Ferrer-Conill, 2018; Bradshaw, 2023), with inductive coding that allowed for new insights to emerge. The coding framework was structured around four analytical dimensions:

- Motivations for adopting emerging technologies.
- Mechanisms of enhancing transparency and verification.
- Operational and ethical challenges in deployment.
- Perceived outcomes in terms of disinformation mitigation and audience trust.

Cross-case comparison was used to identify patterns, variations, and contradictions in technology adoption and its impact across the cases.

This study acknowledges limitations that may influence the interpretation and generalizability of findings. Due to confidentiality and proprietary concerns, access to internal communications, editorial decision-making processes, and quantitative audience analytics was limited. This constraint may affect the depth of understanding regarding newsroom deliberations and performance evaluations. Emerging technologies evolve rapidly, and their implementation levels vary across organisations and timeframes. This creates challenges in comparing case studies with different levels of maturity and in capturing the most recent innovations or discontinuations. Some sources, especially those published by the organisations themselves, may reflect a strategic or promotional bias. Although triangulation with third-party sources was employed, some degree of subjectivity and institutional branding could not be fully eliminated. The analysis focused on organisations with English-language or translated materials, potentially excluding equally innovative practices in regions with less accessible documentation or fewer international collaborations. While audience engagement and trust are central to the study, the research did not include surveys or interviews with audiences themselves, limiting insights into how users interpret and respond to technologically mediated news experiences. Despite these limitations, the case study methodology offers a strong interpretive lens for understanding technological integration in journalism and provides valuable groundwork for future mixed-methods or longitudinal research.

Research questions

1. To what extent have emerging technologies (AI, blockchain, immersive media) been integrated into news media organisations to address disinformation?
2. How do these technologies contribute to enhancing journalistic transparency, trust, and audience engagement in the context of disinformation?

Results/Findings

Case Study 1**:** Reuters – Integrating Artificial Intelligence through Lynx Insight

Reuters has been a pioneer in integrating artificial intelligence (AI) into its newsroom operations to enhance efficiency and mitigate disinformation. In 2017, it launched Lynx Insight, an AI-powered system that uses natural language processing (NLP) and machine learning to analyze large datasets, identify anomalies, and suggest newsworthy developments (Thomson Reuters). The platform supports journalists by automating the generation of data-heavy content, such as financial summaries and sports updates, allowing human reporters to focus on investigative journalism and high-impact stories. This human-machine collaboration model reflects a cybernetic newsroom approach, where AI is utilized as an assistant rather than a replacement. It increases reporting speed, reduces human error, and facilitates timely fact-checking—key factors in countering misinformation. Furthermore, the system's capacity to detect patterns and offer predictive insights contributes to early identification of false or manipulated narratives. Reuters' adoption of AI tools like Lynx Insight illustrates how algorithmic systems, when implemented with transparency and editorial oversight, can improve resilience against disinformation (Barceló-Ugarte, Pérez-Tornero, & Vila-Fumàs, 2021).

Case Study 2: The Washington Post – Combining AI Automation and Immersive Storytelling

The Washington Post employs both AI and immersive technologies in its newsroom innovation strategy. The organisation introduced Heliograf in 2016, a proprietary AI tool designed to generate news stories from structured data (The Washington Post, 2020). Initially used during the Olympics and U.S. elections, Heliograf automates routine reporting and enables broader content production without increasing labor costs. The AI system fills templates with relevant data, enhancing reporting timeliness while freeing journalists to engage in more complex editorial work. Simultaneously, The Washington Post has explored immersive storytelling, for example, through the VR piece "12 Seconds of Gunfire," which portrays a school shooting survivor's trauma (The Washington Post, 2019). Combining animation, original audio, and spatial storytelling, the project encourages empathy and human connection. Although production costs and scalability limit the integration of VR into daily news workflows, such experiments demonstrate the potential of immersive journalism to enhance emotional engagement and combat apathy toward social issues. The Washington Post thus exemplifies how AI and immersive media, when aligned ethically, contribute to journalism's mission of public trust restoration (Floridi et al., 2018).

Case Study 3: The New York Times – Experimentation with Immersive and Spatial Journalism

The New York Times has actively experimented with immersive technologies, particularly VR and augmented reality (AR), to foster audience engagement and immersive literacy. Its NYT VR app, launched in 2015 in partnership with Google Cardboard, delivered 360-degree videos like "The Displaced," depicting the lives of refugee children (The New York Times, 2015). These projects promoted empathy and participatory viewing, offering an alternative to detached news consumption. Classroom guides accompanying these immersive films facilitated integration into education, supporting immersive literacy (The New York Times, 2020). In recent years, the Times has shifted focus to spatial journalism, leveraging 3D mapping, AR overlays, and game engines to produce experimental formats (The New York Times R&D, 2023). Its Research & Development department has explored wearable MR headsets, spatial audio, and real-time environment scanning for journalistic purposes. These initiatives highlight the newsroom's commitment to innovation, even though most projects remain in pilot phases due to technical and financial constraints. The New York Times' immersive experimentation marks a development in experimental, multi-sensory journalistic practices as a multi-sensory, participatory practice capable of fostering trust and resisting manipulation (Steed et al., 2023).

Case Study 4: Al Jazeera – Empathy-Driven Immersive Journalism via AJ Contrast

Al Jazeera's immersive storytelling division, AJ Contrast, offers a model of fully integrated immersive journalism centered on social justice and narrative equity. Established in 2017, AJ Contrast produces VR and 360-degree video reports, particularly focusing on underrepresented voices. One of its hallmark productions, "Yemen's Skies of Terror," documents the devastation of war from the perspective of civilians, filmed by local journalists trained in immersive technologies (AJ Contrast). This bottom-up approach to immersive journalism aims to foster emotional engagement and contextual understanding of distant crises, challenging the desensitization caused by traditional reporting. By embedding immersive practices within its editorial processes and emphasizing co-creation, Al Jazeera fosters transparency, narrative agency, and public engagement. AJ Contrast's work demonstrates how immersive storytelling can counteract dehumanizing narratives and contribute to rebuilding public confidence in news narratives by inviting the viewer into a co-experienced reality (Greber, Vogler, & Eisenegger, 2023).

Case Study 5: Frontline in Focus XR – Immersive Journalism in Conflict Zones
Frontline in Focus XR employs extended reality (XR) tools to document humanitarian crises in hard-to-reach areas, particularly in Syria. The organisation produces immersive stories using VR, AR, and 360-degree video formats, with content generated by trained local reporters. Their narratives cover warzones, archaeological sites, and refugee experiences, enabling viewers to virtually step into conflict-affected environments (Frontline in Focus XR). These productions aim to bypass geopolitical barriers, foster emotional engagement, as observed in immersive media user feedback, and provide trustworthy, on-the-ground information in contexts vulnerable to disinformation. The initiative demonstrates how immersive journalism can serve as both a narrative corrective and a documentary tool in verifying and contextualizing stories from high-risk regions. It supports decentralized, transparent, and participatory news practices, making it a relevant example of immersive technology's application to uphold journalistic integrity and resilience in conflict reporting (Paino Ambrosio & Rodríguez Fidalgo, 2021).

Case Study	**Technology Used**	**Key Features**	**Goals & Outcomes**
Reuters	Artificial Intelligence (AI)	Use of *Lynx Insight* for data analysis, NLP, anomaly detection, and content suggestions.	Real-time fact-checking, automation of routine reports, early detection of misinformation.
The Washington Post	AI & Immersive Media	*Heliograf* for automated reporting; immersive VR story "12 Seconds of Gunfire."	Scalable content production; immersive storytelling for empathy and awareness.
The New York Times	VR, AR, Spatial Journalism	*NYT VR app*, 360° films like *The Displaced*; exploration of AR overlays and MR headsets.	Promote immersive literacy, participatory journalism, and emotional engagement.
Al Jazeera (AJ Contrast)	Immersive Media (VR/360°)	Empathy-driven VR from local perspectives (e.g., *Yemen's Skies of Terror*); co-creation with citizens.	Enhance narrative equity, rebuild trust, and humanize distant crises through immersive journalism.
Frontline in Focus XR	XR (VR, AR, 360°)	Immersive reports from Syria and conflict zones using locally trained reporters and extended reality.	Provide verified, immersive reporting in conflict regions;

			empower local voices and bypass disinformation.

Table 1: Case studies

Discussion

The findings of this study suggest that the integration of emerging technologies, namely artificial intelligence, blockchain, and immersive media—into journalistic practice has progressed in both scale and sophistication. While the adoption patterns vary across media organisations, these technologies have begun to influence how newsrooms address disinformation, particularly in verification and storytelling practices and rebuild trust in an increasingly fragmented and polarized information environment.

Artificial intelligence has proven to be the most widely adopted among the technologies explored. Tools such as Lynx Insight and Heliograf demonstrate how AI can enhance the speed, precision, and efficiency of news production. These systems assist journalists in parsing vast datasets, automating repetitive content, and identifying patterns that may indicate misinformation or emerging narratives. Importantly, their use has shifted from simple automation to collaborative intelligence, where human oversight ensures contextual accuracy and editorial integrity. The increased capacity for real-time verification and scalable content creation makes AI an indispensable tool in the fight against disinformation.

Immersive media technologies, though less ubiquitous, provide methods shown to foster audience engagement and emotional connection. Case studies like Al Jazeera's AJ Contrast and Frontline in Focus XR illustrate how VR and 360-degree video can enhance empathy, amplify marginalized voices, and deepen understanding of complex issues. Immersive storytelling invites audiences into the lived experiences of others, making the news feel personal and participatory. Such formats are particularly effective in contexts where conventional reporting may fail to generate public concern or interest. Despite challenges related to cost, scalability, and ethical considerations, immersive media has demonstrated potential to reshape journalistic narratives in select applications to counteract apathy and desensitization.

Blockchain's adoption remains limited but conceptually transformative. Its affordances, such as content verification, timestamping, and source authentication, align directly with the goals of journalistic transparency and credibility. While case studies indicated limited implementation, the technology has been proposed as a mechanism for enhancing trust, particularly through content verification, especially in environments susceptible to content manipulation. However, high infrastructural demands, energy concerns, and uneven digital literacy present practical obstacles that must be addressed before blockchain can be deployed at scale.

In terms of enhancing journalistic transparency, these technologies contribute through both process and presentation. AI enables clear, data-driven storytelling and fact-checking; immersive media promotes narrative transparency by embedding audiences within the story; and blockchain creates verifiable records of content origin and modification. Each technology, when used responsibly, adds a layer of accountability to news production and dissemination.

These technologies may contribute to conditions that support the restoration of audience trust through these innovations. The key lies in ethical implementation, ensuring AI systems are free of bias, immersive stories respect participants and audiences alike, and blockchain tools do not exclude users due to technical complexity. Transparency in development, inclusivity in design, and interdisciplinary collaboration are essential to mitigating risks.

Finally, audience engagement is evolving. These technologies shift the audience from passive consumers to active participants. AI-driven personalization, immersive co-experiences, and traceable content empower users to question, explore, and engage with information more critically and emotionally. As a result, journalism becomes more interactive, dialogic, and resilient.

Although media organisations primarily operate within national or regional contexts, the challenges they face, such as disinformation, algorithmic bias, and declining trust, are transnational in nature. Disinformation campaigns, AI-generated content, and the use of blockchain or immersive media are not confined by borders; they circulate through global platforms and impact societies worldwide. This mismatch creates a governance gap. Therefore, the advocacy for global frameworks is rooted in the need for shared ethical principles, technological standards, and interoperability across jurisdictions. These frameworks do not aim to homogenize media practices but rather to provide a common ground for ethical innovation, especially in technologies that transcend national boundaries. In doing so, they help ensure that emerging tools serve democratic purposes and journalistic integrity globally, while allowing local media to adapt within their own sociopolitical contexts.

Conclusion and Future Research Directions

This study concludes that emerging technologies, when ethically and strategically integrated, can serve as vital instruments in journalism's evolving toolkit to address the challenges of disinformation, audience disengagement, and institutional distrust. Rather than functioning in isolation, AI, immersive media, and blockchain offer complementary capabilities that, when combined, may significantly reinforce journalism's democratic mission.

Key insights from the case studies suggest that AI enables enhanced verification and productivity; Immersive media fosters emotional resonance and narrative authenticity; Blockchain strengthens content credibility and traceability. However, none of these technologies represent a panacea. Their successful implementation depends on transparent processes, institutional accountability, and inclusive access. Importantly, technology adoption must be accompanied by media literacy efforts, policy support, and cross-sector collaboration to prevent new forms of manipulation, marginalization, or technocratic gatekeeping.

Building on the findings and limitations of this study, several avenues merit further exploration. Future research should investigate how different audiences perceive, interpret, and emotionally respond to AI-generated, blockchain-authenticated, or immersive journalism. Surveys, interviews, and ethnographic studies could uncover trust dynamics, engagement patterns, and comprehension levels. Expanding the study to include non-English-speaking media organisations, particularly in the Global South, could reveal how cultural, economic, and infrastructural variables affect the adoption and perception of emerging technologies. There is a need for multi-year investigations into how immersive journalism or AI verification tools influence media trust, civic participation, or resilience to disinformation over time. Exploring co-design processes between journalists, technologists, and communities can illuminate best practices for creating inclusive, bias-mitigated, and culturally sensitive technological solutions in newsrooms. As disinformation increasingly flows through platforms (e.g., X, YouTube, TikTok), future studies should assess how AI and blockchain tools interface with platform algorithms and moderation policies, and what regulatory mechanisms could harmonize innovation with public interest. Future work should evaluate sustainable business models and funding strategies for deploying immersive or blockchain-based journalism, especially in underfunded media environments. Research could also focus on how journalism education programs incorporate emerging technologies and ethics, particularly among Generation Z journalism students, and whether immersive training enhances their critical awareness and professional readiness.

The convergence of emerging technologies and journalism presents a critical opportunity, but also a profound responsibility, to reimagine the values, practices, and power structures of news media in the digital era. By embracing innovation within ethical, participatory, and inclusive frameworks, journalism can not only survive but flourish as a resilient, transparent, and socially responsive institution in the face of global disinformation threats.

References

AJ Contrast. (n.d.). *About AJ Contrast*. https://ajcontrast.com/about

Anderson, C. W., Bell, E., & Shirky, C. (2015). *Post-industrial journalism: Adapting to the present*. Tow Center for Digital Journalism. https://academiccommons.columbia.edu/doi/10.7916/D8H99HF7/download

Barceló-Ugarte, J. M., Pérez-Tornero, J. M., & Vila-Fumàs, M. (2021). Artificial intelligence in journalism and its ethical implications. *Journalism & Mass Communication Quarterly, 98*(1), (135–154). https://doi.org/10.1177/1077699020947736

Bradshaw, P. (2023). *The data journalism handbook: Towards a critical data practice*. Amsterdam University Press.

Deuze, M., & Beckett, C. (2022). AI and the people: Journalism, automation, and critical consciousness. *Digital Journalism, 10*(9), (1913–1930). https://doi.org/10.1080/21670811.2022.2043676

Erdal, I. J. (2011). Coming to terms with convergence journalism: Cross-media as a theoretical and analytical concept. *Convergence, 17*(2), (213–223). https://doi.org/10.1177/1354856510394641

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V. & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*, (689–707). https://doi.org/10.1007/s11023-018-9482-5

Frontline in Focus XR. (n.d.). About us. https://frontlineinfocusxr.net/

Greber, M., Vogler, D., & Eisenegger, M. (2023). The worthwhileness of immersive journalism: Taking on an audience perspective. *Journalism Studies, 24*(3), (345–364). https://doi.org/10.1080/1461670X.2022.2159261

Kissinger, H., Schmidt, E., & Huttenlocher, D. (2021). *The age of AI and our human future*. Little, Brown and Company.

Kolodzy, J. (2012). *Convergence journalism: Writing and reporting across the news media*. Rowman & Littlefield.

Lee, J., Braud, T., Zhou, P., Wang, L., Xu, D., Lin, Z., & Hui, P. (2021). All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda. *arXiv preprint.* https://arxiv.org/abs/2110.05352

Liu, M. (2023). User-generated content and journalism: Opportunities and tensions. *Digital Media & Society, 1*(1), (288–293).

Niranjanamurthy, M., Nithya, B. N., & Jagannatha, S. (2019). Analysis of blockchain technology: Pros, cons and SWOT. *Cluster Computing, 22*(6), (14743–14757). https://doi.org/10.1007/s10586-018-2405-4

Pain Ambrosio, E., & Rodríguez Fidalgo, M. I. (2021). Immersive communication: Analysis of journalistic discourse in VR environments. *Journal of Media Research, 14*(3), (1–17). https://doi.org/10.24193/jmr.38.2

Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). Sage Publications.

Steed, A., Slater, M., Garau, M., & Sanchez-Vives, M. V. (2023). Immersive literacy and presence: Designing for human response in virtual environments. *Presence: Teleoperators and Virtual Environments, 32*(1), (1–17). https://doi.org/10.1162/pres_a_00375

Sun, Y., Hu, Y., & Wu, L. (2022). Application and ethical analysis of artificial intelligence in journalism. *AI & Society, 37*, (1–20). https://doi.org/10.1007/s00146-021-01178-2

Suh, A., & Prophet, J. (2018). The role of immersive technology in enhancing user experience. *Journal of Business Research, 85*, (77–84). https://doi.org/10.1016/j.jbusres.2017.12.056

The New York Times. (2015). *The Displaced*. https://www.nytimes.com/video/the-displaced

The New York Times. (2020). *Virtual reality curriculum guide*. https://www.nytimes.com/2020/10/29/learning/lesson-plans/virtual-reality-curriculum-guide

The New York Times R&D. (2023). *Exploring mixed reality tools for journalists*. https://rd.nytimes.com/projects/exploring-mixed-reality-tools-for-journalists

The Washington Post. (2019). *12 seconds of gunfire: The reality of school shootings*. https://www.washingtonpost.com/graphics/2019/local/12-seconds-of-gunfire/

The Washington Post. (2020). *Heliograf: AI-powered election coverage*. https://www.washingtonpost.com/pr/2020/10/13/washington-post-debut-ai-powered-audio-updates-2020-election-results/

Thomson Reuters. (n.d.). Reuters brings the human touch to AI. https://www.thomsonreuters.com/en/artificial-intelligence/thomson-reuters-brings-the-human-touch-to-artificial-intelligence.html

Westlund, O., & Ferrer-Conill, R. (2018). *Critical perspectives on algorithmic journalism*. Taylor & Francis.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

The News Weight Coefficient (NeWC) for Measuring the Accuracy of Discrimination Between False and True News with Human and Non-Human

Teodor Răileanu-Olariu[1], Bogdan Oprea[2]

1. Doctoral School of Psychology, Faculty of Psychology and Educational Sciences („Al.I.Cuza" University – Iasi, Romania), teodor.olariu@uaic.ro

2. Head of the Journalism Department, Senior Lecturer, Faculty of Journalism and Communication Sciences (University of Bucharest, Romania), bogdan.oprea@unibuc.ro

Abstract

One of the major challenges in studying the fake news phenomenon in social science research is the selection of equivalent false and true news for academic research purposes. This study proposes a standardized, verifiable, and replicable selection procedure for true and false news using recognized media institutions and fact-checkers as sources. We have developed a calculation algorithm for a coefficient designed to reflect the importance, impact, and dissemination of fake news: the "News Weight Coefficient" (NeWC). Using the signal detection model, this study introduces a method for calculating performance indicators (accuracy, precision, and sensitivity/recall) in distinguishing between true and false news. To validate this, a group of participants (N=386) was asked to evaluate true and false news produced by human authors or generated by AI. The results indicate that participants performed better at identifying false news than true news, but they were more confident in their classifications when they considered a piece of news to be true. Respondents were slightly more sensitive to AI-generated content compared to human-produced content, displaying intuitive characteristics. These findings highlight the need for more intensive and effective media literacy efforts for the general public, alongside the refinement of prebunking methods using artificial intelligence.

Keywords: news weight, fake news, news identification accuracy, artificial intelligence

Introduction

In the post-truth era, where truth is relativized and opinions (or even outright falsehoods) are transformed into "alternative facts", a significant part of the population "lives in an epistemic space that has abandoned the conventional standards of evidence, internal consistency, and factual verification" (Lewandowsky, Ecker, & Cook, 2017). Fake news, if not properly identified, can shape public perceptions, influencing political choices, responses to crises, and trust in public institutions (Lewandowsky et al., 2017). Moreover, fake news is being *weaponized* in hybrid geostrategic conflicts (Allenby, 2017), a relevant example being the Romanian presidential elections in November 2024 (Kirby & Thorpe, 2024).

The academic literature identifies key limitations in the scientific evaluation of news veracity, particularly regarding the domains of news selection, their relevance to the public (or "newsworthiness" as defined by Vasas, 2009), and the structural equivalence between false and true news in terms of topic, subject matter, formulation/tone, source credibility, and internal consistency.

Compounding this situation is the explosive development of artificial intelligence, which introduces substantial challenges in identifying the authorship of news content (Loth, Kappes, & Pahl, 2024). Several studies on the development of automated fake news detection tools have shown that individuals have greater difficulty identifying human-written fake news compared to AI-generated ones (Galli, Masciari, Moscato, & Sperli, 2022). Human narrative language, unlike that of artificial intelligence, is often more complex, subtle, and refined—intentionally designed to manipulate readers' emotions and beliefs (Mishra & Sadia, 2023). Moreover, subtle disinformation techniques such as embedding elements of truth or manipulating context to enhance perceived credibility are far more sophisticated when employed by human authors (Oprea, 2021).

Therefore, the main goal of this research is to control for these limitations and to offer a replicable and accurate method for calculating the equivalence of fake news. Subsequently, we tested its applicability in measuring the accuracy of discrimination between true and false news, authored either by humans or by AI.

Methodology

A sample of Romanian respondents (N = 386; 65.5% women, 34.5% men) completed an online questionnaire evaluating the veracity of sixteen selected news articles, according to the following procedure.

News Selection

The procedure for selecting true and false news (N = 16) for evaluation by respondents involved two distinct sources:
(A) true and false news stories that had already appeared in mainstream media, later evaluated by experts to select the most relevant for the public;
(B) news articles (true and false) generated using artificial intelligence, crafted according to specific conventions of journalism and communication sciences.

To ensure control over *tone* and *journalistic style*, all news items adhered to a minimum standard of journalistic relevance (including a *headline*, an *introductory paragraph*, and a *concluding paragraph* – see Coman, 2009). The use of headlines and lead paragraphs has been successfully employed in previous fake news detection studies (Bago, Rand, & Pennycook, 2020; Allen et al., 2021).

Human-Authored News Selection

A. False news items were selected from recognized fact-checking platforms (*Veridica*, *Factual*, and *AFP*) for the period January–July 2024, corresponding to four distinct domains (*health*, *economy*, *geopolitics*, and *history/nationalism*). The selection process followed these steps:

I. The 29 selected false news items from the time interval were sent to four experts in communication, journalism, and misinformation studies. These experts evaluated the fake news stories according to three criteria:

1. Negative Intentionality – operationalized later in accordance with journalism and communication theory as *the degree of intent to cause harm through the formulation and public dissemination of the false story* (Oprea, 2021).
2. Plausibility – *the potential believability* (and thus *public acceptance*) of the fake news.
3. Spatial-Territorial Relevance – *the potential spread* of the fake news within a specific region (*local*, *regional*, *national*, or *European*), a crucial dimension of *newsworthiness* (Vasas, 2009).

II. The specialists rated each fake news story using a four-point Likert scale with no neutral midpoint ("minimal impact," "some impact," "strong impact," and "MAXIMUM impact").

III. Calculation of the Fake News Weight Coefficient (NeW-C – News Weight Coefficient):

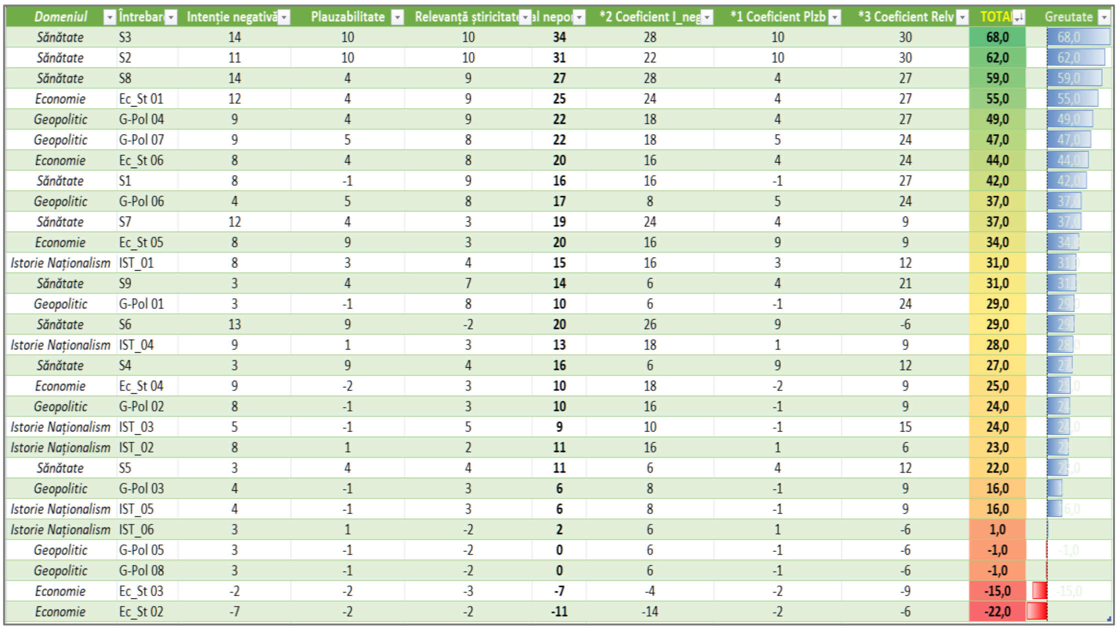
1. The impact levels rated on the Likert scale were scored post-evaluation:
    - Minimal impact = 1 point
    - Some impact = 2 points
    - Strong impact = 3 points
    - Maximum impact = 4 points
2. Combined impact (*I_comb*) was calculated using the formula:
    I_comb = (I_max + I_strong) – (I_some + I_min)
3. Since the criteria differ in relative importance (Oprea, 2021), in order to differentiate between news stories that might receive the same total score but differ in subcategory weights, a *multiplicative effectiveness coefficient* (M_{eff}) was applied:
    - Spatial Relevance: ×3
    - Negative Intentionality: ×2
    - Plausibility: ×1

The combined impact was then multiplied by the corresponding M_{eff}: (I_comb * Meff)

4. The final News Weight Coefficient (NeW-C) was calculated by summing the weighted scores for each criterion:

NeW-C = Σ (I comb*Meff)

IV. One news story with the highest coefficient was selected from each category (N = 4) (see table 1 below).



Domeniul	Întrebar	Intenție negativă	Plauzabilitate	Relevanță știricitat	al nepo	*2 Coeficient I_neg	*1 Coeficient Plzb	*3 Coeficient Relv	TOTA	Greutate
Sănătate	S3	14	10	10	**34**	28	10	30	**68,0**	68,0
Sănătate	S2	11	10	10	**31**	22	10	30	**62,0**	62,0
Sănătate	S8	14	4	9	**27**	28	4	27	**59,0**	59,0
Economie	Ec_St 01	12	4	9	**25**	24	4	27	**55,0**	55,0
Geopolitic	G-Pol 04	9	4	9	**22**	18	4	27	**49,0**	49,0
Geopolitic	G-Pol 07	9	5	8	**22**	18	5	24	**47,0**	47,0
Economie	Ec_St 06	8	4	8	**20**	16	4	24	**44,0**	44,0
Sănătate	S1	8	-1	9	**16**	16	-1	27	**42,0**	42,0
Geopolitic	G-Pol 06	4	5	8	**17**	8	5	24	**37,0**	37,0
Sănătate	S7	12	4	3	**19**	24	4	9	**37,0**	37,0
Economie	Ec_St 05	8	9	3	**20**	16	9	9	**34,0**	34,0
Istorie Naționalism	IST_01	8	3	4	**15**	16	3	12	**31,0**	31,0
Sănătate	S9	3	4	7	**14**	6	4	21	**31,0**	31,0
Geopolitic	G-Pol 01	3	-1	8	**10**	6	-1	24	**29,0**	29,0
Sănătate	S6	13	9	-2	**20**	26	9	-6	**29,0**	29,0
Istorie Naționalism	IST_04	9	1	3	**13**	18	1	9	**28,0**	28,0
Sănătate	S4	3	9	4	**16**	6	9	12	**27,0**	27,0
Economie	Ec_St 04	9	-2	3	**10**	18	-2	9	**25,0**	25,0
Geopolitic	G-Pol 02	8	-1	3	**10**	16	-1	9	**24,0**	24,0
Istorie Naționalism	IST_03	5	-1	5	**9**	10	-1	15	**24,0**	24,0
Istorie Naționalism	IST_02	8	1	2	**11**	16	1	6	**23,0**	23,0
Sănătate	S5	3	4	4	**11**	6	4	12	**22,0**	22,0
Geopolitic	G-Pol 03	4	-1	3	**6**	8	-1	9	**16,0**	16,0
Istorie Naționalism	IST_05	4	-1	3	**6**	8	-1	9	**16,0**	16,0
Istorie Naționalism	IST_06	3	1	-2	**2**	6	1	-6	**1,0**	
Geopolitic	G-Pol 05	3	-1	-2	**0**	6	-1	-6	**-1,0**	-1,0
Geopolitic	G-Pol 08	3	-1	-2	**0**	6	-1	-6	**-1,0**	
Economie	Ec_St 03	-2	-2	-3	**-7**	-4	-2	-9	**-15,0**	-15,0
Economie	Ec_St 02	-7	-2	-2	**-11**	-14	-2	-6	**-22,0**	

Table 1: NeW-C calculation formulas

<u>B. True news stories from fact-checkers</u> were selected in a 'mirrored' manner, real news articles, verified as accurate, were identified from the same sources and period (January–July 2024), within the same domain, and addressing similar topics.

AI-Generated News Selection (True/False)

To create true and false news articles using artificial intelligence, the ChatGPT-4o model (designed for complex tasks) was employed and instructed to generate news items (true/false) in a phased approach:

1. Compilation of a list of the most common fake news and conspiracy theories from the past two years (restricted to the fields of health, Romanian history, economy, and geopolitics).
2. Generation of false news – the AI model was prompted to create fake news articles that adhered to journalistic standards (title, opening paragraph, concluding paragraph), contained a malevolent intent (disinformation), and were based on the compiled list above.
3. Rewriting of the text to be more persuasive and stylistically aligned with recognized mainstream publications, to refine the tone and style of the fake news, making them appear more authentic (cf. Oprea, 2021).
4. Explanation prompt: “Why is this news story false?”
5. Generation of mirrored true news stories, maintaining consistency in domain, tone, journalistic style, and topic.

Prior to inclusion in the online questionnaire, all news stories were randomized using an online randomization tool (https://www.random.org/lists/).

Procedure for Calculating Accuracy and Other Indicators

To compute accuracy and other relevant statistical indicators, we applied *Signal Detection Theory* (Wickens, 2020), which provides a mathematical framework for describing and analysing decision-making under conditions of uncertainty. This model involves constructing a confusion matrix that visualizes the performance of a classification system by presenting the number of correct and incorrect predictions.

	Prediction TRUE NEWS	*Prediction FALSE NEWS*
TRUE News	(TP) 1^+	(FN) 0^-
FALSE News	(FP) 0^+	(TN) 1^-

Table 2: Confusion matrix

The matrix includes the following components:

- *True Positive (TP)*: True news correctly classified as true.
- *True Negative (TN)*: Fake news correctly classified as false.
- *False Positive (FP)*: Fake news incorrectly classified as true.
- *False Negative (FN):* True news incorrectly classified as false.

The indicators used in analysing news evaluation were accuracy, precision, sensitivity (recall) and the F1-Score [17]:

Accuracy is a general measure of classifier performance, indicating the proportion of correct predictions among the total. It reflects how often the classifier makes correct predictions and is particularly useful when classes are balanced.

$$\text{ACC} = \frac{Total\ correct\ predictions}{Total\ predictions} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision indicates how many of the news articles classified as true are actually true. It is the proportion of correctly classified positive instances among all instances classified as positive. This metric is crucial when the cost of false positives is high (e.g., when classifying fake news as true, which contributes to misinformation)

$$PR = \frac{TP}{(TP + FP)}$$

Sensitivity (or *recall*) indicates the model's ability to correctly identify all actual positive cases. It is calculated as the proportion of actual positives correctly identified. This measure is critical when it is essential to capture all relevant instances. In the context of news evaluation, high recall ensures that most true news stories are detected.

---

[17] Since we had an equal number of false and true news items, there was no need to calculate the F1-Score. The remaining three indicators were abbreviated as follows: APS

$$SENS = \frac{TP}{(TP + FN)}$$

To deepen the analysis, we reversed the perspective and treated fake news as the positive class. Consequently, the formulas for performance indicators were modified as follows[18]:

$$PR(False) = \frac{TN}{(TN) + (FN)}$$

$$SENS(False) = \frac{TN}{(TN) + (FP)}$$

$$F1(False) = 2 * \frac{Precision\ (False) * Sensitivity\ (False)}{Precision\ (False) + Sensitivity(False)}$$

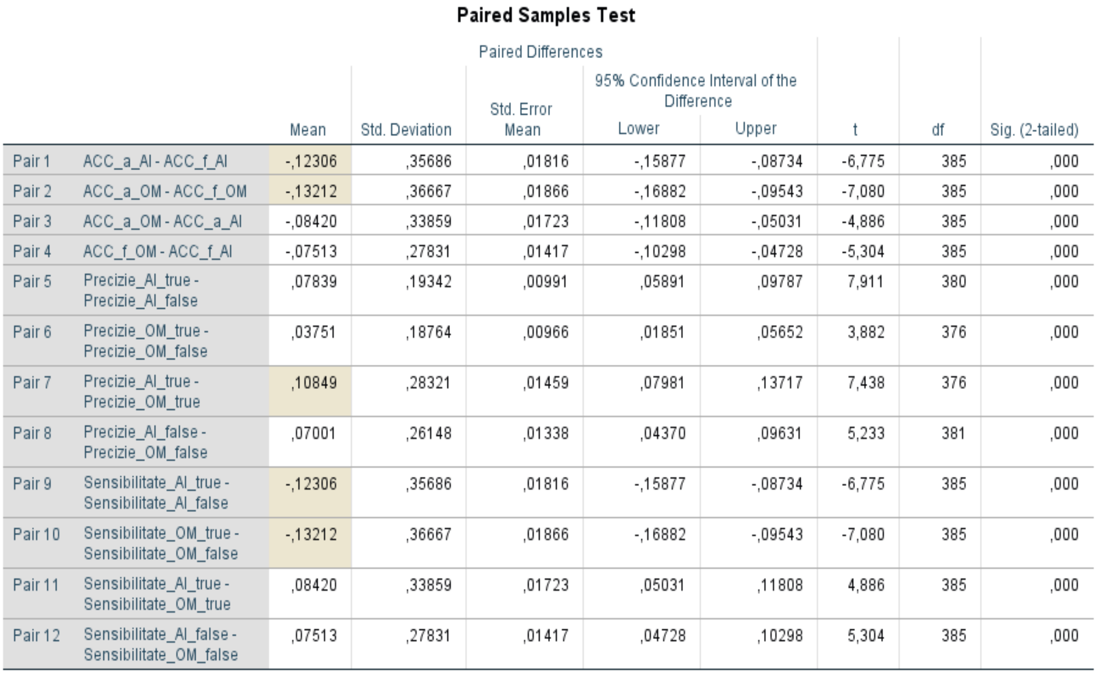
Results

Hypothesis: *"Individuals are more accurate in discriminating fake news produced by artificial intelligence than those produced by human authors."*

The results of the paired-samples t-test analysis revealed statistically significant differences in both the accuracy and precision of identifying true and false news, depending on whether the content was generated by human authors or artificial intelligence. More specifically, respondents demonstrated a higher ability to accurately distinguish between news content generated by humans versus that generated by AI (see Table 3):

---

[18] The calculation of accuracy remains the same regardless of which class is considered positive (true news or false news). Accuracy is an overall measure that indicates the total proportion of correct predictions (both positive and negative) out of all predictions

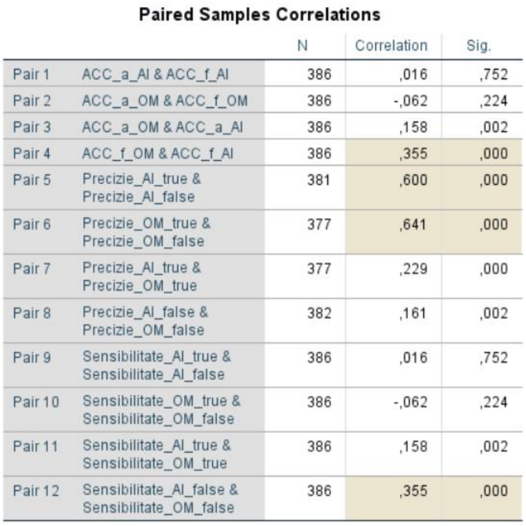


Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	ACC_a_AI - ACC_f_AI	-,12306	,35686	,01816	-,15877	-,08734	-6,775	385	,000
Pair 2	ACC_a_OM - ACC_f_OM	-,13212	,36667	,01866	-,16882	-,09543	-7,080	385	,000
Pair 3	ACC_a_OM - ACC_a_AI	-,08420	,33859	,01723	-,11808	-,05031	-4,886	385	,000
Pair 4	ACC_f_OM - ACC_f_AI	-,07513	,27831	,01417	-,10298	-,04728	-5,304	385	,000
Pair 5	Precizie_AI_true - Precizie_AI_false	,07839	,19342	,00991	,05891	,09787	7,911	380	,000
Pair 6	Precizie_OM_true - Precizie_OM_false	,03751	,18764	,00966	,01851	,05652	3,882	376	,000
Pair 7	Precizie_AI_true - Precizie_OM_true	,10849	,28321	,01459	,07981	,13717	7,438	376	,000
Pair 8	Precizie_AI_false - Precizie_OM_false	,07001	,26148	,01338	,04370	,09631	5,233	381	,000
Pair 9	Sensibilitate_AI_true - Sensibilitate_AI_false	-,12306	,35686	,01816	-,15877	-,08734	-6,775	385	,000
Pair 10	Sensibilitate_OM_true - Sensibilitate_OM_false	-,13212	,36667	,01866	-,16882	-,09543	-7,080	385	,000
Pair 11	Sensibilitate_AI_true - Sensibilitate_OM_true	,08420	,33859	,01723	,05031	,11808	4,886	385	,000
Pair 12	Sensibilitate_AI_false - Sensibilitate_OM_false	,07513	,27831	,01417	,04728	,10298	5,304	385	,000

Table 3: Paired Samples Test results

Respondents exhibit greater *accuracy* in identifying false news compared to true news, both in the case of AI-generated content (M_{difAI}= .123) and human-authored content ($M_{difHuman}$= .132) (see Table 3). Similarly, with respect to *sensitivity*, participants demonstrate a higher capacity to detect false news items than true ones. Notably, *participants are better at identifying true news generated by AI than those created by human authors*. The comparative analysis table shows that, overall, respondents exhibit slightly higher sensitivity to AI-generated news compared to human-authored news.



Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	ACC_a_AI & ACC_f_AI	386	,016	,752
Pair 2	ACC_a_OM & ACC_f_OM	386	-,062	,224
Pair 3	ACC_a_OM & ACC_a_AI	386	,158	,002
Pair 4	ACC_f_OM & ACC_f_AI	386	,355	,000
Pair 5	Precizie_AI_true & Precizie_AI_false	381	,600	,000
Pair 6	Precizie_OM_true & Precizie_OM_false	377	,641	,000
Pair 7	Precizie_AI_true & Precizie_OM_true	377	,229	,000
Pair 8	Precizie_AI_false & Precizie_OM_false	382	,161	,002
Pair 9	Sensibilitate_AI_true & Sensibilitate_AI_false	386	,016	,752
Pair 10	Sensibilitate_OM_true & Sensibilitate_OM_false	386	-,062	,224
Pair 11	Sensibilitate_AI_true & Sensibilitate_OM_true	386	,158	,002
Pair 12	Sensibilitate_AI_false & Sensibilitate_OM_false	386	,355	,000

Table 4: Paired Samples Correlations

The correlation table (see Table 4) reveals a moderate predictive capacity regarding how participants may reduce errors in identifying false news. For example, in approximately 40% of cases, the precision in identifying true news authored by humans significantly predicts a higher precision in identifying false news as well ($r = .64$, $p < .05$). A similar pattern emerges for AI-generated news: in roughly 36% of the population, improved precision in identifying true AI-generated news also leads to greater precision in identifying false news ($r = .60$, $p < .05$).

Conclusions / Discussion

Through the news selection procedure described above, we effectively controlled for several factors known to influence the perception of news veracity (as we find in academic journalism studies). These include:

- *News source* (Pehlivanoglu et al., 2021): All news items were presented as part of a scientific evaluation task, with no mention of any specific outlet or origin.
- *Authorship* (Fan, Wang, & Hu, 2023): The news articles were shown without naming an author; nor were there any cues indicating whether the text was created by a human or by AI.
- *Fact-checking behavior* (ibid.): A control question was introduced to verify whether respondents had previously searched for the news.
- *Consistency* (Kelly, 2019): Potential influence from prior knowledge or personal experience was relatively controlled through the balanced design of both true and false news generated by AI.
- *Level of detail and supporting evid*ence (Worden & Barthel, 2020): News items avoided including statistical data or verifiable sources.
- *Tone and sty*le (Wu & Hooi, 2023): These aspects were partially controlled, particularly in the case of AI-generated news.
- *Social feedback* (Worden & Barthel, 2020): All news items were evaluated individually, anonymously, and in an online setting.

  We consider that the NeW-C (News Weight Coefficient) can be used effectively and reliably in the selection of fake news for research in psychosocial sciences, journalism, and communication studies related to the fake news phenomenon.

Regarding the accurate discrimination between true and false news, the Signal Detection Theory model proved to be highly effective. The statistical analyses revealed that, overall, respondents exhibited slightly greater sensitivity to AI-generated news compared to human-authored news. A possible explanation for this outcome may lie in subtleties of language or depth of narrative, suggesting that certain linguistic and cognitive mechanisms enable participants to "sense" whether a news item, true or false, was generated by artificial intelligence. We may even poetically speculate that respondents are able to "sniff out" AI-generated news due to refined, subconscious processes of linguistic and psychological discrimination.

For approximately 40% of participants, improved precision in identifying true news authored by humans was significantly associated with greater precision in identifying false news. A similar pattern was observed in about 36% of participants with respect to AI-generated news, that is, an increase in precision in identifying true AI-generated news predicted a more accurate detection of false news.

This suggests that exposing individuals to high-quality true news, whether produced by traditional human authors or with the assistance of artificial intelligence, enhances the population’s capacity to accurately identify fake news, regardless of the authorship.

These conclusions align with contemporary prebunking (or inoculation) techniques for combating misinformation, whereby individuals are repeatedly exposed to overtly false (sometimes even exaggerated) examples of fake news in order to boost their capacity to detect maliciously crafted disinformation (Oprea, 2021). However, our findings suggest that prebunking may also be effective through repeated exposure to quality media content, without necessarily requiring examples of fake news. Clearly, this opens a promising avenue for further research and may yield significant practical implications.

The procedure described in the present study is innovative and it has the purpose of providing a scientific selection of news for social sciences research. By using human experts, we tried to ensure a professional assessment of the selected news in order to benefit from lifetime expertise and, also, from creativity of human experts. Once the procedure is validated, the AI capabilities of present time will offer a fast and reliable way of selecting the news for research given the precise and human revised prompts and commands. Of course, the development of such AI instruments will have to be reviewed in future study.

References

Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science advances, 7*(36). https://doi.org/10.1126/sciadv.abf4393

Allenby, B. R. (2017). The Age of Weaponized Narrative, or, Where Have You Gone, Walter Cronkite? *Issues in Science and Technology, 33*(4), 65-70. https://issues.org/the-age-of-weaponized-narrative-or-where-have-you-gone-walter-cronkite/?utm_source=chatgpt.com

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of experimental psychology: general*, *149*(8). https://doi.org/10.1037/xge0000729

Coman, M. (2009). *Manual de jurnalism*. Polirom.

Fan, W., Wang, Y., & Hu, H. (2023). Mimicking Human Verification Behavior for News Media Credibility Evaluation. *Applied Sciences, 13*(17). https://doi.org/10.3390/app13179553

Galli, A., Masciari, E., Moscato, V., & Sperlí, G. (2022). A comprehensive Benchmark for fake news detection. *Journal of Intelligent Information Systems, 59*(1), (237-261). https://link.springer.com/article/10.1007/s10844-021-00646-9?utm_source=chatgpt.com

Kelly, D. (2019). Evaluating the News: (Mis)Perceptions of Objectivity and Credibility. *Political Behavior, 41*, (445–471). https://doi.org/10.1007/s11109-018-9458-4

Kirby, P., & Thorpe, N. (2024). Romania's cancelled presidential election and why it matters. *BBC News*. https://www.bbc.com/news/articles/cx2yl2zxrq1o

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the “post-truth” era. *Journal of Applied Research in Memory and Cognition,* 6(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Loth, A., Kappes, M., & Pahl, M. O. (2024). Blessing or curse? A survey on the Impact of Generative AI on Fake News. *arXiv preprint.*

Mishra, A., & Sadia, H. (2023). A Comprehensive Analysis of Fake News Detection Models: A Systematic Literature Review and Current Challenges. *Engineering Proceedings*, *59*(1). https://doi.org/10.3390/engproc2023059028

Oprea, B. (2021). *Alegerile prezidențiale din România în rețelele de socializare online: Informare, dezinformare, manipulare, fake news*. Universitatea din București. (Lucrare publicată ulterior la ed. Polirom, 2021 – Oprea, B. Fake news si dezinformare online - recunoaște și verifică).

Pehlivanoglu, D., Lin, T., Deceus, F. et al. (2021). The role of analytical reasoning and source credibility on the evaluation of real and fake full-length news articles. *Cognitive Research: Principles and Implications, 6*. https://doi.org/10.1186/s41235-021-00292-3

Vasas, Ferenc (2009). Cum nu se colectează știrile. In M. Coman (Ed.), *Manualul de jurnalism* (3rd rev. ed.). Polirom.

Worden, K., & Barthel, M. (2020). *Many Americans are unsure whether sources of news do their own reporting.* Pew Research Center. https://www.pewresearch.org/journalism/2020/02/11/many-americans-are-unsure-whether-sources-of-news-do-their-own-reporting/

Wu, J., & Hooi, B. (2023). Fake News in Sheep's Clothing: Robust Fake News Detection Against LLM-Empowered Style Attacks. *arXiv preprint.* https://arxiv.org/abs/2310.10830

From the theory of Critical Digital Literacy to Pedagogical Practice in HE: Combating Disinformation in the S.HI.E.L.D. vs Disinfo Project

Panagiota Samioti[1], Konstantinos Sipitanos[2], Mari-Liis Madisson[3], Andreas Ventsel[4], Dimitris Kotzinos[5], Georgios Terzis[6], Christos Gavalas[7]

1. University of Crete 2. University of Crete 3. University of Tartu 4. University of Tartu 5. Cergy Paris University 6. Brussels School of Governance 7. Athens Technology Center

Abstract

Disinformation poses a growing challenge to democratic societies, public trust, and professional knowledge, requiring educational responses that go beyond technical fact-checking or narrow media literacy approaches. This paper presents the design, implementation, and outcomes of the Erasmus+ S.HI.E.L.D. vs Disinfo project, an interdisciplinary initiative aimed at embedding Critical Digital Literacy (CDL) in higher education. Grounded in an action research paradigm, the project draws on policy analysis, literature review, and focus group interviews with professionals from education, journalism, ICT, medicine, and civil society across several European countries. A central outcome of the project is a Quality Assurance Matrix that conceptualizes CDL across three interconnected dimensions: understanding the sociopolitical, ethical, and technological context of disinformation; identifying and critically evaluating misleading content; and applying appropriate pedagogical methodologies. This framework guided the collaborative development and implementation of curricula, teaching scenarios, and workshops across partner higher education institutions. The resulting pedagogical practices and reflections were documented in an Anthology of Educational Scenarios, providing transferable insights for educators. To support sustainability and wider adoption, the project also developed an Educators' E-Toolkit and adapted the Truly Media platform into a learning-oriented environment that enables collaborative verification and critical reflection. Overall, the paper argues that S.HI.E.L.D. vs Disinfo offers a holistic and transferable model for strengthening Critical Digital Literacy in higher education and preparing future professionals to navigate and counter disinformation in complex digital environments.

Keywords: disinformation, critical digital literacy, higher education, action research, pedagogical frameworks

Introduction

In an increasingly complex digital information landscape, disinformation has become a multifaceted threat that undermines democracy, public trust, and knowledge integrity (Lewandowsky et al., 2023). Several experts and professionals are facing severe problems, as concerns the mis- and dis-information being spread daily, such as the information from medical experts and the news that concerns peace and security, climate change, or even religious intolerance. There is a need for Higher Education to prepare the future professionals to identify and tackle disinformation and prepare them beyond isolated fact-checking tools or media literacy slogans (Galan et al. 2021).

The Erasmus+ S.HI.E.L.D. vs Disinfo (Students' Higher Education critical digital Literacy Development against Disinformation) is a systematically organised educational response to disinformation grounded in critical thinking, interdisciplinarity, and contextual sensitivity. In this project, a consortium of five European Universities and two companies (coming from six European Union countries) collaborated to design, test, and disseminate a suite of educational tools and strategies grounded in Critical Digital Literacy (CDL).

The term Critical Digital Literacy in this project refers to the ability to critically understand, evaluate, create, and interact with digital content, technologies, and platforms, with a focus on power structures, biases, and sociopolitical implications (Hinrichsen, J., & Coombs, 2014). Critical Digital Literacy Competencies are the corresponding skills, attitudes, and knowledge that are needed to be critical online. They involve critical reflection, ethical judgement and active participation (European Commission, 2021; Pangrazio, 2016).

The project followed a research-based, co-creative approach, based on the action research tradition (Carr and Kemmis, 1986), and finally produced four major intellectual outputs, all hosted on the project's website: https://shieldvsdisinfo.com/

I. A CDL Competence Framework and A Quality Assurance Matrix

This output provides a comprehensive framework for evaluating and developing educational interventions designed to empower students against disinformation. It is structured into three submatrices: (1) the context of disinformation, (2) the identification of disinformation, and (3) the appropriate pedagogical methodologies. The framework is based on a needs analysis that drew upon EU and national policy texts, as well as interviews with professionals from diverse fields (educators, semioticians, journalists, ICT experts, and medical practitioners). These sources were analysed and compared to identify the critical digital competencies, which in turn informed the quality assurance criteria represented across the three interrelated sub-matrices.

II. A Curriculum Implemented in HE institutions

Based on the above matrix, the partner institutions collaboratively designed and implemented a curriculum as well as teaching scenarios and workshops to enhance curriculum development. The scenarios were implemented in practice at the participating HEIs, and data were collected that provided feedback on the matrix and curriculum, integrating practice with theory.

III. An anthology of educational scenarios

Building on the pedagogical findings, an anthology was created, which included selected scenarios, teaching activities, and students' responses from questionnaires, as well as teachers' entries from their logs. This anthology offers an easy-access depiction of the entire process, but most importantly, it can give clever ideas to all HE faculty who want to teach their students (and future professionals) to resist disinformation through developing their CDL.

IV. Educator's E-toolkit

Designed for scalability and practical implications, a teacher's easy-access and interactive e-toolkit was constructed. This toolkit provides guidelines, relevant tools, and teaching activities, and helps educators to design and evaluate CDL-oriented educational activities in their classes.

By integrating these outputs, S.HI.E.L.D. vs Disinfo offers a holistic and transferable model for combating disinformation through education, basically through developing students' CDL. It bridges theory and practice, academic disciplines and professional fields, and local contexts with the European priority to tackle disinformation.

This paper presents the full journey of the S.HI.E.L.D. vs Disinfo project - from conceptual design and empirical research to its implementation and dissemination.

Needs Analysis

Despite existing literature, comprehensive guidelines for enhancing CDL in higher education institutions remain scarce, particularly those grounded in quality assurance criteria, interdisciplinary approaches, and alignment with labor market needs. The specific Project aims to conceptualize, structure, and propose pedagogical strategies to address these gaps, focusing on CDL at the tertiary level by integrating research-based theory with educational practice. Thus, it adopts an action research paradigm (Stringer & Aragon, 2021), engaging researchers as active participants in addressing real-world challenges and fostering meaningful change.

Our research methodology followed a structured three-phase process, coordinated by the WP2 leader, University of Tartu. The first phase aimed to map key CDL reports at both local and international levels, providing insight into existing measures and strategies. The literature review focused on: (a) EU initiatives to counter disinformation; (b) the current state and recommendations in partner countries - Estonia, Poland, Belgium, Greece, and France - analyzing both achievements and ongoing challenges, as well as proposals to enhance media literacy. The second phase included focus group interviews, which were conducted to critically reflect on the reviewed reports from a practitioner's perspective. These sessions, held in all five partner countries, involved professionals from diverse sectors, such as education, media, government, military, and civil society. In total, 51 participants contributed, enabling us to assess the real-world applicability of CDL measures and identify context-specific concerns.

The reports and focus group interviews highlighted the causes and consequences of disinformation, including the erosion of public trust and increasing societal polarization. A consistent theme across both sources was the range of EU-driven responses, such as media literacy initiatives, digital tools, and targeted projects, aimed at combating disinformation. However, the effectiveness of these measures was frequently questioned. Key limitations included the lack of systematic integration of media literacy into educational curricula, a narrowly defined understanding of media literacy and disinformation, and insufficient attention to contextual factors such as demographic, cultural, linguistic, and generational differences.

Conventional media literacy education often prioritizes technical and cognitive skills, overlooking emotional and psychological dimensions that shape information processing and susceptibility to misinformation. Both reports and focus group participants emphasized the need for more holistic approaches. Although various strategies have been proposed, their implementation remains uneven, hindered by the absence of standardized metrics for cross-national comparison. In response, the third phase of our research translated earlier findings into a matrix of quality criteria for course design and pedagogy, aiming to support a more integrated and effective approach to disinformation education.

Building on insights from the first two phases, the third phase advanced an integrated approach, culminating in the development of a matrix to support more effective teaching methods and curricula for higher education. This process was guided by UNESCO's *Global Standards for Media and Information Literacy Curricula Development Guidelines* (2022), which emphasize the structured distribution of learning outcomes and competencies.

The Matrix

The Quality Assurance Matrix: definition and description

Based on the Need Analysis, a Matrix with three submatrices was constructed. The Matrix is a tool that enhances the design, the evaluation, and the refinement of educational interventions, as well as a pre-curriculum material and an evaluation tool, since it is based on the competences identified in the needs analysis phase. More specifically, this hybrid educational and policy-making tool operates as a practical guide for curriculum development, as a diagnostic tool for assessing learning outcomes, and a policy-aligned framework to support strategic educational planning across institutions.

Structure of the Matrix

The S.HI.E.L.D. vs Disinfo matrix consists of three interconnected submatrices, each addressing a distinct dimension that together form a developmental process in educational practice: (1) understanding the context in which disinformation arises (Table 1), (2) identifying and analyzing disinformation itself (Table 2), and (3) applying appropriate pedagogical methodologies to address it (Table 3).

Each submatrix was represented in a table, where the vertical column lists the quality assurance criteria that any designed educational activity should meet, and the horizontal row lists the future professions of the students from the participating institutions. The number of bullet points displayed in each cell indicates the importance of a given quality assurance criterion for the corresponding professional group. A single bullet denotes that the criterion is important, while three bullets denote that the criterion is very important.

Submatrix 1 focuses on helping students understand the broader sociopolitical, ethical, and technological landscape in which disinformation mostly emerges. It includes criteria concerning the identification of motivations (QA1), analyzing media responsibilities (QA2, QA3), and relevant ethical issues (QA4).

Quality Assurance Criteria	**Educators**	**Journalists**	**Medical Professionals**	**ICT Experts**	**Citizens?**
QA1. Encourage group research on the motives behind disinformation	●●	●●	●●	●●	●●
QA2. Discuss media responsibilities and societal efforts to curtail disinformation	●●	●●	●	●●	●●
QA3. Discuss the relation between media freedom, democracy and disinformation prevalence	●●●	●●●	●	●●	●●
QA4. Discuss ethical issues (e.g., human rights, bias, transparency in data & decision-making)	●●●	●●●	●	●●●	●●
QA5. Understand the scientific process and treatment of scientific evidence	●●●	●●●	●●●	●●	●●
QA6. Discuss online disinformation diffusion (social media, bots, shadow users, etc.)	●●	●●●	●	●●●	●●●

Table 1: Submatrix 1: The Context of Disinformation

Submatrix 2 targets analytical and evaluative skills to detect and respond to disinformation. It promotes the skills connected with corresponding quality assurance criteria such as the differentiation between facts and opinions (QB1), the recognition of characteristics of disinformation (QB3), or the identification of disinformation sources (QB8).

Quality Criteria	**Educators**	**Journalists**	**Medical Professionals**	**ICTExperts**	**Citizens**
QB1: Differentiate facts vs. opinions	●●●	●●●	●●●	●●	●●
QB2: Types of deceptive content	●●●	●●●	●●	●●	●●
QB3: Characteristics of disinformation	●●●	●●●	●●●	●●●	●●●

Quality Criteria	**Educators**	**Journalists**	**Medical Professionals**	**ICTExperts**	**Citizens**
QB4: Judge validity of information	●●●	●●●	●●●	●●●	●●●
QB5: Use of digital tools	●●	●	●	●	●
QB6: Tech aspects of disinfo	●●	●●	●	●	●
QB7: Scientific process & evidence	●	●●●	●●●	●●●	●
QB8: Identify disinfo sources and spreading agents	●	●●●	●●	●●●	●●

Table 2: Submatrix 2: Identification of Disinformation

The final matrix addresses the teaching strategies and learning environments necessary for embedding CDL into Higher Education curricula. It includes, among other practices, the adoption of students' personal experiences in the development of educational material (QC3), the use of collaborative approaches (QC2), and the encouragement of civil activism (QC6).

Quality Assurance Criteria	**Educators**	**Journalists**	**Medical Professionals**	**ICT Experts**	**Citizens?**
QC1. Understand students' information sources and social media usage	●●●	●●●	●●●	●●●	●●●
QC2. Use collaborative methods in combating disinformation and foster community learning	●●●	●●	●●	●●	●●
QC3. Use personal experiences of disinformation as learning materials	●●●	●●●	●●●	●●●	●●●
QC4. Teach multiperspectivity through conflicting narratives and multiple sources	●●●	●●●	●●●	●●●	●●●
QC5. Clarify terms like disinformation, misinformation, fake news, urban myths	●●●	●●●	●	●	●
QC6. Advocate for civic activism and fast responses to misleading online content	●●	●	●	●	●

Table 3: Submatrix 3: Pedagogical Methodologies to tackle disinformation

In summary, the Quality Assurance Matrix developed within the S.HI.E.L.D. vs Disinfo project provides a structured and multidimensional framework for understanding, identifying, and addressing disinformation in higher education. By organising competencies across three interconnected submatrices - context, identification, and pedagogy - it offers educators and institutions a coherent pathway for designing effective educational interventions. This foundation not only supports the evaluation of existing practices but also informs the systematic development of new curricula. Building upon this framework, the next phase of the project focused on translating these insights into concrete educational materials and pedagogical implementations across partner institutions.

Curriculum Development, Collaborative Implementation & Anthology Creation

Building on the structure and insights provided by the Quality Assurance Matrix, the next phase of the S.HI.E.L.D. vs Disinfo project translated its conceptual framework into concrete educational practice. This involved the creation, implementation, and documentation of a robust set of teaching resources designed to foster CDL in higher education.

As the lead institution for Work Package 3 (WP3), the University of Crete (UoC) played a central role in shaping the project's pedagogical strategy. This section outlines the key components developed under UoC's coordination, including the General Curriculum, a set of adaptable sample curricula, the collaborative implementation process across partner institutions, and the creation of a comprehensive Anthology of Educational Scenarios. Together, these outputs operationalize the matrix's principles and offer practical, scalable tools for embedding CDL into diverse academic contexts.

Curriculum Development

UoC led the creation of the General Curriculum, which is organised around twelve key CDL topics, such as distinguishing fact from opinion, recognizing deceptive content, evaluating source credibility, and understanding the ethical implications of disinformation. Each topic includes learning objectives, suggested teaching methodologies, and adaptable educational activities suitable for different academic disciplines and educational levels.

To ensure the General Curriculum's practical application, UoC also developed three sample curricula, aligned with different course formats:

- A full-semester course (e.g. *Using AI to Combat Disinformation*),
- A short seminar (e.g. *Training Teachers as Informed Readers*),
- A practicum for student teachers (e.g. *How Can We Teach Young Children to Combat Disinformation?*).

These models offer detailed lesson plans, learning materials, assessment methods, and teaching strategies, and serve as adaptable blueprints for educators across Europe.

Collaborative Implementation

After the creation of the general curriculum and sample curricula, implementation was undertaken by all partner universities. Each adapted the common framework to their own institutional, disciplinary, and national contexts.

Examples include:

- Vrije Universiteit Brussel implemented a long-format course on disinformation for students in political science and media studies.
- The Medical University of Gdańsk delivered a course titled *Fake News and Medical Conspiracy Theories*, tailored for health science students.

- University of Tartu offered courses in semiotics and strategic communication, drawing on expertise in media studies.
- Cergy Paris Université focused on *Data Protection and Digital Tools*, combining digital ethics with legal and civic awareness.

All implementations reflected a commitment to active, interdisciplinary learning. Teaching strategies included media analysis, group discussion, hands-on engagement with digital tools (e.g. Hoaxy, InVID), and reflective journaling.

The Anthology: A Lasting Resource

UoC also coordinated the production of the project's Anthology, a key WP3 output. This collection presents detailed accounts of course implementations from each participating HEI, including course descriptions, selected activities, students' work, and reflections from engaged faculty members. Anthology serves as both a reflective and practical guide, allowing other HEIs to replicate, adapt, or be inspired by the project's diverse teaching approaches. It captures the collective work of the consortium while ensuring pedagogical coherence, academic depth, and wide applicability.

Through its leadership in WP3, the University of Crete not only shaped the intellectual framework of the S.HI.E.L.D. vs Disinfo project but also ensured that its outputs, that is, curricula and Anthology, remain impactful and usable well beyond the project's timeline. In doing so, UoC has contributed to building a critical, informed, and resilient academic community prepared to face the challenges of digital disinformation.

The Truly Media platform

As part of the S.HI.E.L.D. vs Disinfo project, Truly Media by ATC (a company based in Athens, Greece) was adapted from a professional fact-checking platform into a learning environment, tailored to the needs of journalism education. Originally built to support collaborative verification in real-time newsroom settings, Truly Media's core architecture -structured collections, shared workspaces, and verification workflows- was preserved. However, significant modifications were introduced to support teaching goals and learning outcomes.

The transformation was centered on the idea of bringing the newsroom into the classroom, creating a digital space where students could practice verification with professional tools, while being guided at the same time by academic structures. The most impactful change was the integration of two custom-designed verification checklists developed with university partners. These checklists brought pedagogical depth to the platform and helped students work through complex verification processes step by step, using open-ended prompts that support analytical thinking and media literacy.

The first checklist, *SPACE M CAT*, was co-designed with the Vrije Universiteit Brussel and reflects a rhetorical and contextual approach to source analysis. The acronym stands for Speaker, Purpose, Audience, Context, Exigence, Message, Choices, Appeals, Tone. Within Truly Media, each of these elements appears as an open-ended field, allowing students to reflect on the intent, construction, and delivery of the message under investigation. For example, students might analyze Speaker by researching the author or originator of a claim, and Exigence by exploring the urgency or motivation that may have triggered its publication. This approach is particularly effective in teaching critical reading and argument analysis, while still rooted in the needs of fact-checking practice.

The second checklist, called *Research Criteria*, was developed by the UoC and brings a more traditional information literacy framework into Truly Media. It includes the following fields: Accuracy / Documentation, Authority / Expertise of Author, Primary Source Proximity, Relevance, Purpose / Audience / Motive, Currency / Timeliness, Coverage / Thoroughness, Objectivity / Bias, Language / Persuasion Techniques, and Quality of the Argument. This checklist pushes students to evaluate the credibility of claims by assessing the source's qualifications, evidence base, and rhetorical style. It is especially useful when analyzing scientific misinformation, policy claims, or disinformation narratives where content may appear legitimate on the surface but fails under critical scrutiny.

Integrating these checklists into Truly Media require backend and interface redesigns. Verification sessions now include a dedicated space for students to complete either checklist, directly linked to the content being investigated. Their responses are saved in shared environments, enabling peer collaboration and instructor feedback. These adaptations transformed the platform into more than just a collaborative tool; it became a learning hub.

Additional features that were already part of Truly Media, such as manual content imports, task assignment action buttons, and reporting tools for each student output, made the platform easy to integrate into a classroom workflow and culture. The result is now a platform that prepares students not only to fact-check effectively, but to understand the deeper mechanics of how disinformation works and how it can be countered through thoughtful, evidence-based verification.

Educators across diverse sectors and educational levels face the challenge of integrating complex, interdisciplinary content into their teaching, especially when addressing critical topics like disinformation. To support this effort, S.HI.E.L.D. vs Disinfo project has developed an innovative, user-friendly resource known as the Ee-Toolkit. This guide is designed to help educators effectively incorporate newly developed knowledge, pedagogical strategies, and digital tools into their courses. By using practical teaching scenarios, adaptable materials, and actionable tips, the toolkit empowers lecturers who wish to embed essential topics such as misinformation and media literacy in their curricula.

Tackling disinformation requires a multifaceted and inclusive approach. Solutions must reach educators at all levels and be flexible enough to apply across disciplines and learning environments. The Ee-Toolkit recognizes this need by providing a comprehensive framework that incorporates foundational curriculum elements, such as learning objectives, teaching materials, methodologies, and student assessment tools, while reflecting the diverse experiences and expertise gained by the project's contributors. In doing so, the toolkit ensures that efforts to counter misinformation are thorough, cohesive, and relevant to today's educational landscape.

The Educator's E-Toolkit

Following the design of the Quality Assurance Matrix, the development of targeted curricula, and the collaborative implementation of educational interventions, the S.HI.E.L.D. vs Disinfo project culminated in the creation of the Educators' E-Toolkit, a forward-looking resource specifically designed to support educators in the ongoing integration of CDL into their teaching practice. Designed by the Vrije Universiteit Brussel team (leader of WP4) and created by all participating HEIs, which contributed with relevant material, it owes its appealing form to the creative team of CSI, our partner from Cyprus.

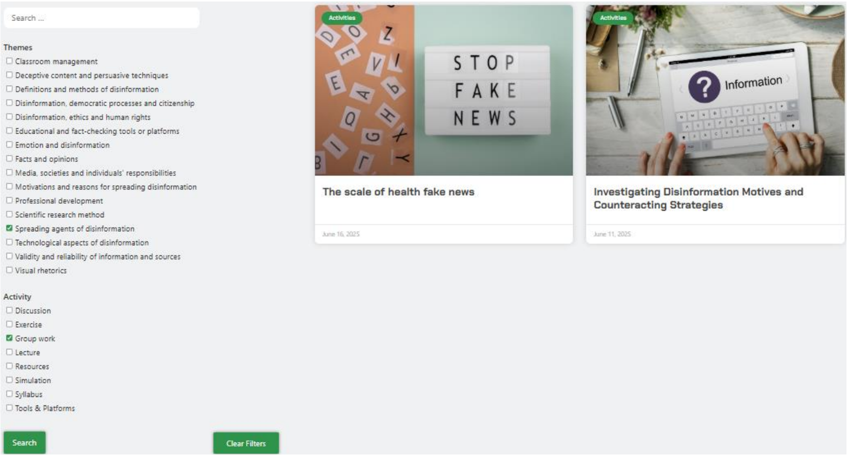
Unlike Anthology, which documented the scenarios, methods, and reflections developed and tested during the implementation stage, the E-Toolkit consists entirely of new educational materials that synthesize the project's key findings and pedagogical principles into adaptable guidance for future use. To further enhance the teaching experience, the toolkit includes supplemental resources that keep educators informed about the latest developments in media literacy and digital education. These resources incorporate interactive media and offer professional development opportunities, promoting continuous learning and adaptation.

More specifically, developed with scalability, interdisciplinary relevance, interactive features, and practical application in mind, the toolkit offers a suite of ready-to-use strategies, design tools, and implementation aids, aimed at faculty working across various disciplines and institutional contexts. It is designed to be modular and customizable, allowing educators to build their own courses, seminars, or lessons by combining suggested Themes and Activities that fit their classroom goals, student profiles, and subject area (Table 4):

Themes	**Activities**
Disinformation & democracy	Group work
Emotion & disinformation	Simulation
Facts vs. opinions	Lecture
Technological tools & platforms	Resources
Ethics & human rights	Discussion
Scientific research & evidence	Exercise
Visual rhetorics & persuasion	Syllabus
Classroom management	Tools & platforms

Table 4: E-Toolkit Builder: Theme × Activity Selector

The example below (Image 1) shows how the E-toolkit works:



Themes
☐ Classroom management
☐ Deceptive content and persuasive techniques
☐ Definitions and methods of disinformation
☐ Disinformation, democratic processes and citizenship
☐ Disinformation, ethics and human rights
☐ Educational and fact-checking tools or platforms
☐ Emotion and disinformation
☐ Facts and opinions
☐ Media, societies and individuals' responsibilities
☐ Motivations and reasons for spreading disinformation
☐ Professional development
☐ Scientific research method
☑ Spreading agents of disinformation
☐ Technological aspects of disinformation
☐ Validity and reliability of information and sources
☐ Visual rhetorics

Activity
☐ Discussion
☐ Exercise
☑ Group work
☐ Lecture
☐ Resources
☐ Simulation
☐ Syllabus
☐ Tools & Platforms

The scale of health fake news
June 16, 2025

Investigating Disinformation Motives and Counteracting Strategies
June 11, 2025

Image 1: S.HI.E.L.D. e-Toolkit Interface

The user has selected:

- Theme: *Spreading agents of disinformation*
- Activity type: *Group work*

Upon applying these filters, the system displays a curated set of educational resources that match the selected criteria. The search yields 2 content results:

1. *The Scale of Health Fake News* (June 16, 2025)
2. *Investigating Disinformation Motives and Counteracting Strategies* (June 11, 2025)

Each resource is presented as a card containing a title, publication date, visual preview, and categorization under “Activities.” These cards link to full teaching units or learning activities that faculty can adapt for use in their own classrooms.

This filtering mechanism reflects the toolkit’s modular and user-centered structure, enabling educators to build context-specific lessons or full curricula without needing to navigate extensive content manually. By combining content filters with pedagogical formats, the toolkit ensures relevance, accessibility, and ease of integration across educational settings and subject domains.

One of the key strengths of the E-Toolkit is its emphasis on cultivating student competencies that go beyond digital literacy. The materials support the development of Critical Digital Literacy skills, alongside vital social and interpersonal abilities. Educators using this toolkit help their students engage in active listening, practice assertiveness, and participate collaboratively through teamwork and perspective-taking. Additionally, there is an emphasis on intercultural communication, negotiation and mediation strategies, and intercultural sensitivity, preparing learners to navigate diverse and dynamic real-world contexts with confidence and empathy.

We believe that the attractive and interactive nature of the e-toolkit will appeal not only to faculty members from various universities, but also to trainers across different fields, such as adult educators, vocational education and training teachers, and secondary school teachers. In this way, the project will continue to have an impact well beyond its formal completion, influencing educational practice among those who choose to integrate CDL development into their teaching, with the aim of strengthening learners’ resilience against disinformation.

Conclusions

The S.HI.E.L.D. vs Disinfo project presents a comprehensive, interdisciplinary, and research-driven response to the growing threat of disinformation in today's digital society. By developing a robust framework grounded in Critical Digital Literacy, the project offers tangible solutions that transcend superficial fact-checking and embrace deeper cognitive, ethical, and sociocultural dimensions.

The integration of the Quality Assurance Matrix, curriculum development, educational scenarios, and digital tools like Truly Media collectively offers a scalable and replicable model for HEIs. These tools not only address the immediate challenges of disinformation but also prepare students to become critically engaged digital citizens who can navigate complex information ecosystems with discernment and responsibility.

Through the collaboration of multiple universities and institutions across Europe, the project validated the need for context-sensitive and professionally relevant approaches to CDL. The engagement with professionals from journalism, semiotics, ICT, medicine, and education during the needs analysis phase ensured that the developed outputs were not only theoretically sound but also practically viable.

The project's e-toolkit further enhances its impact by enabling educators to seamlessly embed CDL into existing curricula, encouraging collaborative learning, ethical reasoning, and civic participation. It recognizes the urgency of equipping future professionals with the skills to critically engage with digital content, understand power structures in contemporary digital societies, and actively counter the spread of disinformation.

In conclusion, S.HI.E.L.D. vs Disinfo has not only produced valuable educational resources but has also laid the groundwork for a sustainable transformation in how higher education addresses digital literacy. Its outcomes offer a blueprint for other educational institutions aiming to cultivate critical, informed, and resilient learners in an era defined by information overload and manipulation.

References

Carr, W. & Kemmis, S. (1986). *Becoming Critical. Education, Knowledge, and Action Research.* Routledge.

European Commission. (2021, April 21). The Digital Competence Framework (DigComp). Digital Skills and Jobs Platform. https://digital-skills-jobs.europa.eu/en/actions/european-initiatives/digital-competence-framework-digcomp

Galan, L. V., Osserman, J., Parker, S., & Taylor, M. (2021). Understanding and countering disinformation: The case for a cognitive, ethical and resilience-based approach in higher education*. Social Sciences*, *10*(8). https://doi.org/10.3390/socsci10080299

Hinrichsen, J., & Coombs, A. (2014). The five resources of critical digital literacy: A framework for curriculum integration. *Digital Culture & Education*, *6*(1), 1–16.

Lewandowsky, S., Cook, J., Fay, N., Lazer, D., & Ecker, U. K. H. (2023). The challenge of misinformation: A focus on science, trust, and democracy. *Nature Reviews Psychology, 2*(1), (14–28). https://doi.org/10.1038/s44159-022-00085-4

Pangrazio, L. (2016). Reconciling the personal with the critical: A challenge for digital literacy education. *Learning, Media and Technology*, *41*(1), (100–114). https://doi.org/10.1080/17439884.2015.1107101

Stringer, E. T., & Aragon, A. O. (2021). *Action Research* (5th ed.). SAGE.

CDL and Critical Creativity: Engaging Students with AI to Counter Disinformation in Higher Education

Panagiota Samioti[1], Rafail Giannadakis[2]
1. Laboratory and Teaching Staff, School of Philosophy, University of Crete, psamioti@uoc.gr
2. Research Assistant, TALOS ERA Chair in AI for SSH, University of Crete, giannadakis.uni@gmail.com

Abstract

This paper presents and reflects on a university-level course designed to foster Critical Digital Literacy (CDL) and strengthen students' resilience to disinformation through the integration of artificial intelligence (AI), digital tools, and creative expression. It outlines the course context and critically examines students' learning, creative engagement, and the pedagogical value of combining digital platforms with Critical Creativity. Drawing on classroom documentation and student reflections, the paper explores how interaction with AI technologies supported metacognitive awareness, enhanced source evaluation, and helped students recognize rhetorical manipulation in digital content. Offered at the University of Crete, as part of the Erasmus+ Project S.HI.E.L.D. vs Disinfo, the course attracted students from the humanities, social sciences, and STEM fields. It promoted CDL by incorporating tools such as ChatGPT and Truly Media, along with three custom-designed analytical tools, forming a comprehensive source evaluation toolkit. Students applied these during in-class activities, such as source evaluation, prompt crafting, and collaborative design, to assess the credibility of academic and social media content and produce informational leaflets. Findings suggest that students learned to recognize bias and AI-generated disinformation while developing responsible, critical use of digital tools as informed citizens.

Keywords: disinformation, critical digital literacy, critical creativity, higher education pedagogy, AI in education

Introduction

Disinformation spreads rapidly across today's digital landscape, affecting sectors like public health, science, and education (González-Pérez, 2020; Tomassi et al., 2025). Addressing these challenges requires evidence-based educational interventions. A recent review by Rau & Premo (2025) shows that programs in misinformation training, information literacy, and fraud awareness can significantly improve individuals' resistance to falsehoods. Furthermore, as Hobbs (2010) emphasized, media and digital literacy go beyond technical skills as they require critical thinking, language awareness, and ethical engagement with information. Hobbs also introduced dimensions such as “reflect” (ethical awareness and social responsibility) and “act” (collaborative problem-solving) (in Jacques et al., 2013).

Building on this broader view, *Critical Digital Literacy (CDL)* emerges as a more targeted framework for addressing disinformation. It involves critically accessing, interpreting, and assessing media content, empowering learners to make informed judgments (Buckingham, 2019; Leaning, 2017). It also includes competencies such as retrieving, analyzing, evaluating, and producing messages across platforms (Buckingham, 2005), all of which are vital for identifying disinformation and engaging with digital content responsibly. By cultivating these skills, educators and students are better equipped to confront misinformation and help safeguard democratic values.

In the S.HI.E.L.D. vs Disinfo Project, CDL is defined as a multidimensional and integrative concept combining media, news, and information literacies (Katsarou et al., under review). As the authors note, no single literacy is sufficient to address digital disinformation. The Project proposes four interconnected CDL dimensions:

- *Critical evaluation and complex skills*: Developing critical thinking, analysis, and communication.
- *Understanding bias and credibility*: Detecting misinformation and assessing source trustworthiness.
- *Ethical concerns*: Encouraging responsible, reflective information use and production.
- *Digital navigation*: Equipping learners to adapt to evolving digital environments.

This paper presents a higher education (HE) course that addresses these dimensions through AI integration, digital tools, and creative activities. While originally designed for students in the Faculty of Letters at the University of Crete, it attracted participants from diverse fields. The course included source evaluation, prompt crafting, and collaborative design projects. Co-authored by the instructor, Panagiota Samioti, and the external observer, Rafail Giannadakis, this paper outlines the course’s design and reflects on students’ learning through qualitative analysis of their activities and outputs. Ultimately, we argue that combining AI tools with critical and creative methods empowers students to navigate the digital public sphere as informed, responsible participants.

Course Context and Details

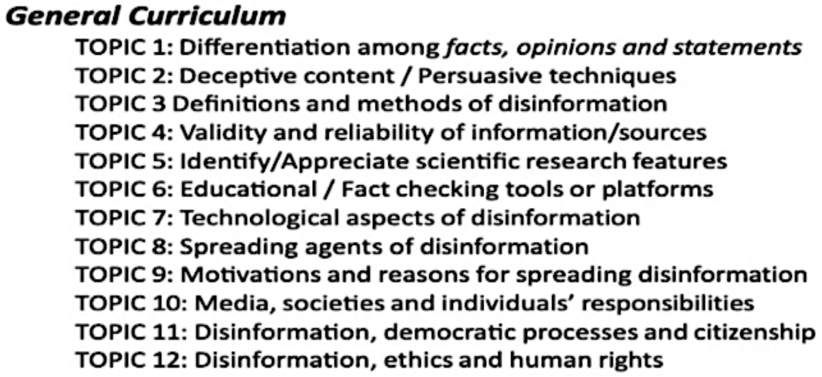
Course Implementation

The course *Academic Language Skills to Combat Disinformation* (Samioti, 2024) was offered during the 2023–2024 Spring Semester at the University of Crete as part of the Erasmus+ Project S.HI.E.L.D. vs Disinfo (Students of Higher Education Critical Digital Literacy Development against Disinformation). This European initiative (2022–2025) aims to develop Critical Digital Literacy (CDL) among university students, equipping them to recognize and respond to disinformation in complex digital environments.

Participating institutions include partners from France, Cyprus, Estonia, Belgium, Poland, and Greece. Key project outputs include a CDL competence framework, quality assurance matrix, general curriculum, scenario anthology, and an Educator's E-toolkit (S.HI.E.L.D. vs Disinfo, 2025b).

The University of Crete, the Greek partner, piloted this course as part of Work Package 3 (WP3), which focuses on translating research into actionable learning materials. WP3's General Curriculum, comprising twelve adaptable learning topics, guided the design of this intervention.

Designed and taught by Panagiota Samioti, with Rafail Giannadakis serving as an external observer, the course was originally intended for undergraduate students in the School of Philosophy. However, it also attracted participants from the humanities, social sciences, and STEM fields, resulting in a diverse cohort of 51 students. The hybrid course (on-site and online) combined academic language development with CDL training, emphasizing the role of language in shaping, resisting, or reproducing disinformation. The course consisted of 6 three-hour sessions, each aligned with targeted learning objectives drawn from the 12 adaptable topics in the S.HI.E.L.D. vs Disinfo General Curriculum (S.HI.E.L.D. vs Disinfo, 2025a; see Figure 1).



General Curriculum

TOPIC 1: Differentiation among *facts, opinions and statements*
TOPIC 2: Deceptive content / Persuasive techniques
TOPIC 3 Definitions and methods of disinformation
TOPIC 4: Validity and reliability of information/sources
TOPIC 5: Identify/Appreciate scientific research features
TOPIC 6: Educational / Fact checking tools or platforms
TOPIC 7: Technological aspects of disinformation
TOPIC 8: Spreading agents of disinformation
TOPIC 9: Motivations and reasons for spreading disinformation
TOPIC 10: Media, societies and individuals' responsibilities
TOPIC 11: Disinformation, democratic processes and citizenship
TOPIC 12: Disinformation, ethics and human rights

Figure 1: The S.HI.E.L.D. vs Disinfo *General Curriculum's 12 adaptable topics*

Course Objectives

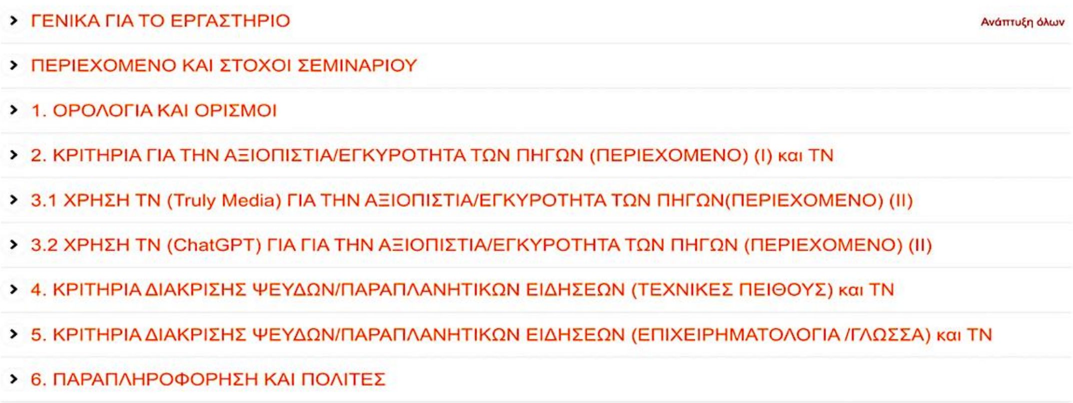
The course aimed to raise awareness of disinformation and equip students with tools to identify fake news features, assess credibility, and recognize rhetorical manipulation. A further goal was to reflect on the usefulness and limitations of AI in this context.

To achieve these aims, the course employed diverse pedagogical strategies, including lectures, Socratic dialogue, role-play, and collaborative group work. Students were engaged in inquiry- and project-based learning, analyzing authentic materials such as news articles, scientific publications, and social media posts. Topics included contested subjects like 5G technology and health risks. These were examined using the Truly Media platform, a collaborative verification environment that enabled students to annotate and evaluate digital content (Truly Media, n.d.). ChatGPT was also used to develop prompt-engineering skills and reflect on AI's evaluative capacities. Assessment combined formative (e.g., checklists, reflections) and summative tools (e.g., final projects, post-course questionnaires).

Course Content

The course followed a deliberately scaffolded three-part structure (Figure 2):

- Part 1 (session 1) introduced conceptual foundations.
- Part 2 (sessions 2-5) focused on applied analysis through a staged source-evaluation toolkit.
- Part 3 (session 6) emphasized reflection and civic action.



› ΓΕΝΙΚΑ ΓΙΑ ΤΟ ΕΡΓΑΣΤΗΡΙΟ

Ανάπτυξη όλων

› ΠΕΡΙΕΧΟΜΕΝΟ ΚΑΙ ΣΤΟΧΟΙ ΣΕΜΙΝΑΡΙΟΥ

› 1. ΟΡΟΛΟΓΙΑ ΚΑΙ ΟΡΙΣΜΟΙ

› 2. ΚΡΙΤΗΡΙΑ ΓΙΑ ΤΗΝ ΑΞΙΟΠΙΣΤΙΑ/ΕΓΚΥΡΟΤΗΤΑ ΤΩΝ ΠΗΓΩΝ (ΠΕΡΙΕΧΟΜΕΝΟ) (I) και ΤΝ

› 3.1 ΧΡΗΣΗ ΤΝ (Truly Media) ΓΙΑ ΤΗΝ ΑΞΙΟΠΙΣΤΙΑ/ΕΓΚΥΡΟΤΗΤΑ ΤΩΝ ΠΗΓΩΝ(ΠΕΡΙΕΧΟΜΕΝΟ) (II)

› 3.2 ΧΡΗΣΗ ΤΝ (ChatGPT) ΓΙΑ ΓΙΑ ΤΗΝ ΑΞΙΟΠΙΣΤΙΑ/ΕΓΚΥΡΟΤΗΤΑ ΤΩΝ ΠΗΓΩΝ (ΠΕΡΙΕΧΟΜΕΝΟ) (II)

› 4. ΚΡΙΤΗΡΙΑ ΔΙΑΚΡΙΣΗΣ ΨΕΥΔΩΝ/ΠΑΡΑΠΛΑΝΗΤΙΚΩΝ ΕΙΔΗΣΕΩΝ (ΤΕΧΝΙΚΕΣ ΠΕΙΘΟΥΣ) και ΤΝ

› 5. ΚΡΙΤΗΡΙΑ ΔΙΑΚΡΙΣΗΣ ΨΕΥΔΩΝ/ΠΑΡΑΠΛΑΝΗΤΙΚΩΝ ΕΙΔΗΣΕΩΝ (ΕΠΙΧΕΙΡΗΜΑΤΟΛΟΓΙΑ /ΓΛΩΣΣΑ) και ΤΝ

› 6. ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ ΚΑΙ ΠΟΛΙΤΕΣ

Figure 2: Overview of the course modules as presented on the Moodle platform

Part 1 (session 1): Conceptual foundations

Students compared scientific, policy-based, and AI-generated definitions of disinformation to establish shared terminology and identify differences in emphasis (Topic 3 of the S.HI.E.L.D. vs Disinfo *General Curriculum*).

Part 2 (sessions 2-5): Applied analysis and toolkit development

In Part 2, students developed the *CLEAR* source-evaluation toolkit (Figure 3), which integrates three complementary tools, each targeting a different layer of analysis within the evaluation process: source credibility (*PROSPECT*), rhetorical persuasion (*APPEAL*), and argument quality (*ELEVATE*):

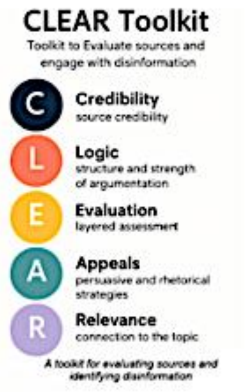
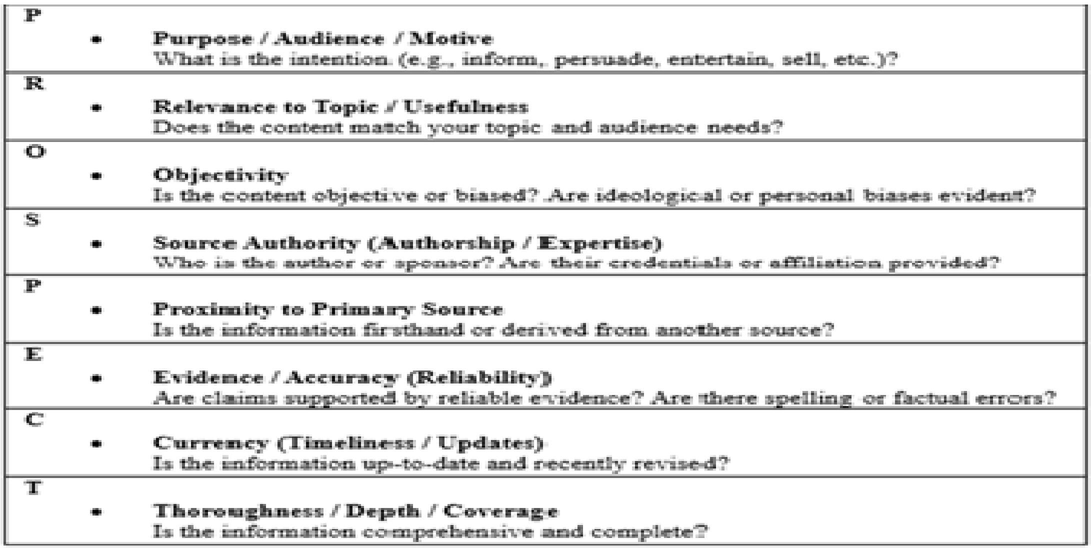

CLEAR Toolkit
Toolkit to Evaluate sources and engage with disinformation
C
Credibility
source credibility
L
Logic
structure and strength of argumentation
E
Evaluation
layered assessment
A
Appeals
persuasive and rhetorical strategies
R
Relevance
connection to the topic
A toolkit for evaluating sources and identifying disinformation

Figure 3: CLEAR Toolkit (Samioti, 2024)

Session 2-3: PROSPECT (source credibility)

Students used *PROSPECT* (*Protocol for Resource Objectivity, Source, Purpose, Evidence, Currency, and Thoroughness*) (Samioti, 2024; see Table 1) to evaluate authorship, objectivity, evidence quality, currency, and overall completeness. The tool builds on the *CRAAP* test (Blakeslee, 2004) by incorporating criteria such as proximity to primary sources. In session 3, students applied *PROSPECT* to selected digital texts using ChatGPT and the Truly Media collaborative verification environment.

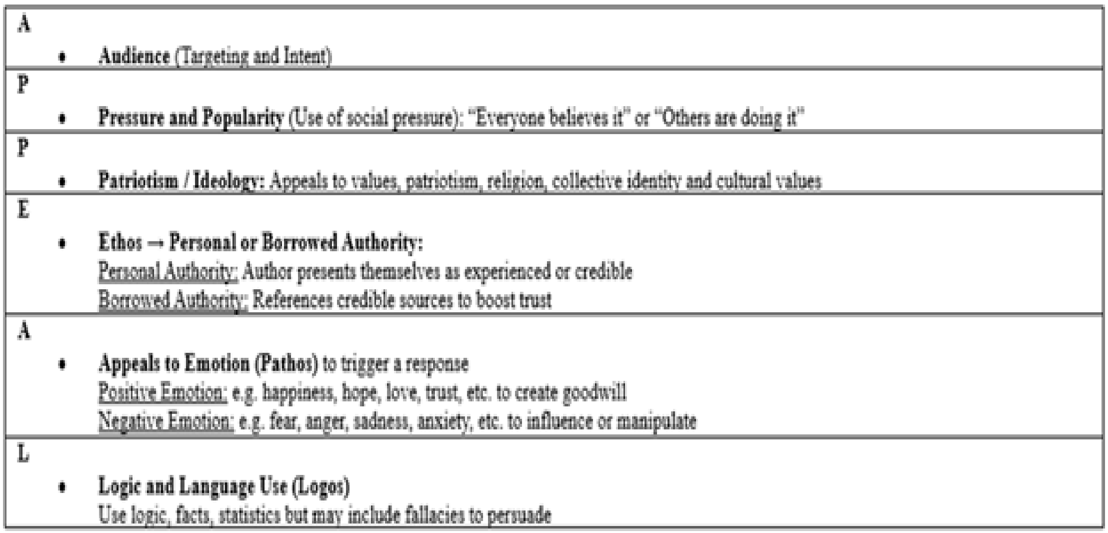


P	• **Purpose / Audience / Motive** What is the intention (e.g., inform, persuade, entertain, sell, etc.)?
R	• **Relevance to Topic / Usefulness** Does the content match your topic and audience needs?
O	• **Objectivity** Is the content objective or biased? Are ideological or personal biases evident?
S	• **Source Authority (Authorship / Expertise)** Who is the author or sponsor? Are their credentials or affiliation provided?
P	• **Proximity to Primary Source** Is the information firsthand or derived from another source?
E	• **Evidence / Accuracy (Reliability)** Are claims supported by reliable evidence? Are there spelling or factual errors?
C	• **Currency (Timeliness / Updates)** Is the information up-to-date and recently revised?
T	• **Thoroughness / Depth / Coverage** Is the information comprehensive and complete?

Table 1: PROSPECT tool

Session 4: APPEAL (rhetorical strategies)

Students next applied *APPEAL* (*Audience Targeting and Intent, Pressure and Popularity, Patriotism/Ideology, Ethos, Appeals to Emotion, and Logic and Language Use*) (Samioti, 2024; see Table 2) to identify rhetorical strategies and persuasive techniques in texts, including emotional appeals, ideological positioning, and credibility cues.

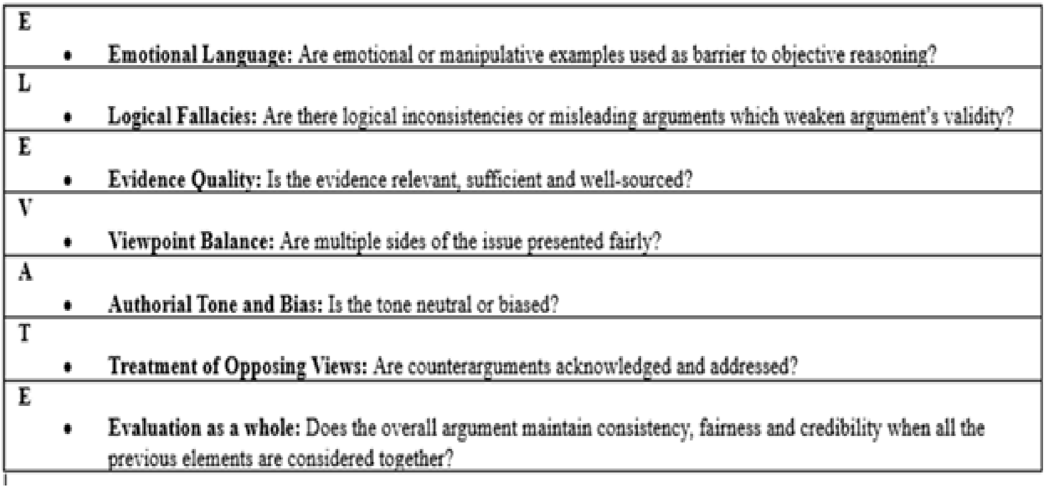


A	• **Audience** (Targeting and Intent)
P	• **Pressure and Popularity** (Use of social pressure): "Everyone believes it" or "Others are doing it"
P	• **Patriotism / Ideology:** Appeals to values, patriotism, religion, collective identity and cultural values
E	• **Ethos → Personal or Borrowed Authority:** Personal Authority: Author presents themselves as experienced or credible Borrowed Authority: References credible sources to boost trust
A	• **Appeals to Emotion (Pathos)** to trigger a response Positive Emotion: e.g. happiness, hope, love, trust, etc. to create goodwill Negative Emotion: e.g. fear, anger, sadness, anxiety, etc. to influence or manipulate
L	• **Logic and Language Use (Logos)** Use logic, facts, statistics but may include fallacies to persuade

Table 2: APPEAL tool

Session 5: ELEVATE (argument and evidence quality)

In this session, students used *ELEVATE* (*Emotional Language, Logical Fallacies, Evidence, Viewpoint Balance, Authorial Tone, Treatment of Opposition, and Evaluation as a Whole*) (Samioti, 2024; see Table 3) to assess argument structure, fallacies, balance, and how opposing views are represented.



E	• **Emotional Language:** Are emotional or manipulative examples used as barrier to objective reasoning?
L	• **Logical Fallacies:** Are there logical inconsistencies or misleading arguments which weaken argument's validity?
E	• **Evidence Quality:** Is the evidence relevant, sufficient and well-sourced?
V	• **Viewpoint Balance:** Are multiple sides of the issue presented fairly?
A	• **Authorial Tone and Bias:** Is the tone neutral or biased?
T	• **Treatment of Opposing Views:** Are counterarguments acknowledged and addressed?
E	• **Evaluation as a whole:** Does the overall argument maintain consistency, fairness and credibility when all the previous elements are considered together?

Table 3: ELEVATE tool

Although the three tools overlap in areas such as bias, evidence, and tone, each provides a distinct analytical lens. Used together as *CLEAR* (*Credibility, Logic, Evaluation, Appeals, Relevance*), they supported a layered evaluation process that guided students from source credibility to rhetorical analysis and ultimately to argumentation. Students compared their analyses during quizzes, peer discussions, and digital source assessments with AI-assisted outputs and reflected on where AI could support interpretation and where it was unreliable without critical human judgment. This staged progression fostered progressively deeper analytical engagement with digital texts while positioning the *CLEAR* toolkit as adaptable and open to refinement through action research in diverse educational contexts.

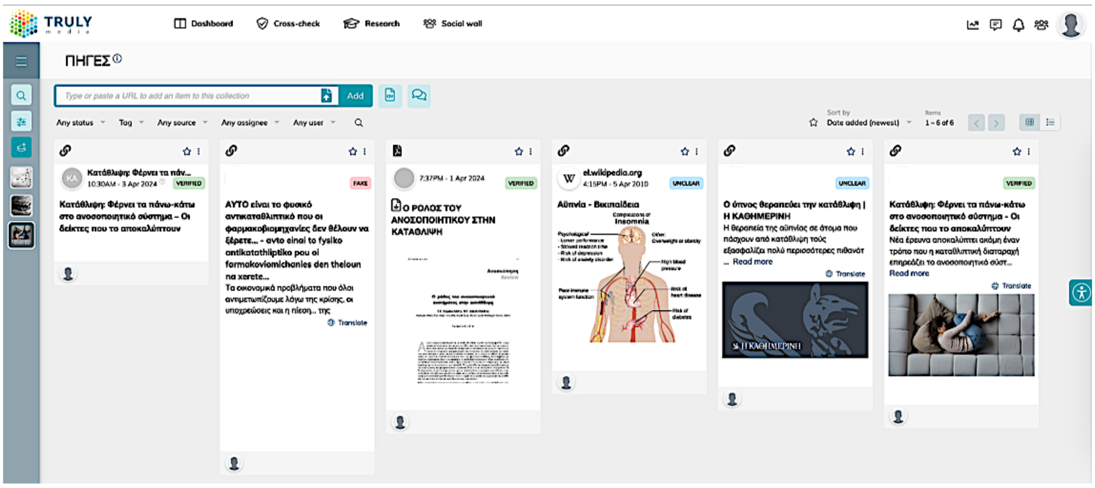
Part 3 (session 6): Reflection and civic action

In Part 3 (session 6), students synthesized their learning by exploring disinformation's societal impact and proposing civic responses. They collaboratively designed informational materials, such as leaflets or posters, applying evaluation criteria from the course. This final activity emphasized ethical responsibility and real-world engagement.

Sample Activities

To put these principles into practice, the course used scaffolded activities combining collaborative analysis and creative production, aligned with CDL goals.

First, students used the Truly Media platform (Truly Media, n.d.; see Figure 4) to collaboratively annotate scientific articles, news items, and social media posts. They applied the CLEAR Toolkit in group discussions and recorded their findings. This demonstrated that even neutral-seeming content can contain misleading elements, highlighting the importance of evidence-based scrutiny.



Κατάθλιψη: Φέρνει τα πάν...
10:30AM - 3 Apr 2024

Κατάθλιψη: Φέρνει τα πάνω-κάτω στο ανοσοποιητικό σύστημα – Οι δείκτες που το αποκαλύπτουν

ΑΥΤΟ είναι το φυσικό αντικαταθλιπτικό που οι φαρμακοβιομηχανίες δεν θέλουν να ξέρετε... - avto einai to fysiko antikatathliptiko pou oi farmakoviomichanies den theloun na xerete...
Τα οικονομικά προβλήματα που όλοι αντιμετωπίζουμε λόγω της κρίσης, οι υποχρεώσεις και η πίεση... της

7:37PM - 1 Apr 2024

Ο ΡΟΛΟΣ ΤΟΥ ΑΝΟΣΟΠΟΙΗΤΙΚΟΥ ΣΤΗΝ ΚΑΤΑΘΛΙΨΗ

el.wikipedia.org
4:15PM - 5 Apr 2010

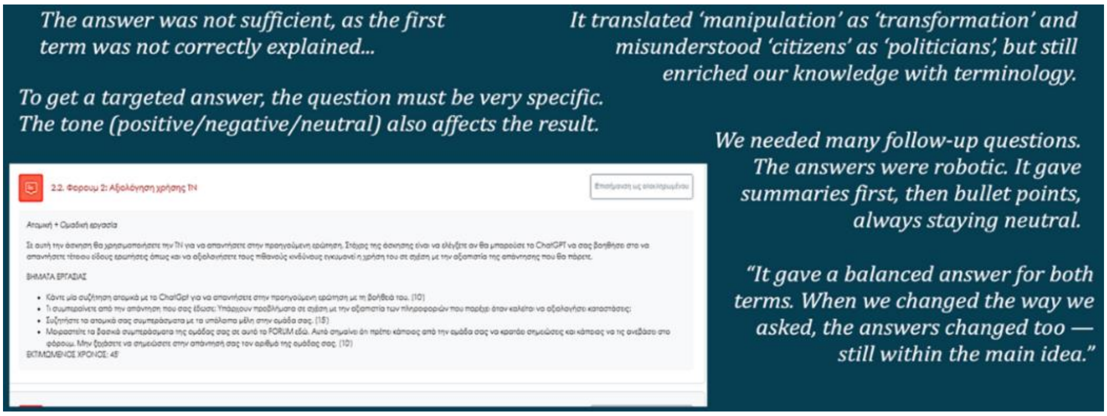
Αϋπνία - Βικιπαίδεια

Ο ύπνος θεραπεύει την κατάθλιψη | Η ΚΑΘΗΜΕΡΙΝΗ
Η θεραπεία της αϋπνίας σε άτομα που πάσχουν από κατάθλιψη τούς εξασφαλίζει πολύ περισσότερες πιθανότ ...

Κατάθλιψη: Φέρνει τα πάνω-κάτω στο ανοσοποιητικό σύστημα - Οι δείκτες που το αποκαλύπτουν
Νέα έρευνα αποκαλύπτει ακόμη έναν τρόπο που η καταθλιπτική διαταραχή επηρεάζει το ανοσοποιητικό σύστ...

Figure 4: Articles, Wikipedia lemmas, and social media posts annotated by the students

In the second activity, students documented prompts, compared outputs, and critically reflected on the interaction with ChatGPT (OpenAI, 2023). Working in groups, students assigned roles (e.g., AI operator, note-taker) and evaluated the same content with AI assistance. As shown in Figure 5, many discovered that vague or emotionally charged prompts resulted in superficial or misdirected responses, while precise, neutral language elicited more targeted and balanced answers. For example, one group observed that ChatGPT translated "manipulation" as "transformation" and misread "citizens" as "politicians", revealing both linguistic limitations and the system’s interpretive bias. Others noted that the AI's tone remained consistently neutral, often defaulting to summary formats that avoided evaluative positions. These observations prompted students to engage in multiple follow-up queries, experimenting with prompt tone and specificity to shape the output, a process they described as revealing and iterative. While they recognized AI's usefulness in expanding terminology and synthesizing information, they also acknowledged its inability to detect rhetorical nuance or source credibility without user intervention. This exercise ultimately highlighted a core CDL insight: AI can support, but not replace, critical human judgment in assessing information reliability and intent**.**



The answer was not sufficient, as the first term was not correctly explained...

It translated ‘manipulation’ as ‘transformation’ and misunderstood ‘citizens’ as ‘politicians’, but still enriched our knowledge with terminology.

To get a targeted answer, the question must be very specific. The tone (positive/negative/neutral) also affects the result.

We needed many follow-up questions. The answers were robotic. It gave summaries first, then bullet points, always staying neutral.

“It gave a balanced answer for both terms. When we changed the way we asked, the answers changed too — still within the main idea.”

2.2. Φορουμ 2: Αξιολόγηση χρήσης ΤΝ

Ατομική + Ομαδική εργασία

Σε αυτή την άσκηση θα χρησιμοποιήσετε την ΤΝ για να απαντήσετε στην προηγούμενη ερώτηση. Στόχος της άσκησης είναι να ελέγξετε αν θα μπορούσε το ChatGPT να σας βοηθήσει στο να απαντήσετε τέτοιου είδους ερωτήσεις όπως και να αξιολογήσετε τους πιθανούς κινδύνους εγκυμονεί η χρήση του σε σχέση με την αξιοπιστία της απάντησης που θα πάρετε.

ΒΗΜΑΤΑ ΕΡΓΑΣΙΑΣ

- Κάντε μια συζήτηση ατομικά με το ChatGpt για να απαντήσετε στην προηγούμενη ερώτηση με τη βοήθειά του. (10)
- Τι συμπεραίνετε από την απάντηση που σας έδωσε; Υπάρχουν προβλήματα σε σχέση με την αξιοπιστία των πληροφοριών που παρέχει όταν καλείται να αξιολογήσει καταστάσεις;
- Συζητήστε τα ατομικά σας συμπεράσματα με τα υπόλοιπα μέλη στην ομάδα σας. (15)
- Μοιραστείτε τα βασικά συμπεράσματα της ομάδας σας σε αυτό το FORUM εδώ. Αυτό σημαίνει ότι πρέπει κάποιος από την ομάδα σας να κρατάει σημειώσεις και κάποιος να τις ανεβάσει στο φόρουμ. Μην ξεχάσετε να σημειώσετε στην απάντησή σας τον αριθμό της ομάδας σας. (10)

ΕΚΤΙΜΩΜΕΝΟΣ ΧΡΟΝΟΣ: 45'

Figure 5: 2nd assignment example on the course’s Moodle platform (bottom left), and student quotes

Critical Creativity

Task Design and Pedagogical Foundations

Alongside analytical skills, the course integrated a creative dimension to boost engagement and deepen understanding, drawing on the concept of *Critical Creativity*. More specifically, the final course assignment, conducted in session 6, marked a shift from analysis to application. Students collaborated to create leaflets or posters targeting either peers or the general public, focusing on disinformation or source evaluation. This *project-based learning (PBL)* task was inquiry-driven, collaborative, and rooted in real-world communication. Students took on roles such as project coordinator, AI operator, content editor, and designer. They applied the evaluation tools from earlier sessions, used generative AI (e.g., ChatGPT) to develop content, and created materials using platforms like Canva and PowerPoint.

This approach supported not only the practical application of analytical tools but also key goals of CDL, such as engaging in real-world issues, communicating responsibly online, and fostering civic resilience against disinformation.

Student Outputs and Critical Creativity

The assignment offered an authentic context for students to apply the CLEAR Toolkit while exercising creative expression. It exemplified *Critical Creativity*, a concept where analytical and imaginative thinking intersect. Students had to engage their audience, collaborate digitally, and communicate complex ideas in accessible ways.

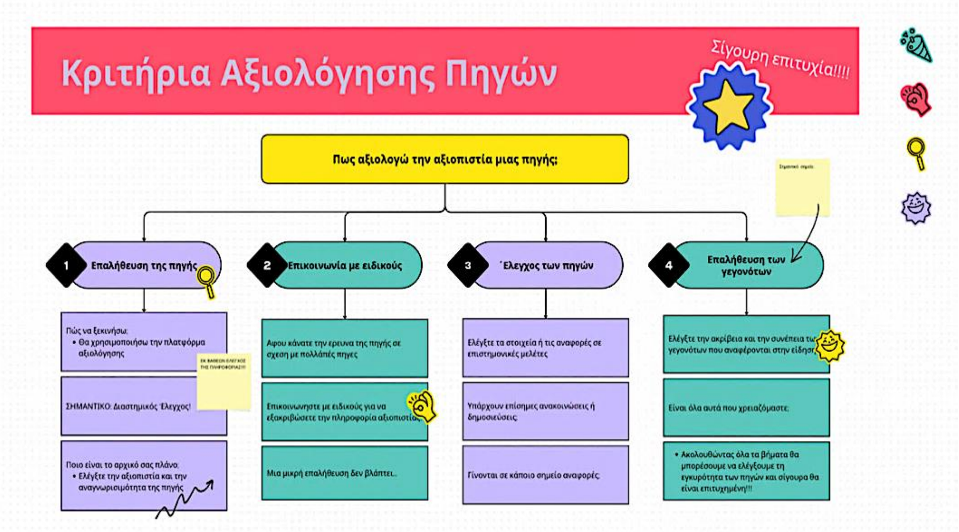
Although not introduced as a formal instructional model, Critical Creativity emerged during post-course reflection as a useful lens for interpreting student work. Their collaborative output demonstrated not only critical analysis but also creative synthesis, particularly evident in the design of their informational leaflets. Through these artifacts, students expressed nuanced understanding, audience awareness, and meaningful application of course concepts.

This interpretation aligns with research emphasizing the interplay between creativity and criticality. Creative tasks often depend on higher-order skills like reasoning, analysis, and reflection (Maor et al., 2023). Similarly, the U.S. State Department's *Critical Creativity in Action* framework defines it as "creative expression that demonstrates deeper thinking and nuanced understanding of content" (Bureau of Educational and Cultural Affairs, 2022). These ideas connect directly to students' engagement in the final project, where creativity became the mode through which they processed, applied, and communicated their learning. Frameworks like Torrance's (1974) model - highlighting fluency, flexibility, originality, and elaboration - further support this integration, especially when addressing complex challenges like disinformation.

Selected students' artifacts

Below are four student-designed artifacts (see Figures 6–9), each showing how learners synthesized course content through Critical Creativity:

- Figure 6 shows a flowchart which guides readers through steps like verifying sources and fact-checking. The structured format and accessible tone show pedagogical thinking and a strong grasp of evaluation concepts.



Κριτήρια Αξιολόγησης Πηγών

Σίγουρη επιτυχία!!!!

Πως αξιολογώ την αξιοπιστία μιας πηγής;

1 Επαλήθευση της πηγής

Πώς να ξεκινήσω;
- Θα χρησιμοποιήσω την πλατφόρμα αξιολόγησης

ΣΗΜΑΝΤΙΚΟ: Διαστημικός Έλεγχος!

Ποιο είναι το αρχικό σας πλάνο;
- Ελέγξτε την αξιοπιστία και την αναγνωρισιμότητα της πηγής

2 Επικοινωνία με ειδικούς

Αφου κάνατε την ερευνα της πηγής σε σχεση με πολλάπές πηγες

Επικοινωνήστε με ειδικούς για να εξακριβώσετε την πληροφορία αξιοπιστίας

Μια μικρή επαλήθευση δεν βλάπτει..

3 ΄Ελεγχος των πηγών

Ελέγξτε τα στοιχεία ή τις αναφορές σε επιστημονικές μελέτες

Υπάρχουν επίσημες ανακοινώσεις ή δημοσιεύσεις;

Γίνονται σε κάποιο σημείο αναφορές;

4 Επαλήθευση των γεγονότων

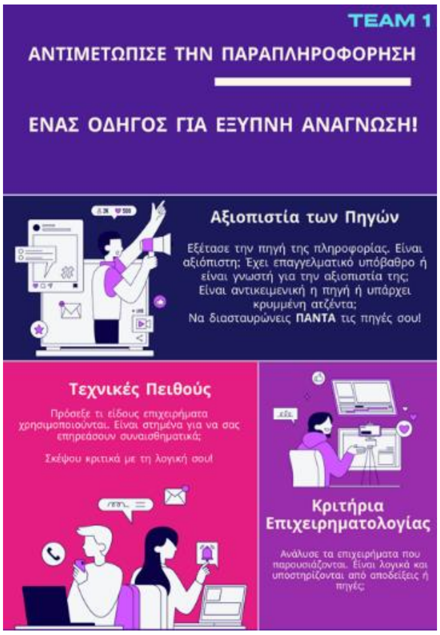
Ελέγξτε την ακρίβεια και την συνέπεια των γεγονότων που αναφέρονται στην είδηση;

Είναι όλα αυτά που χρειαζόμαστε;

- Ακολουθώντας όλα τα βήματα θα μπορέσουμε να ελέγξουμε τη εγκυρότητα των πηγών και σίγουρα θα είναι επιτυχημένη!!!

Figure 6: Source Evaluation Criteria

- The leaflet in Figure 7 merges source reliability, persuasion techniques, and logical reasoning into a simple, empowering guide. It highlights emotional manipulation and promotes logical consistency, using straightforward language and minimal design.



TEAM 1

ΑΝΤΙΜΕΤΩΠΙΣΕ ΤΗΝ ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ

ΕΝΑΣ ΟΔΗΓΟΣ ΓΙΑ ΕΞΥΠΝΗ ΑΝΑΓΝΩΣΗ!

Αξιοπιστία των Πηγών

Εξέτασε την πηγή της πληροφορίας. Είναι αξιόπιστη; Έχει επαγγελματικό υπόβαθρο ή είναι γνωστή για την αξιοπιστία της; Είναι αντικειμενική η πηγή ή υπάρχει κρυμμένη ατζέντα;
Να διασταυρώνεις **ΠΑΝΤΑ** τις πηγές σου!

Τεχνικές Πειθούς

Πρόσεξε τι είδους επιχειρήματα χρησιμοποιούνται. Είναι στημένα για να σας επηρεάσουν συναισθηματικά;

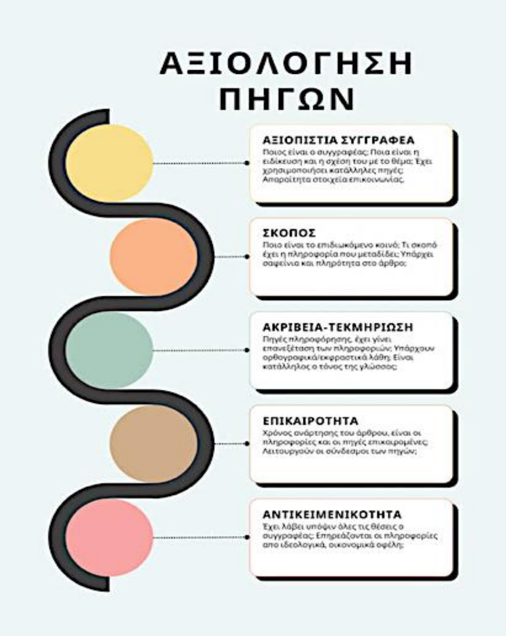
Σκέψου κριτικά με τη λογική σου!

Κριτήρια Επιχειρηματολογίας

Ανάλυσε τα επιχειρήματα που παρουσιάζονται. Είναι λογικά και υποστηρίζονται από αποδείξεις ή πηγές;

Figure 7: A Guide to Smart Reading

- The flowchart in Figure 8 integrates vocabulary like “objectivity” and “timeliness,” presenting evaluation as a series of reflective questions. Though visually basic, it demonstrates conceptual depth and a strong grasp of trustworthiness indicators learnt throughout the course.



ΑΞΙΟΛΟΓΗΣΗ ΠΗΓΩΝ

ΑΞΙΟΠΙΣΤΙΑ ΣΥΓΓΡΑΦΕΑ
Ποιος είναι ο συγγραφέας; Ποια είναι η ειδίκευση και η σχέση του με το θέμα; Έχει χρησιμοποιήσει κατάλληλες πηγές; Απαραίτητα στοιχεία επικοινωνίας.

ΣΚΟΠΟΣ
Ποιο είναι το επιδιωκόμενο κοινό; Τι σκοπό έχει η πληροφορία που μεταδίδει; Υπάρχει σαφείνια και πληρότητα στο άρθρο;

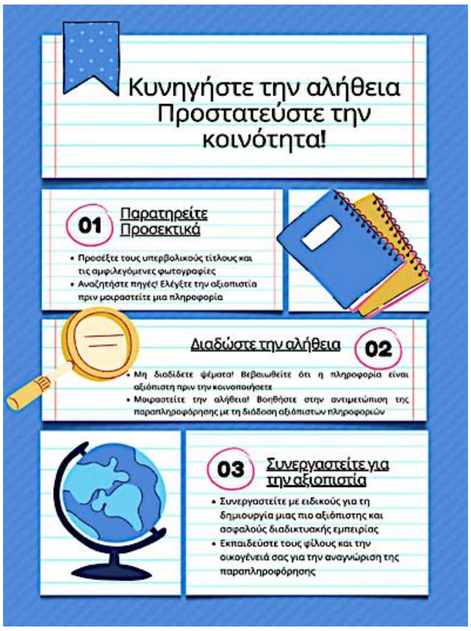
ΑΚΡΙΒΕΙΑ-ΤΕΚΜΗΡΙΩΣΗ
Πηγές πληροφόρησης, έχει γίνει επανεξέταση των πληροφοριών; Υπάρχουν ορθογραφικά/εκφραστικά λάθη; Είναι κατάλληλος ο τόνος της γλώσσας;

ΕΠΙΚΑΙΡΟΤΗΤΑ
Χρόνος ανάρτησης του άρθρου, είναι οι πληροφορίες και οι πηγές επικαιρομένες; Λειτουργούν οι σύνδεσμοι των πηγών;

ΑΝΤΙΚΕΙΜΕΝΙΚΟΤΗΤΑ
Έχει λάβει υπόψιν όλες τις θέσεις ο συγγραφέας; Επηρεάζονται οι πληροφορίες απο ιδεολογικά, οικονομικά οφέλη;

Figure 8: Sources' evaluation

- Figure 9: A public-facing poster uses short, imperative phrases like "Observe Carefully" and "Spread the Truth" to promote civic responsibility. Its design appeals to communal values, translating abstract principles into actionable messages for everyday digital engagement.



Κυνηγήστε την αλήθεια
Προστατεύστε την
κοινότητα!

01 Παρατηρείτε Προσεκτικά

- Προσέξτε τους υπερβολικούς τίτλους και τις αμφιλεγόμενες φωτογραφίες
- Αναζητήστε πηγές! Ελέγξτε την αξιοπιστία πριν μοιραστείτε μια πληροφορία

Διαδώστε την αλήθεια 02

- Μη διαδίδετε ψέματα! Βεβαιωθείτε ότι η πληροφορία είναι αξιόπιστη πριν την κοινοποιήσετε
- Μοιραστείτε την αλήθεια! Βοηθήστε στην αντιμετώπιση της παραπληροφόρησης με τη διάδοση αξιόπιστων πληροφοριών

03 Συνεργαστείτε για την αξιοπιστία

- Συνεργαστείτε με ειδικούς για τη δημιουργία μιας πιο αξιόπιστης και ασφαλούς διαδικτυακής εμπειρίας
- Εκπαιδεύστε τους φίλους και την οικογένειά σας για την αναγνώριση της παραπληροφόρησης

Figure 9: Chase the truth

These artifacts reflect more than content knowledge. They represent students' ability to communicate persuasively and ethically using digital tools. By combining critical analysis, AI-assisted drafting, and peer collaboration, students gained insight into both the potential and limitations of AI, the demands of audience-aware design, and the ethics of digital participation. Through this, they enacted the core principles of CDL and demonstrated the transformative potential of Critical Creativity in higher education.

Conclusion

The course intervention offers key insights for educators integrating CDL and AI tools. First, the course went beyond a basic introduction to disinformation by creating a structured, hands-on environment where students developed CDL through inquiry, reflection, and creative engagement. Namely, students engaged with concepts such as bias, manipulation, and source reliability not only in theoretical terms but also through applied exercises using real-world digital content and tools like ChatGPT and Truly Media.

By the end of the course, students demonstrated that they had not only grasped key terminology and theoretical frameworks concerning disinformation but had also developed the capacity to critically analyze and evaluate the reliability of information across diverse media environments. They became increasingly aware of how both media and AI can perpetuate bias, partiality, or even disinformation, and how these risks can be mitigated through rigorous evaluation, reflective judgment, and responsible use of digital technologies.

Importantly, the course fostered student agency through collaborative work and the use of digital tools for content creation. The final project - designing informational brochures for peers or the public - highlighted the effectiveness of the pedagogical approach. These student-designed materials synthesized key course concepts in a communicative and engaging format. The process of planning, drafting, and designing these outputs invited students to reframe and reinterpret what they had learned, depending on their chosen target audience. This aligns closely with the framework of Critical Creativity, where “learners use creative expression to demonstrate deeper thinking and nuances of understanding content” (Bureau of Educational and Cultural Affairs, Office of English Language Programs, 2022).

Ultimately, the course served as a powerful example of how higher education can empower students to address disinformation both critically and constructively. It demonstrated that by blending theory with digital practice, and by integrating AI tools through a lens of ethical responsibility, educators can foster both media literacy and civic engagement. In an era marked by information overload, algorithmic influence, and growing distrust in media, this combination of critical thinking, creativity, and digital fluency is not only timely but also essential.

References

Blakeslee, S. (2004). The CRAAP test. *LOEX Quarterly, 31*(3), Article 4. https://commons.emich.edu/loexquarterly/vol31/iss3/4

Buckingham, D. (2005). *The media literacy of children and young people. A review of the research literature* on behalf of Ofcom. Retrieved from: https://www.researchgate.net/publication/253736824_The_Media_Literacy_of_Children_and_Young_People

Buckingham, D. (2019). Teaching media in a 'post-truth' age: fake news, media bias and the challenge for media/digital literacy education / La enseñanza mediática en la era de la posverdad: fake news, sesgo mediático y el reto para la educación en materia de alfabetización mediática y digital. *Culture and Education*, *31*(2), 213-231. https://doi.org/10.1080/11356405.2019.1603814

Bureau of Educational and Cultural Affairs, Office of English Language Programs. (2022). *Critical Creativity in Action: Teacher's Manual*. U.S. Department of State. https://americanenglish.state.gov/files/ae/resource_files/critical_creativity_in_action_-_teachers_manual.pdf

González-Pérez, Y. (2020). Disinformation in the information age. *European Journal of Hospital Pharmacy*, *27*(6). https://doi.org/10.1136/ejhpharm-2020-002213

Jacques, J., Fastrez, P., De Smedt, T. (2013). *Organizing media as social objects. An exploratory assessment of a core media literacy competence*. https://www.researchgate.net/publication/251573827_Organising_Media_as_Social_Objects_an_exploratory_assessment_of_a_core_media_literacy_competence

Hobbs, R. (2010). *Digital and Media Literacy: A Plan of Action*. The Aspen Institute. https://files.eric.ed.gov/fulltext/ED523244.pdf

Katsarou, E.; Sipitanos, K.; Samioti, P.; Ventsel, A.; Madisson, M., L.; Terzis, G.; Kotzinos, D.; Waszak Z. P.; Gavalas, C. (publication expected in 2026). *A Quality Assurance Matrix* for Educational Design to Combat Disinformation in Higher Education*, Journal of Media Literacy Education.*

Leaning, M. (2017). *Media and Information Literacy: An Integrated Approach for the 21st Century*. Chandos Publishing.

Maor, R., Paz-Baruch, N., Grinshpan, N., Milman, A., Mevarech, Z., Levi, R., Shlomo, S., & Zion, M. (2023). Relationships between metacognition, creativity, and critical thinking in self-reported teaching performances in project-based learning settings. *Thinking Skills and Creativity, 50*. https://doi.org/10.1016/j.tsc.2023.101425

OpenAI. (2023). ChatGPT (GPT-3.5). https://chat.openai.com/

Rau, M. A., & Premo, A. E. (2025). Systematic Review of Educational Approaches to Misinformation. *Educational Psychology Review*, *37*(2). https://doi.org/10.1007/s10648-025-10012-8

Samioti, P. (2024). *Academic Language Skills to Combat Disinformation.* Spring Term Course 2024, S.HI.E.L.D. vs Disinfo Project. Elearn Platform, UoC

S.HI.E.L.D. vs Disinfo (2025a) WP3_R1: CURRICULA. In *Deliverables* https://shieldvsdisinfo.com/deliverables/ (Accessed on 2nd July 2025).

S.HI.E.L.D. vs Disinfo (2025b) *Educators E-toolkit*. https://shieldvsdisinfo.com/educators-e-toolkit/ (Accessed on 2nd July 2025).

Tomassi, A., Falegnami, A., & Romano, E. (2025). *Disinformation in the Digital Age: Climate Change, Media Dynamics, and Strategies for Resilience*. *Publications*, *13*(2). https://doi.org/10.3390/publications13020024

Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking: Norms-Technical Manual. Research Edition. Verbal Tests, Forms A and B. Figural Tests, Forms A and B.* Princeton, NJ: Personnel Press.

Truly Media. (n.d.). Truly Media Web-based collaboration platform for verifying digital content. https://www.truly.media/

Source Evaluation in a Digital World: From Information Consumers to Critical Evaluators

Zoi A. Traga Philippakos
University of Tennessee, Knoxville

Abstract

The proliferation of social media, widespread information sharing, and the integration of AI tools such as ChatGPT in education have amplified the need for learners to critically evaluate the credibility of information. This paper advocates for the deliberate development of source evaluation processes and strategies that are transferable across contexts, enabling students to assess both online and AI-generated content effectively. We examine how educators can embed these practices into instruction to promote critical thinking, uphold academic integrity, and cultivate responsible digital literacy.

Keywords: source evaluation, reading online, writing

The Imperative for Critical Reading and Source Evaluation in the Age of AI

The rapid integration of artificial intelligence tools such as ChatGPT and other large language models (LLMs) into educational settings is reshaping how students access and engage with information. These technologies offer powerful capabilities but also introduce complex challenges related to the reliability and credibility of the content produced. While concerns about AI's generation of factually incorrect or fabricated information are valid, they represent only part of a broader digital literacy challenge. Today's learners are immersed in a vast and varied information ecosystem that includes search engines, social media platforms, websites of differing trustworthiness, and AI-generated materials. Within this complex environment, the ability to critically evaluate sources has emerged as an essential academic and life skill, requiring intentional cultivation by educators and learners alike. Dimensions such as authenticity, accuracy, and relevance are central to establishing the quality of content (Huang et al., 2025). However, articulating these dimensions at a theoretical level is insufficient without providing learners with practical, applicable tools to implement critical evaluation in real-world academic and everyday scenarios.

Traditional approaches to source evaluation, such as the CRAAP test (Currency, Relevance, Authority, Accuracy, and Purpose) developed by Blakeslee (2004), have been instrumental in guiding learners through systematic assessment of information credibility. Further, a variety of other pedagogical strategies have been proposed and implemented to foster critical evaluation skills (e.g., lateral reading; Wineburg & McGrew, 2017). The rise of AI-generated content also necessitates specific educational responses tailored to the unique characteristics of this medium. Unlike traditional sources, AI-generated texts can synthesize vast amounts of information into seamless narratives, but they may also introduce subtle inaccuracies or biased perspectives without clear attribution. Therefore, critical evaluation must include an understanding of the limitations and potential errors inherent in AI outputs. This includes teaching students to question not only the content itself but also the underlying algorithms and data sets that inform AI responses.

Effective integration of critical evaluation instruction into curricula requires a multifaceted approach. Educators must develop learners' metacognitive skills to reflect on their own information processing habits and biases. Furthermore, embedding evaluation strategies into authentic tasks, for instance, research projects or collaborative discussions, provides students with meaningful opportunities to practice and internalize these skills. Additionally, fostering a classroom culture that values inquiry, skepticism, and ethical responsibility supports the ongoing development of critical literacy.

As the digital landscape continues to evolve, so too must educational strategies aimed at preparing learners to navigate it responsibly. The convergence of AI technologies and the persistent challenges of misinformation underscores the urgency of equipping students with robust, adaptable tools for critical reading and source evaluation. By addressing these needs, educators can help ensure that learners are not only consumers of information but also discerning evaluators capable of contributing thoughtfully to academic, professional, and civic conversations.

Pedagogical Approaches to Developing Critical Evaluation Skills

Addressing these challenges requires well-designed pedagogical approaches that are age-appropriate, scaffolded, and contextually relevant. Educators play a pivotal role in creating learning experiences that build students' capacities to think critically about information sources.

Explicit Instruction and Modeling

One foundational strategy is explicit instruction. During explicit and systematic instruction, teachers clearly demonstrate evaluation techniques, articulate reasoning processes, and provide guided practice. For example, educators might analyze a news article or AI-generated text in class, highlighting how to check for author credentials, publication date, source citations, and potential biases (see also Philippakos, 2018). Modeling this process makes abstract criteria tangible and equips students with a cognitive framework for independent evaluation. Students may proceed with their instructor in collaborative implementation of those criteria and through a gradual release of responsibility proceed with evaluation of sources they locate for their research project. This collaborative implementation may take the form of working as a group with the instructor supporting the application of the taught strategies or in small groups.

Peer collaboration provides opportunities for students to articulate their reasoning, challenge one another's interpretations, and refine their evaluative criteria. Students may engage in discussions over specific sources, debates over content read and its accuracy (while they apply the criteria to prove that something is or it is not accurate), or peer review activities. The latter when it utilizes specific criteria and guidelines can support learners' revisions as well as deeper understanding of the crtieria (see Philippakos & MacArthur, 2016).

Strategic Reading and Rereading Across Modalities

The core of the Developing Strategic Writers approach emphasizes the importance of teaching students metacognitive strategies—conscious, goal-directed reading and rereading behaviors—that enable them to actively construct meaning and critically engage with texts (Philippakos & MacArthur, 2016). These strategies include previewing content to set purposes, questioning the author's intent and credibility, monitoring comprehension, and revisiting text sections to clarify or verify information.

In the context of digital information and AI-generated outputs, these strategies must be adapted and extended. Online sources such as websites, blogs, and social media posts often combine text with images, hyperlinks, and embedded videos. Students benefit from learning to evaluate each element critically, recognizing how multimedia components can influence meaning and perception. For example, a video accompanying a news article might add emotional appeal that shapes readers' interpretation but may not be fact-checked or neutral.

Similarly, AI-generated text requires an additional layer of scrutiny. Since AI systems produce content algorithmically, students must practice rereading to detect inconsistencies, factual errors, or biased framing that may not be evident on first reading. Encouraging multiple passes—such as initial comprehension followed by verification through external fact-checking or lateral reading—aligns with the DSW emphasis on strategic rereading as a tool for deepening understanding and critical inquiry.

Critical Questioning

Critical questions can support curiosity and skepticism by engaging students gathering, evaluating, and synthesizing of diverse information sources. For example, in our work within the Developing Strategic Learners program, we engage students in the examination of published information (e.g., news) from different outlets for them to examine how information is presented, the tone of the presentation, the word usage, the accuracy of information (and how accuracy can be examined, and what bias may exist.

Digital Literacy Integration

Embedding evaluation skills within broader digital literacy instruction recognizes that source evaluation is part of responsible technology use. Lessons might include the identification of misinformation techniques and practicing ethical digital citizenship (Hobbs, 2017). Digital literacy curricula can also introduce tools like fact-checking websites, browser extensions, or AI detection software that assist learners in real-time source verification.

Challenges and Considerations in Technology Integration

While technology offers powerful support for strategy instruction, educators must remain mindful of potential challenges. Access and equity issues can limit students' opportunities to engage with digital tools. Furthermore, an overreliance on technological scaffolds without foundational strategy instruction risks superficial engagement.

Therefore, integrating technology must be accompanied by intentional pedagogy that centers on developing deep comprehension, critical thinking, and ethical information use. Blending traditional literacy practices with technology-enhanced strategies creates a balanced approach that prepares learners for complex information environments

Teacher Professional Development for Technology-Supported Strategy Instruction

Professional development (PD) plays a central role in enabling teachers to effectively integrate instructional approaches, particularly within the evolving landscape of AI-generated content and multimedia sources. Many educators report feeling unprepared to navigate and teach digital literacy competencies, especially those involving complex technologies like AI tools or multimodal texts (Trust & Whalen, 2021). Traditional teacher preparation programs often lack focused training on technology integration for literacy instruction, and ongoing PD opportunities may not keep pace with emerging digital challenges (also see Traga Philippakos et al., 2022).

Without adequate support, teachers risk relying on outdated methods or avoiding technology-enhanced strategies altogether, which limits students' opportunities to develop critical evaluation skills essential for today's information environment. Furthermore, teachers themselves must become critical consumers of digital information to model effective strategies and foster a classroom culture of inquiry and skepticism. Research-based guidelines on PD can support PD on evaluation and on AI integration in literacy tasks. One time or inconsistent (in scope and sequence) workshops are insufficient for developing deep understanding and pedagogical change. Instead, multi-session, scaffolded PD over time allows teachers to gradually build skills and apply them in practice, reflecting and adjusting based on classroom experience (Desimone & Garet, 2015). Further, PD should provide explicit instruction on both digital literacy content, such as evaluation criteria for online and AI-generated sources and writing practices for reading and writing to be connected. This includes demonstrating how to teach rereading techniques and writing strategies as well as technology tool integration. Teachers benefit from hands-on experiences using relevant technology tools (e.g., fact-checking resources) and from opportunities to collaboratively examine (develop and revise and think through) plans tailored to their grade levels and subjects. In essence, peer collaboration within PD encourages sharing of best practices, problem-solving around challenges, and collective growth.

Supporting Teacher Confidence and Instructional Adaptability

Professional development that is ongoing, content based, collaborative and reflective as well as practical and contextual can supports teacher self-efficacy in their ability to implement technology-supported strategy instruction that can potentially support student outcomes. When teachers feel capable and prepared, they are more likely to experiment with new tools, embed critical evaluation across subjects, and foster engagement with texts. Additionally, as digital tools and AI applications evolve rapidly, PD must emphasize adaptability. Equipping teachers with foundational pedagogical frameworks rather than specific AI or digital tool training, assures that teaches can respond flexibly to future innovations. This adaptability supports sustained integration of critical literacy skills regardless of technological changes.

Institutional Support and Infrastructure

For PD efforts to be successful, institutional commitment is essential. Schools and districts must allocate time, resources, and technical support for ongoing teacher learning (also see Traga Philippakos, 2022). Leadership support can help embed technology-supported strategy instruction into curricular priorities, assessment systems, and professional learning communities.

Moreover, addressing equity considerations and making sure that all educators have access to necessary devices, internet connectivity, and software are critical to prevent gaps in access and in student learning. Teachers equipped through targeted PD to integrate technology-enhanced strategy instruction are better positioned to guide students in becoming discerning digital citizens (Leu et al., 2017).

Conclusion

In an era characterized by unprecedented access to information and the growing presence of artificial intelligence tools such as ChatGPT, the ability to critically evaluate and reread digital and AI-generated content has become an indispensable academic and life skill. The challenges posed by the complexity, volume, and variable credibility of information sources require learners to move beyond passive consumption toward active, strategic engagement with texts in all formats, including online articles, multimedia, and AI outputs.

Integrating technology thoughtfully into literacy instruction offers powerful scaffolds that can enhance students’ strategic reading and evaluation practices. Digital tools for annotation, fact-checking, and multimedia interaction facilitate active engagement and deepen learners’ critical inquiry. The Developing Strategic Writers attempts to embed these strategies within authentic, meaningful tasks that promote transfer across subjects and formats.

Equally crucial is the provision of sustained, content-focused, and practical PD that empowers teachers to confidently implement technology-supported strategy instruction. Such PD fosters adaptability in the face of rapidly evolving digital tools, equips educators to model critical consumption of information, and supports equitable access to digital literacy learning. Overall, preparing learners to responsibly navigate the complex information landscape of today and tomorrow demands a comprehensive approach—one that integrates rigorous source evaluation, strategic rereading and critical evaluation, and teacher capacity building.

References

Blakeslee, S. (2004). The CRAAP Test, *LOEX Quarterly*, *31*(3), Article 4. https://commons.emich.edu/loexquarterly/vol31/iss3/4

Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society & Education*, *7*(3), 252–263. https://doi.org/10.25115/psye.v7i3.515

Hobbs, R. (2017). Approaches to teacher professional development in digital media literacy education. In B. De Abreu, P. Mihailidis, A. Lee, J. Melki & J. McDougall (Eds). I*nternational Handbook of Media Literacy Education* (pp. 88 – 113) Routledge.

Kim, J., & Lee, Y. (2024). Accuracy Evaluation of Tree Images Created Using Generative Artificial Intelligence. *Journal of Digital Landscape Architecture*, *2024*(9), 1029–1037. https://doi.org/10.14627/537752098

Leu, D.J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2017). New Literacies: A Dual-Level Theory of the Changing Nature of Literacy, Instruction, and Assessment. *Journal of Education*, *197*(2), 1–18. https://doi.org/10.1177/002205741719700202

Philippakos. Z. A., MacArthur, C. A. & Coker, D. L. (2015). *Developing strategic writers through genre instruction: Resources for grades 3-5.* Guilford Press.

Traga Philippakos, Z. A. (2025, July 27). Developing Strategic Learners: Source evaluation for credible writing. *Revisiting Disinformation: Critical Media Literacy Approaches.* Paper presented at the S.HI.E.L.D. vs Disinfo project conference, Crete, Greece.

Trust, T., & Whalen, J. (2021). Emergency remote teaching with technology during the COVID-19 pandemic: using the whole teacher lens to examine educator's experiences and insights. *Educational Media International*, *58*(2), 145–160. https://doi.org/10.1080/09523987.2021.1930479

Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record (1970)*, *121*(11). https://doi.org/10.1177/016146811912101102

The Immunity Illusion: Media Literacy in an Age of Identity-Driven Disinformation

Massimo Flore
Independent Researcher, massimo.flore@outlook.com

Abstract

Traditional media literacy approaches focusing on analytical skills demonstrate limited effectiveness against contemporary disinformation, particularly when information contradicts deeply entrenched identity commitments and belief systems. Based on research from cognitive science, political psychology, and communication studies, complemented by practitioner experience teaching media literacy courses, this article develops an integrative theoretical framework explaining why conventional media literacy education often fails when confronted with identity-relevant misinformation. To be effective, media literacy must address four interconnected dimensions: analytical processing, identity-protective cognition, emotional responses, and narrative persuasion, all operating within structured belief networks that filter information processing. This framework offers theoretically-grounded strategies for strengthening resistance against manipulation techniques that exploit psychological vulnerabilities, with particular attention to applications within political communication scenarios.

Keywords: media literacy, identity-protective cognition, political polarization, disinformation resistance, affective polarization

Introduction

Digital media proliferation has fundamentally revolutionized the information ecosystem, simultaneously democratizing knowledge access while enabling unprecedented spread of disinformation (Wardle & Derakhshan, 2017). Media literacy education has consequently emerged as a primary intervention strategy, usually emphasizing critical thinking, source evaluation, and verification skills (Buckingham, 2019). However, these traditional approaches have shown inconsistent effectiveness when confronted with sophisticated disinformation strategies that exploit fundamental psychological processes (Bulger & Davison, 2018; Guess et al., 2020).

This effectiveness gap becomes most evident when misinformation aligns with preexisting beliefs, social identities, or partisan affiliations. Research in political psychology demonstrates that people process information in fundamentally different ways depending on whether it relates to identity-relevant beliefs compared to neutral content (Kunda, 1990; Taber & Lodge, 2006). Even individuals with advanced education and critical thinking skills consistently show vulnerability to manipulative narratives that align with their political identities (Kahan et al., 2012; Iyengar et al., 2019).

These findings reveal a critical limitation in conventional media literacy approaches: the assumption that providing individuals with analytical skills and accurate information is sufficient to correct misperceptions and build resistance to manipulation. This "information deficit model" overlooks how identity protection, emotional processing, and narrative framing fundamentally shape information evaluation regardless of analytical capabilities (Lewandowsky & van der Linden, 2021).

While previous scholars have identified various psychological barriers to effective media literacy (Cooke, 2018; Mihailidis, 2018), the field lacks an integrative theoretical framework that systematically connects these psychological dimensions to specific educational interventions, particularly for addressing politically polarized content. This article addresses this gap by developing a comprehensive framework integrating four dimensions of information processing relevant to media literacy: analytical evaluation, identity-protective cognition, emotional responses, and narrative persuasion.

Methodology

This theoretical paper employs a qualitative synthesis methodology combining systematic literature review, critical discourse analysis, and reflective practitioner insights. The literature review examined empirical studies from cognitive science, political psychology, and communication studies published between 2000-2023, focusing on psychological mechanisms influencing information processing beyond analytical reasoning. Critical discourse analysis identified points of consensus and contention across disciplinary boundaries regarding misinformation susceptibility. Practitioner insights from teaching media literacy courses (2021-2023) helped identify gaps between theoretical constructs and classroom applications.

This triangulated methodology emphasizes theoretical integration grounded in established research while acknowledging limitations. The framework lacks direct experimental validation, practitioner insights were not collected using rigorous qualitative research methods, and proposed educational interventions require empirical verification in controlled studies.

The Limitations of Traditional Media Literacy Approaches

Traditional media literacy education emerged from a primarily cognitive tradition, focusing on source evaluation, fact verification, and logical analysis (Potter, 2004). However, empirical research increasingly demonstrates these limitations, particularly in politically charged contexts. Several studies report that conventional media literacy interventions show modest or inconsistent impact on misinformation susceptibility when addressing politically polarized content (Guess et al., 2020; Vraga & Bode, 2017).

These limitations follow a fundamental misconception: the assumption that misinformation acceptance primarily results from information deficits or analytical failures. However, substantial evidence from political psychology indicates that misinformation susceptibility is often caused by motivated reasoning processes that serve identity protection and psychological needs (Flynn et al., 2017; Kahan, 2017).

In politically polarized environments, these limitations become more pronounced. Research on affective polarization demonstrates that emotional attitudes toward opposing political groups increasingly drive information processing (Iyengar et al., 2019). Mason (2018) illustrates how political identities have become increasingly aligned with other social identities, creating "mega-identities" that make identity-threatening information psychologically expensive to embrace.

Four-Dimensional Framework for Media Literacy

Identity-Protective Cognition in Political Information Processing

When information threatens beliefs connected to valued political identities, individuals experience psychological distress that motivates defensive processing (Sherman & Cohen, 2006). Kahan et al. (2017) demonstrate that individuals often use their cognitive abilities to defend beliefs tied to their political or cultural identities. When confronted with identity-threatening information, individuals engage in motivated reasoning to preserve group affiliation.

The political science literature on negative partisanship further illuminates these dynamics. Abramowitz and Webster (2016) show how increasing antipathy toward opposing political groups drives information processing to the extent that processing out-group information becomes psychologically threatening.

Beyond identity protection, individuals' broader belief systems play a central role in shaping information consumption. Drawing on models by Rokeach (1960) and Stroud (2011), belief systems function as structured, hierarchical networks that act as interpretive filters for political information, prioritizing coherence and emotional salience over factual alignment.

Emotional Dimensions of Political Information

Traditional media literacy education often assumes a predominantly rational model of information processing, emphasizing analytical critical thinking over emotion's role in political belief development. Research in affective science demonstrates that emotional responses frequently precede and influence cognitive evaluation of information, particularly for content designed to provoke fear, outrage, or moral indignation (Damasio, 1994; Barrett, 2017).

Brady et al. (2017) demonstrated that morally and emotionally charged political content spreads faster and farther in social networks than messages without emotional and moral connotations. This affective spreading helps explain why emotionally provocative political misinformation often achieves greater reach and impact than neutral, factually accurate content.

Narrative Persuasion in Political Communication

Sophisticated political disinformation strategies often employ narrative structures that bypass critical resistance through storytelling. Research on narrative persuasion shows that information framed within narrative formats is processed differently than information presented as isolated facts or arguments (Green & Brock, 2000; Hamby et al., 2018).

Transportation theory establishes that effective narratives transport audiences into story worlds, reducing critical distance and counter-arguing against embedded persuasive messages (Green & Brock, 2000). Traditional fact-checking approaches typically break narratives into isolated claims, addressing each claim individually while missing the holistic power of narrative persuasion.

Educational Applications

Based on the theoretical perspectives outlined above, specific educational strategies should address each dimension of information processing:

Analytical Dimension

Building upon traditional analytical literacy requires integrating meta-cognitive awareness. Educational interventions should include structured analysis of one's own political media diet, cross-cutting content exposure exercises, and decision journaling techniques that document reasoning processes when evaluating political claims.

Identity Dimension

Addressing identity-protective cognition requires helping individuals recognize when political identity protection motivates their information processing. Media literacy education should incorporate identity complexity awareness exercises, targeted self-affirmation interventions before engaging with politically threatening information, and structured exposure to politically diverse viewpoints within safe learning environments.

Emotional Dimension

Since emotional reactions significantly influence susceptibility to political misinformation, emotional literacy should enable individuals to recognize and regulate emotional responses to political content. This includes emotional awareness training, analysis of emotional appeals across ideological perspectives, and emotional distancing techniques for cognitive reappraisal.

Narrative Dimension

Narrative persuasion requires developing narrative literacy - the ability to analyse narrative structures in political communication and construct alternative narrative frameworks. Educational approaches should include political narrative structure analysis, counter-narrative development, and case study analysis of successful political misinformation narratives.

Policy Implications

This framework has significant implications for policymakers seeking to counter political disinformation. Educational policy should recognize that media literacy needs to move beyond conventional digital literacy frameworks. Public information campaigns addressing politically charged topics should incorporate insights from identity-protective cognition research. Platform design regulations could incorporate insights about identity-protective cognition and emotional dynamics. Civic education initiatives should integrate media literacy approaches that specifically address the role of identity, emotion, and narrative in political information processing.

Conclusion

Traditional media literacy approaches focusing primarily on analytical skills have shown inconsistent effectiveness against contemporary political disinformation, particularly when misleading content aligns with individuals' identity commitments and emotional predispositions. The integrative framework presented addresses this limitation by conceptualizing media literacy as a multidimensional practice engaging four interconnected aspects of information processing: analytical evaluation, identity-protective cognition, emotional responses, and narrative persuasion.

This framework makes three contributions to media literacy research and practice. First, it synthesizes insights from cognitive science, political psychology, and communication studies that have previously remained confined within disciplinary boundaries. Second, it explicitly connects psychological mechanisms with specific educational interventions, providing a more robust theoretical foundation for media literacy pedagogy. Third, it proposes educational approaches that address traditional analytical models' limitations while building upon their strengths.

The primarily theoretical nature represents both the framework's contribution and key limitation. Although based on established research across multiple disciplines, the integrated approach requires empirical validation. Future research should empirically test this framework's effectiveness across different educational settings and politically diverse populations through controlled experiments comparing traditional media literacy interventions with approaches incorporating identity, emotional, and narrative dimensions.

References

Abramowitz, A. I., & Webster, S. (2016). The rise of negative partisanship and the nationalization of U.S. elections in the 21st century. *Electoral Studies*, *41*, 12-22.

Albertson, B., & Gadarian, S. K. (2015). *Anxious politics: Democratic citizenship in a threatening world*. Cambridge University Press.

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, *47*(2), 241-262. https://doi.org/10.1017/S0007123415000186

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313-7318. https://doi.org/10.1073/pnas.1618923114

Buckingham, D. (2019). *The media education manifesto*. Polity Press.

Bulger, M., & Davison, P. (2018). The promises, challenges, and futures of media literacy. *Journal of Media Literacy Education*, *10*(1), 1-21. https://doi.org/10.23860/JMLE-2018-10-1-1

Cohen, G. L., Sherman, D. K., Bastardi, A., Hsu, L., McGoey, M., & Ross, L. (2007). Bridging the partisan divide: Self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation. *Journal of Personality and Social Psychology*, *93*(3), 415-430. https://doi.org/10.1037/0022-3514.93.3.415

Cooke, N. A. (2018). *Fake news and alternative facts: Information literacy in a post-truth era*. American Library Association.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam.

Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, *38*(S1), 127-150. https://doi.org/10.1111/pops.12394

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*(5), 701-721. https://doi.org/10.1037/0022-3514.79.5.701

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27), 15536-15545. https://doi.org/10.1073/pnas.1920498117

Hamby, A., Brinberg, D., & Jaccard, J. (2018). A conceptual framework of narrative persuasion. *Journal of Media Psychology*, *30*(3), 113-124. https://doi.org/10.1027/1864-1105/a000190

Hobbs, R. (2010). *Digital and media literacy: A plan of action*. Aspen Institute.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*, 129-146. https://doi.org/10.1146/annurev-polisci-051117-073034

Jones, M. D., & Song, G. (2014). Making sense of climate change: How story frames shape cognition. *Political Psychology*, *35*(4), 447-476. https://doi.org/10.1111/pops.12113

Kahan, D. M. (2017). The expressive rationality of inaccurate perceptions. *Behavioral and Brain Sciences*, 40. https://doi.org/10.1017/S0140525X1500064X

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54-86. https://doi.org/10.1017/bpp.2016.2

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732-735. https://doi.org/10.1038/nclimate1547

Kreiss, D., & McGregor, S. C. (2023). A review and provocation: On polarization and platforms. *New Media & Society*, *26*(1), 556-579.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498. https://doi.org/10.1037/0033-2909.108.3.480

Levendusky, M. (2018). Americans, not partisans: Can priming American national identity reduce affective polarization? *Journal of Politics*, *80*(1), 59-70. https://doi.org/10.1086/694985

Lewandowsky, S., & van der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, *32*(2), 348-384.

Mason, L. (2018). *Uncivil agreement: How politics became our identity*. University of Chicago Press.

Mihailidis, P. (2018). Civic media literacies: Re-imagining engagement for civic intentionality. *Learning, Media and Technology*, *43*(2), 152-164. https://doi.org/10.1080/17439884.2018.1428623

Potter, W. J. (2004). *Theory of media literacy: A cognitive approach*. Sage Publications.

Rokeach, M. (1960). *The open and closed mind: Investigations into the nature of belief systems and personality systems*. Basic Books.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology*, *38*, 183-242. https://doi.org/10.1016/S0065-2601(06)38004-5

Stroud, N. J. (2011). *Niche news: The politics of news choice*. Oxford University Press.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755-769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, *39*(5), 621-645. https://doi.org/10.1177/1075547017731776

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.

Critical Literacy as a Tool for Approaching the Rhetoric of Multimodal Texts in Secondary Education: A Teaching Approach to Media Analysis in the Classroom

Anna Paraskevaidi
PhD, Secondary School Teacher annaparask@gmail.com

Abstract

This paper presents teaching practices implemented with 16-year-old Greek high school students, aiming to develop their critical literacy skills through the analysis and redesign of multimodal advertisements. Drawing on the theoretical framework of Critical Literacy and Multimodality, the activities engaged students in analyzing advertising's role in spreading misinformation, normalizing stereotypes and influencing consumer beliefs. Through rhetorical analysis, counter-advertisement creation and historical comparisons, students learned to question persuasive strategies, recognize hidden ideologies and respond critically. These student-centered practices promote agency and social awareness, helping learners shift from passive media consumption to active, reflective and socially responsible communication.

Keywords: critical media literacy, advertisements, disinformation, Artificial Intelligence, secondary education

Introduction

In recent years, there has been growing concern about how advertising manipulates adolescents, especially through digital media and social networks (Levine, 2023; Raffoul, Ward, Santoso, Kavanaugh & Austin, 2023; Pellegrino & Stasi 2024). Advertisements for energy drinks, beauty products and supplements, promote unrealistic expectations and encourage uncritical consumption, often affecting young people's health, values and self-image (Radesky, Chassiakos, Ameenuddin, & Navsaria, 2020; Van Reijmersdal & van Dam, 2020; Packer, Croker, Goddings, Boyland, Stansfield, Russell & Viner, 2022; Lajnef, 2023).

Advertising and misinformation are closely linked with false advertisements making exaggerated claims, while misleading ones using unprecise language or incomplete information to deceive consumers (Kumar & Shah, 2018). Advertising often uses emotional manipulation, fear or pseudoscience to appear credible, presents commercial messages as authentic through native and influencer ads, reinforces stereotypes shaping narrow views on gender, age, race, and identity and embeds value systems linking products to ideals like beauty, strength, and success.

With AI blurring the lines between reality and simulation, media literacy education is now more essential than ever to support students critically interpret, question and navigate the media. Due to its widespread presence across various platforms such as social media, YouTube, gaming apps and public spaces, advertising serves as a multimodal medium suitable for developing students' critical and digital literacy by revealing how different modes work together to shape reality and influence behavior.

Theoretical Framework: Multimodality and Critical Literacy

The term "Critical Literacy" (Janks, 2010; 2012) began to be systematically employed in language education in the early 1990s (Κουτσογιάννης, 2014). The concept of *critical literacy* is interpreted in diverse ways across different academic traditions, including genre-based pedagogy, the rhetorical tradition in North America, the multiliteracies framework among others. We can identify two major schools of thought related to Critical Literacy, grounded in distinct theoretical foundations: The one is North American version of Critical Literacy, rooted in the rhetorical tradition and the cognitive research on reading comprehension (Luke, 2012; Mitsikopoulou, 2013).

When the liberal-humanist view of literacy intersects with the rhetorical tradition, critical literacy focuses on uncovering fixed meaning and analyzing how language shapes persuasive arguments. Teaching practices focus on rhetorical strategies and logical structure, prioritizing textual analysis over sociopolitical context (Luke & Freebody, 1999; Mitsikopoulou, 2013). In contrast, a more socially and politically conscious approach to critical literacy rooted in Freire's radical pedagogy (Freire,1970), critical linguistics and studies on literacy issues (Κουτσογιάννης, 2014), views language as deeply connected to economic, social and cultural contexts. Reading extends beyond decoding texts to reading the world, questioning dominant narratives, uncovering hidden ideologies and analyzing how power, stereotypes and inequality are reproduced through language.

When integrated into multimodal analysis (O'Halloran, 2004; Unsworth, 2006), critical literacy intersects with theories of multimodality and semiotics (Bezemer & Kress, 2008, 2015; Jewitt, Bezemer & O'Halloran, 2016; Jewitt, 2008, 2009; Jewitt & Kress, 2003; Kress, 2008; Kress & van Leeuwen, 1996, 2001; Luke & Freebody, 1999; Serafini, 2013), highlighting how different semiotic resources interact to persuade targeted audiences, legitimize social inequalities and serve political or commercial interests (Djonov & Zhao, 2013; Duffelmeyer & Ellertson, 2005), while also helping students better understand how social phenomena such as racism, injustice and power are constructed and represented across multiple semiotic modes (Ajayi, 2015; Albers, Harste & Vasquez, 2015; Duffelmeyer & Ellertson, 2005; Huang, 2015; Way, 2014).

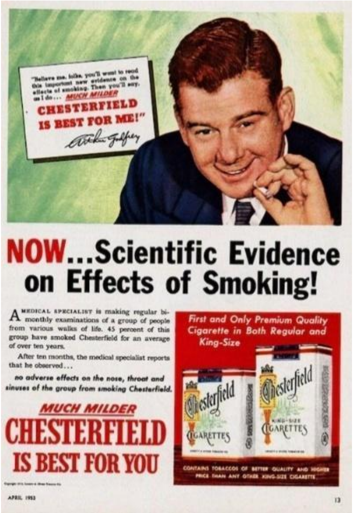
Critical multimodal pedagogy empowers students to re-design texts and create meaning using semiotic resources strategically, expressing their interests, challenging dominant discourses and envisioning alternative futures (Kress, 2010; Ajayi, 2009; Djonov & Zhao, 2013). In a rapidly changing world, education should go beyond analyzing texts to foster creativity, socially engaged communication and the development of students as active, critical meaning-makers - crucial for responsible, reflective citizenship (Kress, 2010).

Group-based student activities that foster critical literacy

The teaching activities presented here draw on diverse approaches to critical literacy and were implemented in the Modern Greek language class with a randomly selected sample of 20 first-grade high school students (aged 15–16) with varying cognitive levels. These activities aimed to equip students with tools to deconstruct manipulative strategies, question media intentions and make responsible choices as consumers and future citizens. Their effectiveness was assessed through pre- and post-tests measuring students' ability to identify persuasive techniques and to produce AI-assisted texts with well-documented analyses—skills not previously evident in the class.

Rhetorical Analysis: Develop awareness of rhetorical strategies used in texts and promote critical evaluation.

Through scaffolded questions, students analyze persuasive strategies in multimodal ads and uncover underlying ideologies, values and intentions. Drawing upon different semiotic resources, they identified techniques such as logical (logos), emotional (pathos), appeal to the credibility (Ethos) of the speaker (Walton, Reed & Macagno, 2008; Rapp, 2002), as well as challenges to opponent' s Ethos (Walton, 1998) and appeals to authority via celebrity or expert endorsements (Goodwin, 2011). They also examined ethically problematic ads that mislead through harmful product promotion, fake reviews, false expert opinions, hidden fees or exaggerated benefits (Figures 1 and 2). This process enhances their understanding of multimodal meaning-making, sharpens critical engagement with media, and raises awareness of strategies that shape audience perception.



"Believe me, folks, you'll want to read this important new evidence on the effects of smoking. Then you'll say, as I do... MUCH MILDER CHESTERFIELD IS BEST FOR ME!"

NOW...Scientific Evidence on Effects of Smoking!

A MEDICAL SPECIALIST is making regular bi-monthly examinations of a group of people from various walks of life. 45 percent of this group have smoked Chesterfield for an average of over ten years.

After ten months, the medical specialist reports that he observed...

no adverse effects on the nose, throat and sinuses of the group from smoking Chesterfield.

MUCH MILDER CHESTERFIELD IS BEST FOR YOU

First and Only Premium Quality Cigarette in Both Regular and King-Size

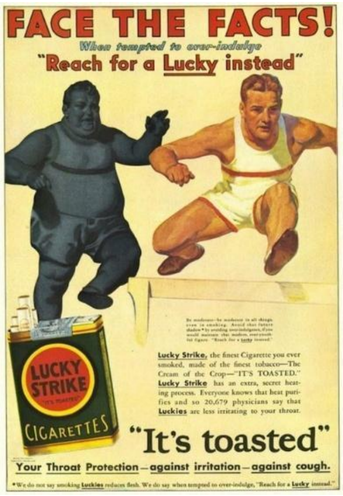
CONTAINS TOBACCOS OF BETTER QUALITY AND HIGHER PRICE THAN ANY OTHER KING-SIZE CIGARETTE

APRIL 1953

13

Figure 1

A vintage cigarette advertisement



FACE THE FACTS!

When tempted to over-indulge

"Reach for a Lucky instead"

Lucky Strike, the finest Cigarette you ever smoked, made of the finest tobacco—The Cream of the Crop—"IT'S TOASTED." Lucky Strike has an extra, secret heating process. Everyone knows that heat purifies and so 20,679 physicians say that Luckies are less irritating to your throat.

"It's toasted"

Your Throat Protection—against irritation—against cough.

*We do not say smoking Luckies reduces flesh. We do say when tempted to over-indulge, "Reach for a Lucky instead."

Figure 2

A 1930s Lucky Strike magazine

Source: Gazzetta (2015, August 7), Vintage διαφημίσεις τσιγάρων (pics)
Retrieved May 30, 2025, from https://www.gazzetta.gr/plus/article/784046/vintage-diafimiseis-tsigaron-pics

advertisement UK, 1930s.
Source: Bridgeman Images
Retrieved from Bridgeman Images. (n.d.). Retrieved May 30, 2025, from https://www.bridgemanimages.com/

Integrating Artificial Intelligence tools into text design activities, introduces students to novel and transformative approaches to critical text analysis. The AI-generated posters presented below, designed for a teenage audience, address social media addiction through emotional (Figure 3) and logical appeals (Figure 4). By creating such posters, students become familiar with the use of rhetorical techniques and develop a deeper understanding of how persuasive messages are constructed and conveyed.

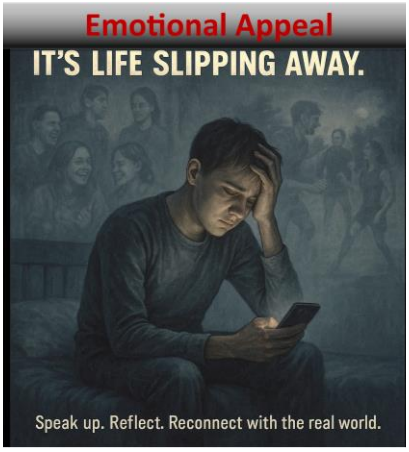


Figure 3
It's life is slipping away
Poster for the prevention of internet and social media teen's addiction.
AI generated image

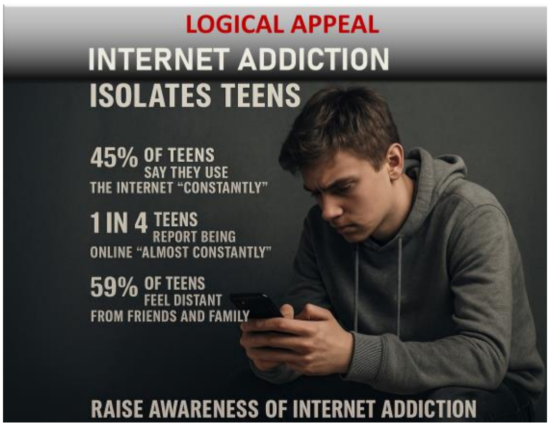
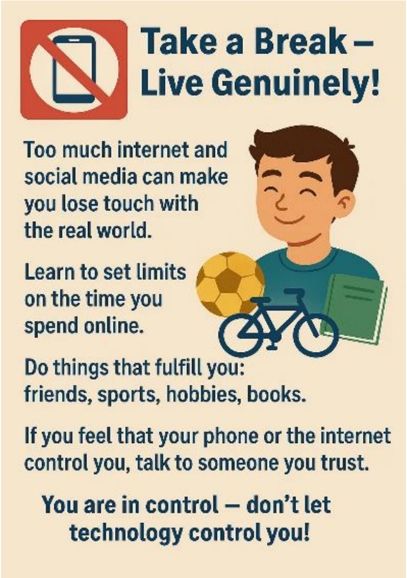

LOGICAL APPEAL
INTERNET ADDICTION
ISOLATES TEENS
45% OF TEENS SAY THEY USE THE INTERNET "CONSTANTLY"
1 IN 4 TEENS REPORT BEING ONLINE "ALMOST CONSTANTLY"
59% OF TEENS FEEL DISTANT FROM FRIENDS AND FAMILY
RAISE AWARENESS OF INTERNET ADDICTION

Figure 4
Internet addiction isolates teens
Poster for the prevention of internet and social media teen's addiction.
AI generated image

By analyzing two AI-created multimodal texts - one for teenagers (Figure 5) and one for parents (Figure 6) - students see how language and images shift by audience, recognize misleading techniques and build critical thinking to question ads and resist misinformation on issues like teen social media addiction.



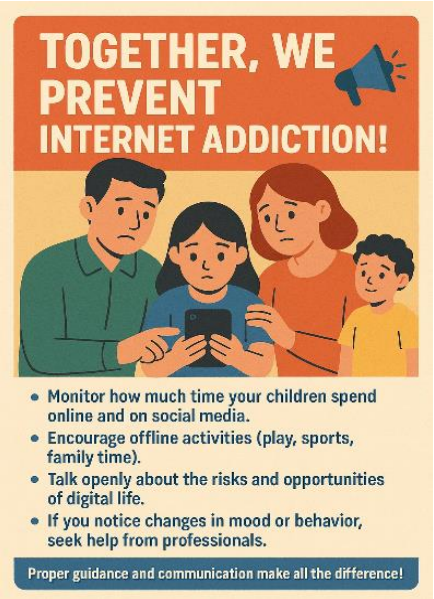
Too much internet and social media can make you lose touch with the real world.

Learn to set limits on the time you spend online.

Do things that fulfill you: friends, sports, hobbies, books.

If you feel that your phone or the internet control you, talk to someone you trust.

Figure 5
A poster for teen prevention of internet and social media addiction addressed to students.
AI-generated image

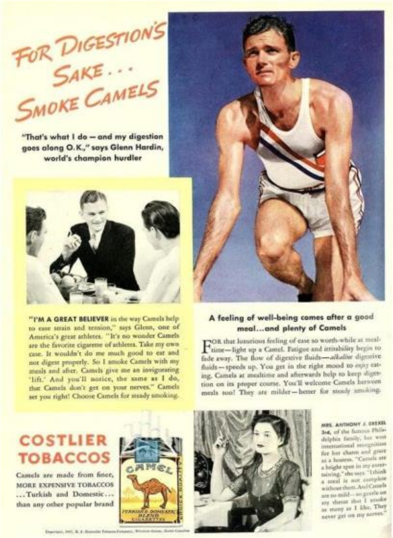


- Monitor how much time your children spend online and on social media.
- Encourage offline activities (play, sports, family time).
- Talk openly about the risks and opportunities of digital life.
- If you notice changes in mood or behavior, seek help from professionals.

Figure 6
A poster for teen prevention of internet and social media addiction addressed to parents.
AI-generated image

Focusing on the role of the historical and sociocultural context in shaping texts.
Critical literacy emphasizes the importance of historical, social and cultural contexts in shaping both the construction and meaning of texts (Luke & Freebody, 1999; Janks, 2010). More specifically, it recognizes that texts reflect the values, beliefs and power relations inherent in their sociocultural environments (Street, 2003). To enhance students' awareness of how historical context influences the content, structure, perspective and underlying values of a text, we engage them in the critical analysis of advertisements promoting the same product across different time periods (Comber & Simpson, 2001). The aim is to support students in thinking critically, questioning assumptions, and gradually developing their own informed perspective on the world (Janks, 2010; Luke, 2013).

By analyzing and comparing the images presented below (Figures 7, 8, 9 and 10), for instance, students gain a deeper understanding of how advancements in medical knowledge about smoking, evolving social attitudes toward smoking and stricter governmental policies, have influenced the way smoking and smokers are portrayed in advertising and anti-smoking campaigns. Smoking is no longer associated with independence, strength, health or personal charm, but rather with irresponsibility and a direct threat to health and life.



FOR DIGESTION'S SAKE... SMOKE CAMELS

"That's what I do — and my digestion goes along O.K.," says Glenn Hardin, world's champion hurdler

"I'M A GREAT BELIEVER in the way Camels help to ease strain and tension," says Glenn, one of America's great athletes. "It's no wonder Camels are the favorite cigarette of athletes. Take my own case. It wouldn't do me much good to eat and not digest properly. So I smoke Camels with my meals and after. Camels give me an invigorating 'lift.' And you'll notice, the same as I do, that Camels don't get on your nerves." Camels set you right! Choose Camels for steady smoking.

A feeling of well-being comes after a good meal...and plenty of Camels

FOR that luxurious feeling of ease so worth-while at mealtime—light up a Camel. Fatigue and irritability begin to fade away. The flow of digestive fluids—*alkaline* digestive fluids—speeds up. You get in the right mood to enjoy eating. Camels at mealtime and afterwards help to keep digestion on its proper course. You'll welcome Camels between meals too! They are milder—better for steady smoking.

COSTLIER TOBACCOS

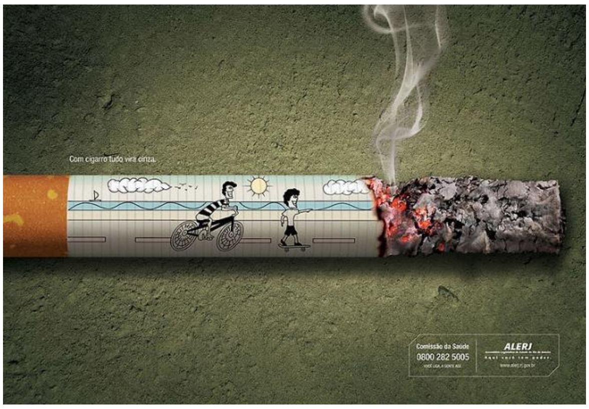
Camels are made from finer, MORE EXPENSIVE TOBACCOS ...Turkish and Domestic... than any other popular brand

MRS. ANTHONY J. DREXEL 3rd, of the famous Philadelphia family, has won international recognition for her charm and grace as a hostess. "Camels are a bright spot in my entertaining," she says. "I think a meal is not complete without them. And Camels are so mild—so gentle on my throat that I smoke as many as I like. They never get on my nerves."

Figure 7

For digestion's sake…smoke camels

Center for the Study of Tobacco and Society. (n.d.). Magazine advertisement by the R.J. Reynolds Tobacco Company for Camel cigarettes (1937) Used under fair use for educational purposes. University of Alabama CSTS.



Com cigarro tudo vira cinza.

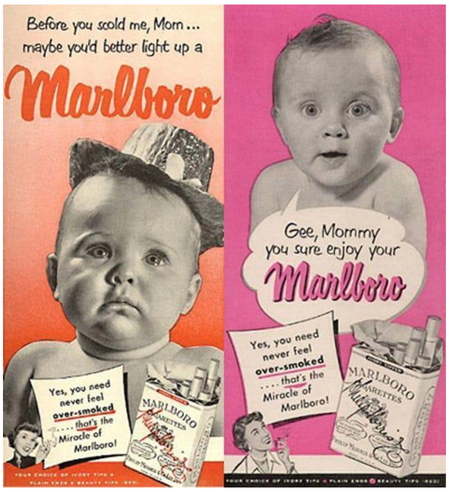
Comissão da Saúde
0800 282 5005
ALERJ

Figure 8

With cigarettes, your life goes to ashes.

Antismoking campaign poster

Reprinted from Antismoking campaigns that shock, by LiFO, 2012 Used under fair use for educational purposes LiFO, (2012).



Before you scold me, Mom... maybe you'd better light up a Marlboro

Yes, you need never feel over-smoked ...that's the Miracle of Marlboro!

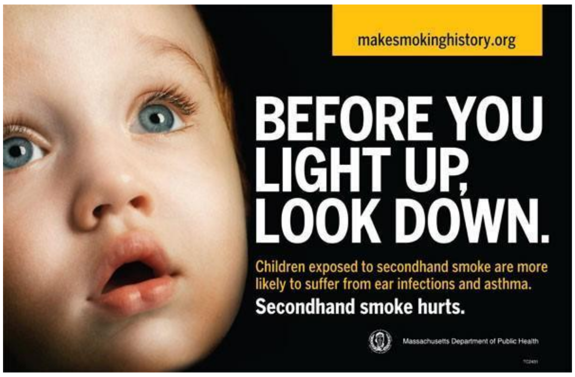
Gee, Mommy you sure enjoy your Marlboro

Yes, you need never feel over-smoked ...that's the Miracle of Marlboro!

Figure 9

Vintage cigarette ad of Marlboro

Used under fair use for educational purposes. Business Insider (2012).



makesmokinghistory.org

BEFORE YOU LIGHT UP, LOOK DOWN.

Children exposed to secondhand smoke are more likely to suffer from ear infections and asthma.

Secondhand smoke hurts.

Massachusetts Department of Public Health

Figure 10

Anti-Smoking Campaign

Massachusetts Health Promotion Clearinghouse

Used under fair use for educational purposes.

Furthermore, students can design advertisements addressed to imagined audiences of the present, the past (Figure 11) or even the future (Figure 12). During the design process, students are encouraged to modify parameters such as the historical context, the text’s purpose, the target audience and the values a text reflects, observing how these changes impact its language, tone and overall design.

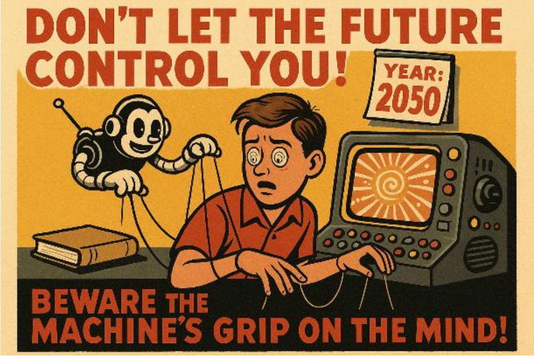

DON’T LET THE FUTURE CONTROL YOU!
YEAR: 2050
BEWARE THE MACHINE’S GRIP ON THE MIND!

Figure 11
Don’t let the future control you.
A retro-style awareness poster
AI generated image

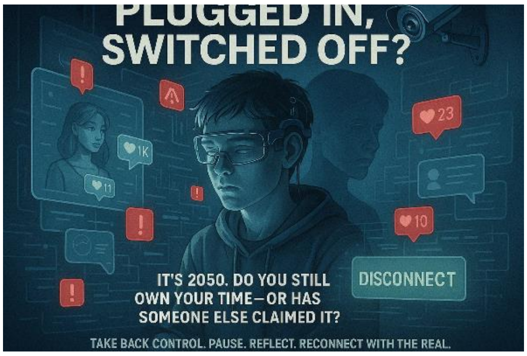
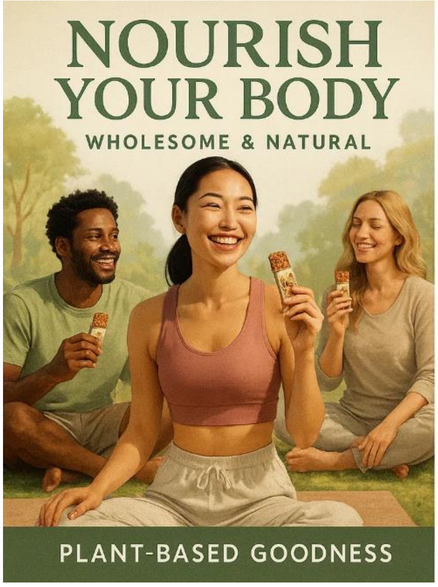

PLUGGED IN, SWITCHED OFF?
1K
11
23
10
IT'S 2050. DO YOU STILL OWN YOUR TIME—OR HAS SOMEONE ELSE CLAIMED IT?
DISCONNECT
TAKE BACK CONTROL. PAUSE. REFLECT. RECONNECT WITH THE REAL.

Figure 12
Plugged in, switched off?
A futuristic poster set in the year 2050
AI generated image

Advertising “Debate”: Exploring Dual Discourses

Students can develop critical literacy skills by designing, analyzing and comparing ads that promote similar products in different ways, linking these products to various lifestyle ideals, values and target audiences. For example, one AI-created advertisement (Figure 13) produced by students during a debate-based activity, depicts a diverse group in a natural setting, emphasizing community, well-being and a balanced lifestyle. Conversely, another ad (Figure 14) highlights luxury, personal success and self-reward through the image of an individual in an urban nightscape. By comparing these ads, students gain insight into the persuasive techniques used and the social values promoted.



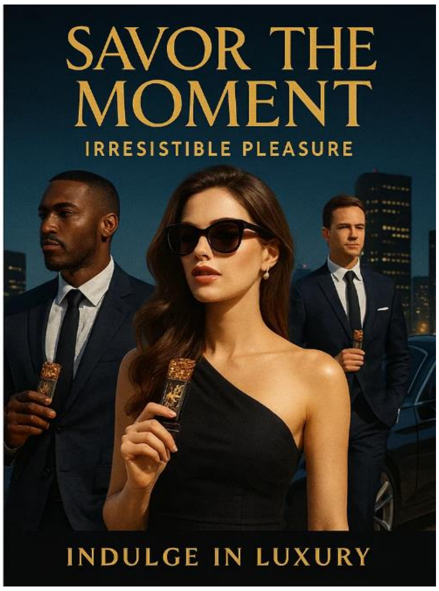


Figure 13

Nourish your body. Wholesome & Natural

OpenAI. (2025). AI generated image

Figure 14

Savor the Moment. Irresistible Pleasure

OpenAI. (2025). AI generated image

Media Truth Seekers: Evaluating Claims in Advertising

One of the key principles of Critical Literacy is to encourage students to approach texts as socially and ideologically constructed artifacts, exploring the accuracy and validity of product claims made in the advertisements, questioning the underlying assumptions and intentions embedded within them (Figures 15 and 16). Through systematic research and verification of information using credible sources, students become aware of how texts reflect and shape reality. Besides, students can actively participate in challenging deceptive or unethical advertisements through creative activities, such as writing critical responses and designing counter-texts (Figure 17) or videos to challenge misleading claims.

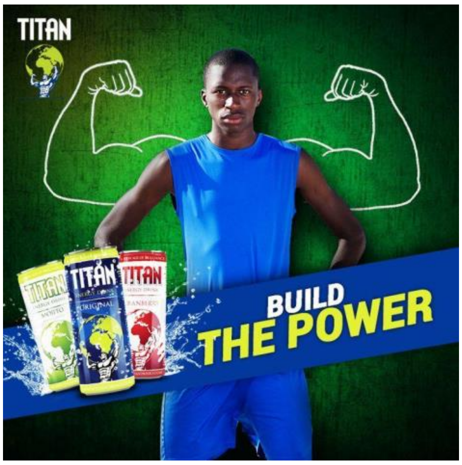
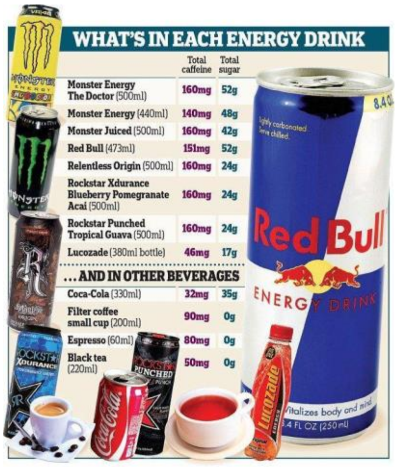


Figure 15

Build the power image from a Pinterest pin. Adapted from Pinterest (n.d.). ©Unknown.



WHAT'S IN EACH ENERGY DRINK

	Total caffeine	Total sugar
Monster Energy The Doctor (500ml)	160mg	52g
Monster Energy (440ml)	140mg	48g
Monster Juiced (500ml)	160mg	42g
Red Bull (473ml)	151mg	52g
Relentless Origin (500ml)	160mg	24g
Rockstar Xdurance Blueberry Pomegranate Acai (500ml)	160mg	24g
Rockstar Punched Tropical Guava (500ml)	160mg	24g
Lucozade (380ml bottle)	46mg	17g
...AND IN OTHER BEVERAGES		
Coca-Cola (330ml)	32mg	35g
Filter coffee small cup (200ml)	90mg	0g
Espresso (60ml)	80mg	0g
Black tea (220ml)	50mg	0g

Figure 16

What's in each energy drink?

From Grove, J. (2018, August 30). *Daily Mail.* https://www.dailymail.co.uk/news/article-6115457/May-faces-backlash-proposed-ban-sale-energy-drinks-schoolchildren.html

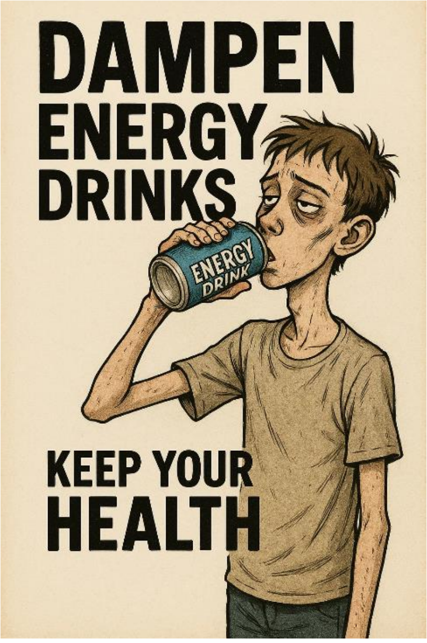


Figure 17
OpenAI. (2025). A counter advertisement of energy drinks challenging stereotypes AI-generated image.

Moreover, students can be encouraged to critically examine influencer advertisements on platforms such as Instagram, TikTok and YouTube, investigating the truthfulness and credibility of marketing claims made by celebrities, as well as the beauty stereotypes and consumer ideals embedded in such media. Students can also design critical versions of influencer ads (Figure 18), aiming to reveal their unrealistic expectations and deceptive nature, while reconsidering how media shape self.

"As it's shown" **"As it is"**

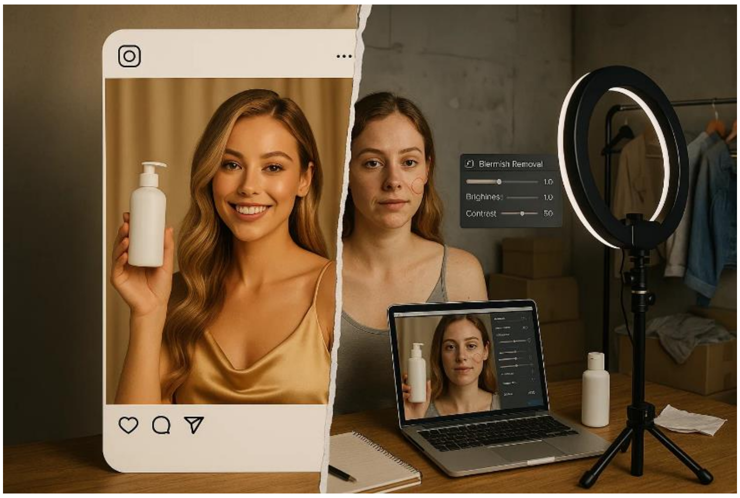


Figure 18
"As it'shown" vs "As it is".
OpenAI. (2025). Realistic image represents a "deconstructed" version of a social media influencer advertisement. [AI-generated image].

Changing point of view -Rewriting the Narrative to Challenge Representations and Express Alternative Perspectives of Reality.

As presented earlier, students can critically engage with advertisements by challenging dominant narratives - e.g., representing excluded groups, rewriting slogans, changing images or creating parodies. After identifying absent or stereotypically portrayed social groups in ads, they creatively redesign ads to include their perspectives and voices. For example, the images below (Figures 20 and 21), were created by students as part of a teaching activity focused on deconstructing gender stereotypes in a Greek toy TV- commercial. After analyzing the commercial, students redesigned a key snapshot (Figure 19) to challenge traditional gender roles.

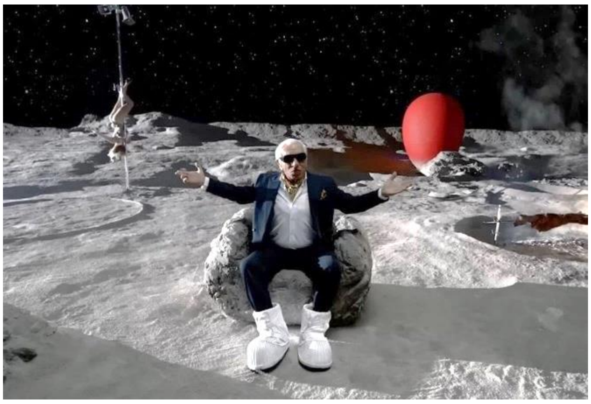


Figure 19

A still from the 2024 original Easter advertisement of the Greek toy company.

(Lifo Newsroom, 2023) © Lifo

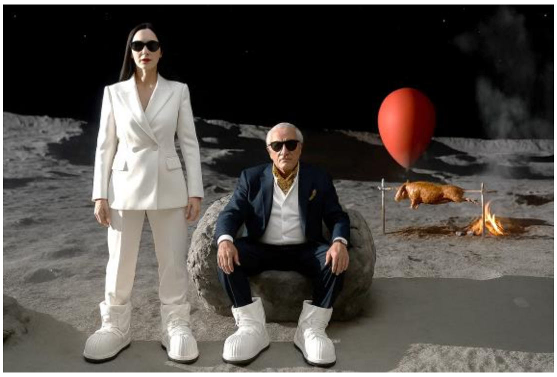


Figure 20

AI- recreated version of the original still from the 2024 Easter advertisement of the Greek toy company

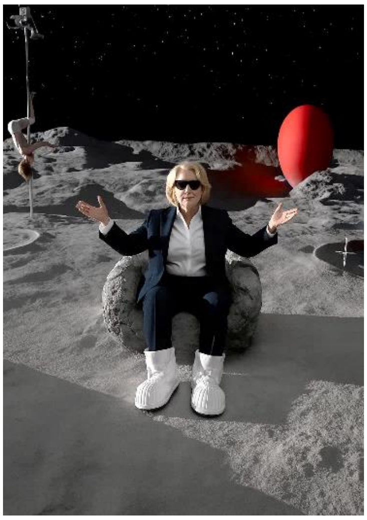


Figure 21

AI- recreated version of the original still from the 2024 Easter advertisement of the Greek toy company

The image below is taken from a Greek TV commercial by a well-known toy company featuring the slogan "Hit, hit like a man" (Figure 22). Though referencing a traditional Easter custom, the phrase can be interpreted as reinforcing gender-based violence.

Following classroom analysis, students used AI tools to redesign the image, aiming to challenge gender stereotypes. In one version (Figure 23), the slogan becomes "Show me respect like a real man," promoting respect over aggression. In another (Figure 24), a full gender-role reversal presents a male figure with the line "Hit like a woman," subverting traditional expectations.

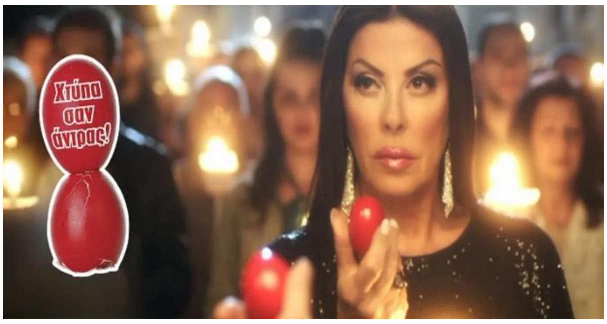


Figure 22

A still from the original advertisement featuring the controversial slogan *"Χτύπα σαν άντρας." ("Hit like a man")*

Source: Voria (2016, April 22). Αποσύρεται από τη διαφήμιση της Jumbo το «χτύπα σαν άντρας» [Still image]. Retrieved from https://www.voria.gr/article/aposirete-apo-ti-diafimisi-tis-jumbo-to-chtipa-san-antras

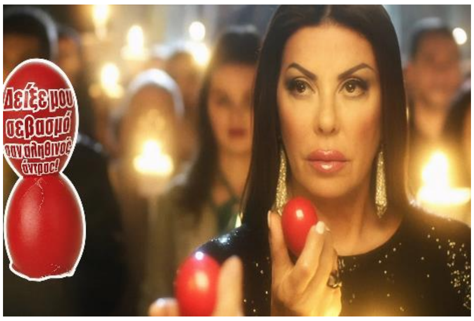


Figure 23

AI-recreated version of the original photo with the slogan *"Respect me like a real man".*

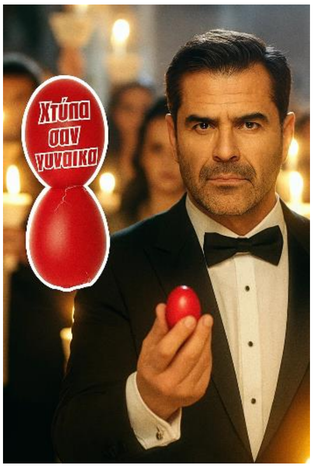


Figure 24

AI-recreated version of the original photo with the slogan *"Hit me like a woman".*

Conclusion

The implementation of these activities demonstrates how integrating AI into classroom practice can enhance critical literacy. For teachers, structured prompt-creation tasks provide a practical framework to combine language development with critical thinking and creative digital production. At the policy level, such approaches highlight the need for curricula that explicitly develop students' digital and critical literacy skills, preparing them to navigate increasingly complex media environments and engage thoughtfully with the world.

Beyond these benefits, through such activities students gain practical experience in creating effective prompts - a demanding design process that requires not only language skills in description and narration, but also critical literacy as well as visual and multimodal literacy skills. More specifically, by creating effective prompts, students improved their language abilities through precise and creative use of vocabulary, while strengthening their visual imagination by translating concepts into words. Furthermore, creating effective prompts empowered students' communication skills, as they should clearly define their goals and target audience and draw on multimodal resources of meaning-making to critically express their personal perspectives, values and ideologies. In addition, making prompts to challenge a representation, supported students' critical thinking through recognizing and questioning assumptions and creating well-reasoned, reflective responses. This process also fostered problem-solving skills, as students experimented, refined and improved their prompts.

Future research could explore the long-term impact of these interventions on students' ability to critically engage with media, evaluate information and apply these skills across diverse contexts, informing evidence-based strategies for integrating AI and digital literacies into education and contributing to a society guided not by fear or uncritical reliance on technology, but by informed, reflective and critical thinking.

References

Ajayi, L. (2009). English as a second language learners' exploration of multimodal texts in a junior high school. *Journal of Adolescent & Adult Literacy, 52* (7), 585-595.

Ajayi, L. (2015). Critical Multimodal Literacy: How Nigerian Female Students Critique Texts and Reconstruct Unequal Social Structures. *Journal of Literacy Research, 47* (2), 216-244.

Albers, P., Harste, J. C., & Vasquez, V. M. (2015). Critical and Multimodal Literacy Curricula. In M., Hamilton, R., Heydon, K., Hibbert, & R., Stooke (Eds). *Negotiating spaces for literacy learning: multimodality and governmentality* (pp. 115-130). Bloomsbury Publishing

Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication, 25*(2), 166–195. https://doi.org/10.1177/0741088307313177

Bezemer, J., & Kress, G. (2015). *Multimodality, learning and communication: A social semiotic frame*. Routledge.

Comber, B., & Simpson, A. (2001). Negotiating critical literacies in classrooms. *Reading Research Quarterly, 36*(2), 186–196.

Djonov, E., & Zhao, S. (2013). From multimodal to critical multimodal studies through popular discourse. In E. Djonov & S. Zhao (Eds.), *Critical multimodal studies of popular discourse* (pp. 1–14). Routledge.

Duffelmeyer, B. B., & Ellertson, A. (2005). WAC, WID, ECAC, CAC, CXC, LAC–VAC? Critical visual literacy: Multimodal communication across the curriculum. *Across the disciplines: Interdisciplinary perspectives on language, learning, and academic writing. https://wac.colostate.edu/docs/atd/visual/duffelmeyer_ellertson.cfm*

Freire, P. (1970). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). Continuum. (Original work published 1968)

Goodwin, J. (2011). Accounting for the appeal to the authority of experts. *Argumentation, 25*(3), 285–296. https://doi.org/10.1007/s10503-011-9206-5

Huang, S. Y. (2015). The intersection of multimodality and critical perspective: multimodality as subversion. *Language Learning & Technology, 19* (2), 21– 37. http://scholarspace.manoa.hawaii.edu/bitstream/10125/44428/1/19_03_action1.pdf

Janks, H. (2010). *Literacy and power*. Routledge.

Janks, H. (2012). The importance of critical literacy. *English Teaching: Practice and Critique, 11*(1), 150- 163. https://files.eric.ed.gov/fulltext/EJ970245.pdf

Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education, 32*(1), 241–267. https://doi.org/10.3102/0091732X07310586

Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. Routledge.

Jewitt, C., Bezemer, J., & O’Halloran, K. (2016). *Introducing multimodality*. Routledge.

Jewitt, C., & Kress, G. (Eds.). (2003). *Multimodal literacy*. Peter Lang.

Κουτσογιάννης, Δ. (2014). Κριτικοί Γραμματισμοί: Διεθνής Εμπειρία και Ελληνική Πραγματικότητα, in Ε. Γρίβα, Δ. Κουτσογιάννης, Κ. Ντίνας, Α. Στάμου, Α. Χατζηπαναγιωτίδη, & Σ. Χατζησαββίδης (eds), Πρακτικά Πανελλήνιου Συνεδρίου "Ο Κριτικός Γραμματισμός στη σχολική πράξη". Δράμα: http://www.nured.uowm.gr/drama/PRAKTIKA.html

Kress, G. (2008). *Literacy in the new media age*. Routledge.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G., & Van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. Hove, UK: Psychology Press.

Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.

Kumar, S., & Shah, N. (2018). False information on web and social media: A survey. *arXiv preprint.* https://arxiv.org/abs/1804.08559

Lajnef, K. (2023). The effect of social media influencers on teenagers’ behavior: An empirical study using cognitive map technique. *Current Psychology, 42*(22), 19364–19377. https://doi.org/10.1007/s12144-023-04273-1

Levine, S. (2023). *Protecting kids from stealth advertising in digital media: FTC staff perspective*. Federal Trade Commission. https://www.ftc.gov/system/files/ftc_gov/pdf/p214505kidsadvertisingstaffperspective092023.pdf

Luke, A., & Freebody, P. (1999). Further notes on the four resources model. *Reading Online, 3*(2). http://www.readingonline.org/articles/fourresources/

Luke, A. (2012). Critical Literacy: Foundational Notes, *Theory Into Practice*, *51*(1), 4-11.

Mitsikopoulou, B. (2013). *Rethinking online educational resources: Ideologies, pedagogies and identities.* Paradigm Publishers.

Luke, A. (2013). Defining critical literacy. In *Moving critical literacies forward* (pp. 19–31). Routledge.

O’Halloran, K. L. (Ed.). (2004). *Multimodal discourse analysis: Systemic functional perspectives*. Continuum.

Packer, J., Croker, H., Goddings, A. L., Boyland, E. J., Stansfield, C., Russell, S. J., & Viner, R. M. (2022). Advertising and young people’s critical reasoning abilities: Systematic review and meta-analysis. *Pediatrics*, *150*(6). https://doi.org/10.1542/peds.2022-057780

Pellegrino, A., & Stasi, A. (2024). A bibliometric analysis of the impact of media manipulation on adolescent mental health: Policy recommendations for algorithmic transparency. *Online Journal of Communication and Media Technologies*, *14*(4).https://doi.org/10.30935/ojcmt/202453

Radesky, J., Chassiakos, Y. L. R., Ameenuddin, N., & Navsaria, D. (2020). Digital advertising to children. *Pediatrics*, *146*(1). https://doi.org/10.1542/peds.2020-1681

Rapp, C. (2002). Aristotle’s rhetoric. In E. Craig (Ed.), *Routledge encyclopedia of philosophy*. Routledge.

Raffoul, A., Ward, Z. J., Santoso, M., Kavanaugh, J. R., & Austin, S. B. (2023). Social media platforms generate billions of dollars in revenue from US youth: Findings from a simulated revenue model. *Public Library of Science ONE*, *18*(12). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0295337

Serafini, F. (2013). *Reading the visual: An introduction to teaching multimodal literacy*. Teachers College Press.

Street, B. (2003). What’s ‘new’in new literacy studies? Critical approaches to literacy in theory and practice. *Current issues in comparative education*, *5*(2), 77–91.

Unsworth, L. (2006). *Analyzing multimodal texts*. Continuum.

Van Reijmersdal, E. A., & van Dam, S. (2020). How age and disclosures of sponsored influencer videos affect adolescents’ knowledge of persuasion and persuasion. *Journal of youth and adolescence*, *49*(7), 1531-1544. https://doi.org/10.1007/s10964-020-01291-6

Walton, D. (1998). *Ad hominem arguments*. University of Alabama Press.

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. Cambridge University Press.

Way, K. (2014). *Speaking Back to Structure: Critical Multimodal Media Literacy & The Politics of School Reform.* (Doctoral Dissertation, University of Massachusetts, Amherst). http://scholarworks.umass.edu/dissertations_2/136/

List of Figures

Figure 1: Gazzetta Team. (2015, August 7). *Vintage διαφημίσεις τσιγάρων (pics)*. Gazzetta.gr. https://www.gazzetta.gr/plus/article/784046/vintage-diafimiseis-tsigaron-pics

Figure 2: No Artist Known. (1930s). *Lucky Strike magazine advert, UK*. Bridgeman Images. https://www.bridgemanimages.com/en/noartistknown/lucky-strike-magazine-advert-uk-1930s/print/asset/3446634

Figure 3: OpenAI. (2025). *It's life slipping away.* Poster for the prevention of internet and social media teen's addiction. AI generated image

Figure 4: OpenAI. (2025), *Internet addiction isolates teens.* Poster for the prevention of internet and social media teen's addiction. AI generated image

Figure 5: OpenAI. (2025) *A poster for teen prevention of internet and social media addiction addressed to students* [Digital image]. AI-generated image. (n.d.).

Figure 6: OpenAI. (2025). *A poster for teen prevention of internet and social media addiction addressed to parents* [Digital image]. AI-generated image. (n.d.).

Figure 7: Center for the Study of Tobacco and Society. (n.d.). *Magazine advertisement by the R.J. Reynolds Tobacco Company for Camel cigarettes (1937)* [Advertisement]. University of Alabama CSTS. Used under fair use for educational purposes.

Figure 8: LiFO. (2012). *With cigarettes, your life goes to ashes* [Antismoking campaign poster]. In *Antismoking campaigns that shock*. Reprinted under fair use for educational purposes. https://www.lifo.gr/

Figure 9: Business Insider. (2012). *Vintage cigarette ad using a baby to market Marlboro as a calming product for women* [Advertisement]. Used under fair use for educational purposes.

Figure 10: Massachusetts Health Promotion Clearinghouse. (n.d.). *Before You Light Up Look Down* [Poster]. Massachusetts Department of Public Health. https://massclearinghouse.ehs.state.ma.us/PROG-TOB/TC2431kit.html

Figure 11: OpenAI. (2025). A retro-style awareness poster,AI generated image

Figure 12: OpenAI. (2025). A futuristic poster set in the year 2050 AI generated image

Figure 13: OpenAI. (2025). *Nourish your body. Wholesome & Natural*. AI generated image

Figure 14: OpenAI. (2025). *Savor the Moment. Irresistible Pleasure*. AI generated image

Figure 15: Unknown author. (n.d.). *Titan Energy Drink power image* [Pinterest pin]. Pinterest. Adapted by author.

Figure 16: Daily Mail. (2018, March 6). May faces backlash over proposed ban on sale of energy drinks to schoolchildren. *Daily Mail*. https://www.dailymail.co.uk/news/article-6115457/May-faces-backlash-proposed-ban-sale-energy-drinks-schoolchildren.html

Figure 17: OpenAI. (2025). *A counter advertisement of energy drinks challenging stereotypes* AI-generated image.
Figure 18 : OpenAI. (2025). Realistic image that represents a "deconstructed" version of a social media influencer advertisement. AI-generated image.
Figure 19: LiFO Newsroom. (2023, March 21). *Jumbo: Αντιδράσεις για την πασχαλινή διαφήμιση με Μπουγά – «Vegan, μακριά κι αγαπημένοι»* [Magazine advertisement screenshot]. LiFO.gr. https://www.lifo.gr/lifoland/its-viral/jumbo-antidraseis-gia-tin-pashalini-diafimisi-me-mpoyga-vegan-makria-ki
Figure 20: OpenAI. (2025). *AI- recreated version of the original still from the 2024 Easter advertisement of the greek toy company.*
Figure 21: OpenAI. (2025) *AI- recreated version of the original still from the 2024 Easter advertisement of the greek toy company.*
Figure 22: Voria. (2016, April 22). *Αποσύρεται από τη διαφήμιση της Jumbo το «χτύπα σαν άντρας»* [Still image]. https://www.voria.gr/article/aposirete-apo-ti-diafimisi-tis-jumbo-to-chtipa-san-antras
Figure 23: OpenAI. (2025). *AI-recreated version of the original photo with the slogan "Respect me like a real man".*
Figure 24: OpenAI. (2025) *AI-recreated version of the original photo with the slogan "Respect me like a real man".*

Enhancing Media Critical Thinking against Disinformation with a Socratic Chatbot: A Methodology

Panayiotis Kapetanakis[1], George Petasis[2], Dora Katsamori[3], Eirini Papachristou[4], Giannis Paparigopoulos[5]
NCSR ‘Demokritos’, Greece
1. pkapetanakis@iit.demokritos.gr 2. petasis@iit.demokritos.gr
3.dkatsamori@iit.demokritos.gr 4. e.papachristou@iit.demokritos.gr
5.ie.paparrigopoulos@iit.demokritos.gr

Abstract

Existing anti-disinformation interventions like fact-checking are often insufficient against manipulative content that exploits cognitive biases and emotional triggers. This paper presents a boosting intervention, the Socratic Coach, an AI chatbot developed by NCSR ‘Demokritos’ in the EU TITAN project. This ‘media thinking coach’ is designed to enhance users’ Media Critical Thinking (MCT), a hybrid of Media and Information Literacy (MIL) and Critical Thinking (CT). To ensure pedagogical consistency and avoid AI ‘hallucinations’, the system uses Retrieval-Augmented Generation (RAG) architecture. Unlike belief-centric bots that risk psychological reactance, the Socratic Coach directs inquiry toward actual media content, using a non-confrontational approach that emphasizes the recognition of manipulation tactics, through tactic-specific questioning and MIL heuristics, accuracy nudges, and metacognitive prompts. We present findings from a pilot study (N=12). Key findings confirm the Socratic method is a promising pedagogical tool but reveal a demand for AI transparency, a mismatch between the tool’s coaching function and user expectations for a fact-checker, and a need for more interactive dialogue. The study demonstrates the potential of combining Socratic pedagogy with controlled AI to create a transparent, tactic-labeled coaching tool that complements, not replaces, fact-checking by shifting the focus to manipulation recognition.

Keywords: critical thinking, media literacy, disinformation, AI chatbot, Socratic Method, psychological inoculation

Introduction

In the post-truth era, information disorders threaten public health, democratic integrity, and social cohesion by fueling vaccine hesitancy, eroding institutional trust, and amplifying polarization (McIntyre, 2018; Roozenbeek et al., 2023). Unlike simple falsehoods, disinformation agents leverage manipulative tactics—such as emotional appeals, conspiracism, and rhetorical framing—that exploit cognitive biases, distort reasoning, and resist binary fact-checking (Ecker et al., 2022; Kapantai et al., 2021). These strategies thrive in high-speed digital environments, spreading faster than corrections, reinforcing echo chambers and entrenching belief persistence (Flaxman et al., 2016; Vosoughi et al., 2018).

Existing interventions (Kozyreva et al., 2024) remain insufficient (Roozenbeek et al., 2023). Solutions like legislation, fact-checking and algorithmic moderation struggle with scalability, speed, and declining trust, sometimes provoking scepticism or reactance (Graves, 2018; Hoes et al., 2024; Nyhan & Reifler, 2010; Park & Yoon, 2025). They also fail to counter the illusory truth effect or superficial scrolling habits and reduce manipulation discernment (DeVerna et al., 2023; Kaye & Johnson, 2024; Udry & Barber, 2024). On the other hand, individual 'boosting' strategies—including generic critical thinking, gamified inoculation tools (e.g., *Bad News*), procedural frameworks like SIFT, and media literacy programs—show promise (Jones-Jang et al., 2021; Lantian et al., 2021; Orhan, 2023; Roozenbeek & van der Linden, 2019). However, they are often short-lived, context-dependent, and fragile when misinformation appeals to strong emotions or identity (Pennycook & Rand, 2021; Ziemer & Rothmund, 2024). In addition, without domain-specific heuristics, engagement strategies, and epistemic virtues such as open-mindedness, critical thinking alone may fail to transfer and can reinforce motivated reasoning (Kwek et al., 2023).

This highlights the need for a hybrid, AI-driven approach combining MIL heuristics and critical thinking (Caufield & Wineburg, 2023; McGrew & Breakstone, 2023; Paul & Elder, 2020) to improve durability, scalability, and the transfer of manipulativeness discernment skills across contexts (Bailin et al., 2019; Hitchcock, 2024; Horn & Veermans, 2019). This novel integration of MIL and critical thinking shapes what we term Media Critical Thinking (MCT), i.e., the active ability to discern manipulative from neutral media content (Maertens et al., 2023). To mitigate motivated reasoning and identity-protective cognition, MCT must also cultivate critical thinking dispositions through reflective self-monitoring (Koppel et al., 2023; Singh et al., 2025).

To operationalize MCT, we present the Socratic Coach, an AI-driven boosting intervention developed by NCSR Demokritos in the EU Horizon Europe TITAN project. Unlike automated fact-checkers or belief-focused Socratic bots (Costello et al., 2024; Meyer et al., 2024), the coach guides users to evaluate actual media content, emphasizing manipulation detection rather than worldview confrontation. In particular, the coach leverages AI to scale personalized guidance on manipulativeness discernment (Maertens et al., 2023; Geissler et al., 2025), critical reasoning (Machete & Turpin, 2020; Paul & Elder, 2016, 2020; Sartori et al., 2022), dispositions like humility and open-mindedness (Facione, 2015; Koppel et al., 2024; Swami et al., 2014), and metacognitive self-monitoring (Cohen et al., 2013; Lutzke et al., 2019).

Methodology

Psychological Pillars and MCT Outcomes

The intervention's primary goal is to develop Media Critical Thinking (MCT). Its methodology rests on three interrelated psychological pillars, which collectively support the development of MCT skills:

Nudging Analytical Thinking – Prompts users to slow down and engage System 2 reasoning, reflecting on content framing and cognitive processes (Kahneman, 2011; Swami et al., 2014).

Dialogical Inoculation – Builds cognitive 'antibodies' via passive prebunking and active dialogical prebunking (Socratic dialogue), encouraging users to recognize manipulation and generate their own counterarguments (Cook et al., 2017; Harjani et al., 2022).

Fostering Self-Monitoring – Promotes metacognitive reflection, helping users internalize critical dispositions and transfer analytical skills to future media encounters (Cohen et al., 2013; Gabaree, 2022; Lee & Ramazan, 2021).

These pillars operationalize MCT by integrating manipulativeness discernment, reflective reasoning, epistemic virtues, and the ability to critically evaluate sources and evidence.

The Knowledge Base

At the core of the Socratic Coach lies a robust Knowledge Base (KB), constructed using the 'Tactic Profiling' methodology. This framework systematically identifies and categorizes the warning signs of manipulation tactics, drawing from contemporary misinformation taxonomies and empirical research (Harjani et al., 2022; Kapantai et al., 2021; Kozyreva et al., 2020). Each tactic profile is structured around six elements: a precise definition, narrative structures, rhetorical ploys and logical fallacies, cognitive biases and emotional triggers exploited, intended agendas and targets, and preemptive refutations or forewarnings. Currently, the KB focuses on six primary tactics—Discrediting, Polarization, Conspiracism, Trolling, Pseudoscience, and Science Denialism—and highlights associated rhetorical, cognitive, and narrative cues, including ad hominem, emotional manipulation, sensationalist framing, clickbait, cherry-picked evidence, and diversionary reasoning.

Beyond tactic-specific content, the KB incorporates Media & Information Literacy (MIL) heuristics guiding users to evaluate sources across authorship, URL and platform characteristics, funding transparency, editorial policies, and activity patterns, informed by procedures such as SIFT (Caufield & Wineburg, 2023; McGrew & Breakstone, 2023). Learning is further supported through accuracy prompts featuring rationale-based nudges and forewarnings (Sharevski et al., 2022; Zavolokina et al., 2024), interactive MIL material tailored to varying literacy levels (Guess et al., 2020; Panizza et al., 2022) and metacognitive prompts (Singh et al., 2025).

A Socratic 'Inoculation Engine'

At the core of the Socratic Coach is a Dialogical Engine that transforms tactic-focused questioning into both active MCT and 'dialogical inoculation'. The process involves Socratic inquiry—interpreting messages, testing assumptions, evaluating reasoning, spotting fallacies, and considering alternatives (Facione, 1990, 2015; Neenan, 2008; Paul & Elder, 2012)—and is reinforced by MIL heuristics for assessing sources and content. Users not only learn to recognize manipulative strategies but are also shown why they are manipulative and how to refute them through counterarguments (Cook et al., 2017). By embedding preemptive refutations and explanations, the engine strengthens users' ability to generate their own counterarguments, cultivating cognitive resistance, analytical skill, and reflective habits (Harjani et al., 2022).

Dialogue Design and Dual Function

The dialogue uses a two-block structure to combine prebunking with active media-critical thinking (MCT) and dialogical inoculation. Block 1 provides passive prebunking via infographics and micro-lessons, highlighting disinformation warning signs with real-world examples from the Tactic Profiles. Block 2 applies Socratic questioning to guide structured analysis of both source ("WHO") and content ("WHAT"), prompting reflection on assumptions, reasoning, and manipulative strategies to detect manipulation and practice metacognitive skills.

This design allows the coach to serve as both an educational tool and a real-time assistant. As a training tool, it builds cognitive resilience, fosters self-monitoring, and guides users in evaluating source credibility and persuasive techniques. As a plugin, it analyzes media against the knowledge base and issues contextual alerts (e.g., "This article uses discredited tactics by labeling experts as corrupt"). Socratic metacognitive prompts act as cognitive speed bumps, shifting users from intuitive to deliberative System 2 thinking (Kahneman, 2011; Paul & Elder, 2012; Singh et al., 2025) while encouraging reflection and counterargument generation (Harjani et al., 2022).

AI Model Rationale and Design Considerations

The Socratic chatbot ensures pedagogical consistency and reduces hallucination by using a large language model (LLM) within a Retrieval-Augmented Generation (RAG) framework (Gao et al., 2024). All materials, including Socratic questions and tactic profiles, are stored as vector embeddings in a curated Chroma knowledge base. The LLM can interact with users by referencing this base as a source and context. For the LLMs, LLaMa 3.1 8B Instruct from Meta (initially) and Gemma 3 12B Instruct from Google (currently) were deployed through LangChain and VLLM. These models were selected for being state-of-the-art, open-source, and multilingual, supporting a 128K token context for efficient data analysis. Gemma 3 is currently being utilized as it better supports languages other than English, e.g., Greek. Combining LLM with RAG enables engaging and coherent multi-turn Socratic dialogue, ensuring robust, safe, and transparent interactions aligned with the chatbot's cognitive coaching goals. Larger LLMs from OpenAI were not chosen due to licensing costs and reproducibility concerns important for an EU-funded educational initiative.

Evaluation Methodology

Research Questions

The evaluation of the Socratic chatbot was guided by four research questions:

1. Expectations: What are users’ expectations when using an interactive AI-driven chatbot for evaluating news content?
2. Relevance: How relevant do users find a Socratic chatbot’s questions in helping them detect manipulation and critically analyze media items?
3. Experience and Usability: How do users evaluate the experience, usability, and value of interacting with an AI-driven Socratic chatbot?
4. Pedagogical Alignment: Does the chatbot’s dialogue design align with users' needs for refuting manipulative arguments and enhancing their critical thinking skills?

Data Collection and Analysis

To answer these questions, we employed a mixed-methods design centered on a qualitative, co-creation and evaluation workshop. The 2-hour session took place on October 15, 2024, at the National Centre for Scientific Research "Demokritos" (NCSR-D) in Athens, Greece. The study followed a pre-test/post-test structure across four stages:

1. Pre-Test: Participants completed a demographic and media use questionnaire along with a baseline score on the Manipulative Online Content Recognition Inventory (MOCRI; Maertens et al., 2023).
2. Brainstorming session: A focus group discussion elicited user expectations and initial views on misinformation interventions.
3. Hands-On Interaction: Participants directly engaged with the chatbot prototype.
4. Post-Test: Participants completed the MOCRI again and answered open-ended and Likert-scale questions about their experience.

Data from the focus group and open-ended questions were transcribed and thematically analyzed to identify recurring themes related to user experience and suggestions for improvement. Due to the small sample size (N=12) and post-test attrition (N=8 at post-test), no inferential statistical analysis was performed on the MOCRI data. Instead, the quantitative data collection served as a successful methodological pilot for a future, larger-scale evaluation.

Participant Selection

The workshop involved a convenience sample of 12 volunteers from various departments within the NCSR ‘Demokritos’ research center (see Appendix). This sample was not intended to be representative of the general population. The aim was to obtain detailed, high-literacy feedback on a prototype in early development. The sample was 66.7% female (n=8), with 50% aged 41–50 (n=6). Most participants held graduate or professional degrees (75%, n=9).

Findings

Given the exploratory nature of this research, the following results should be read as qualitative insights to guide refinement and larger-scale testing, not as generalizable conclusions.

Finding 1: Users demand transparency to build trust. A consistent theme was the need for meta-information about the tool beyond its clean interface. Participants' trust hinged on understanding the development team, funding, and training data, and they valued "evaluation from other users" as social proof. The emphasis on backend disclosure suggests that credibility depends on transparency rather than surface-level design choices.

Finding 2: Socratic pedagogy is promising but mismatches fact-checker expectations. Most participants (9 of 12) found the chatbot's questions "Quite Relevant" or "Very Relevant," describing them as a "refreshing approach" that encouraged deeper reflection. Several commented that they "learnt new things" and appreciated the active role. However, some still expected a conclusive verdict: "I would like to have an overall answer whether the article is intended to misinform us." This highlights a tension between the tool's coaching role and user desire for closure, making expectation management essential.

Finding 3: Usable, but dialogue needs more interactivity. Participants rated clarity positively but noted fatigue, reporting they "started to get bored/tired after 15 minutes." Conversational issues included looping, abrupt answers, and repetition. A central critique was limited agency: "Allow me to place questions and guide the conversation according to my own remarks." They also suggested integrating hints, prompts, and concise summaries of takeaways to sustain engagement.

Finding 4: Pre/post-test evaluation is viable. The use of MOCRI proved feasible for capturing changes in manipulativeness discernment. For example, one participant who misclassified a neutral headline in the pre-test correctly identified it post-test, illustrating learning effects. While anecdotal, such shifts suggest that Socratic coaching can foster discernment and demonstrate that MOCRI is a sufficiently sensitive instrument for future, larger-scale testing.

Discussion

This study evaluated the AI-driven, Socratic Coach. Our findings align with the literature on the limits of existing interventions: users demand transparency (Finding 1) in line with AI trust constraints (Graves, 2018; Park & Yoon, 2025), and report cognitive fatigue and a desire for simple verdicts (Findings 2–3), reflecting effort-dependent limits of boosting strategies (Jeong et al., 2012; Ziemer & Rothmund, 2024). Against this backdrop, the Socratic Coach offers a complementary path: it provides Socratic guided discovery and reflective thinking to cultivate manipulativeness discernment via tactic-grounded questioning, accuracy nudges, and metacognitive prompts.

Findings and Positioning

Transparency as a Prerequisite for Trust: Participants' trust hinged on provenance and operational clarity, mirroring documented trust erosion toward opaque interventions. To model epistemic accountability, the Socratic Coach will surface meta-information in-product, including its development team, funding, RAG architecture, and data sources. This positions the Socratic Coach apart from AI fact-checkers, which face challenges of hallucination and authority-based skepticism (Hoes et al., 2024).

Socratic Pedagogy vs. Demand for Fact-checks: While most users valued Socratic questioning, some wanted a binary answer. Our solution is a middle path: the Coach presents tactical labels (e.g., conspiracism) paired with rationale and optional microlessons. This preserves pedagogical aims (Paul & Elder, 2012): it provides the verdict-like clarity users seek and functions as a technique-based prebunk (Cook et al., 2017). This approach also distinguishes the Socratic Coach from gamified inoculation tools like *Bad News* (Roozenbeek & van der Linden, 2019) and belief-centric Socratic bots like *DebunkBot* (Costello et al., 2024) and *Street Epistemologist* (Meyer et al., 2024), which use role-play or target users' beliefs, respectively.

Interactivity, Effort, and Flow: Reports of cognitive fatigue and constrained agency echo critiques that boosting is often effortful. Because our approach explicitly tries to slow down System-1 processing, we will support the effort with accuracy labels, metacognitive prompts, branching controls.

MOCRI as an Evaluation Instrument: Pre/post shifts suggest that MOCRI is sensitive to change, supporting the argument that domain-specific measures outperform generic CT scales for misinformation outcomes (Horn & Veermans, 2019). Future trials will retain MOCRI as the primary outcome, adding secondary measures of dispositions and transfer, with delayed post-tests to examine durability.

Limitations, Applications, and Future Directions

The study's findings are illustrative, not generalizable, due to the small, non-representative sample (n=12, highly educated). The co-creation setting may have induced social desirability bias. Key limitations emerged from feedback: the semi-scripted dialogue was perceived as rigid, and a clearer onboarding process is needed to manage the expectation that the tool is a fact-checker rather than a cognitive coach. The knowledge base's reliance on pre-defined tactic profiles is also inherently static and requires continuous updates.

Despite these limitations, the Socratic Coach offers a model for a scalable cognitive infrastructure. Its modular design allows for integration into various contexts, serving as a tool for educators, a pre-publication audit system for newsrooms, a public resource for policymakers and civic NGOs, and as a browser plugin for the general public. "Future work will develop AI-driven psychoeducational interventions—using Socratic prompts and behavioral nudges—to raise awareness of cognitive and socio-affective drivers of misinformation and support changes in evaluative behavior. A randomized controlled trial will assess the Socratic coach as a whole. Broader impact will require pursuing accessibility for diverse populations and strengthening trust in AI through transparent labeling.

References

Bailin, S., Case, R., Coombs, J.R., & Daniels, L.B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies, 31*(3), 285–302. https://doi.org/10.1080/002202799183133

Caulfield, M., & Wineburg, S. (2023). *Verified: How to think straight, get duped less, and make better decisions about what to believe online.* University of Chicago Press.

Cohen, J. S., Edmunds, J. M., Brodman, D. M., Benjamin, C. L., & Kendall, P. C. (2013). Using self-monitoring: Implementation of collaborative empiricism in cognitive-behavioral therapy. *Cognitive and Behavioral Practice*, *20*(4), 419–428. https://doi.org/10.1016/j.cbpra.2012.07.001

Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *Public Library of Science ONE*, *12*(5). https://doi.org/10.1371/journal.pone.0175799

Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, *385*(6714). https://doi.org/10.1126/science.adq1814

DeVerna, M. R., Yan, H. Y., Yang, K.-C., & Menczer, F. (2023). Fact-checking information from large language models can decrease headline discernment. *arXiv preprint.* https://doi.org/10.48550/arXiv.2308.10800

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., & Fazio, L. K. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*, 13–29. https://doi.org/10.1038/s44159-021-00006-y

Facione, P.A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Research findings and recommendations prepared for the Committee on Pre-College Philosophy of the American Philosophical Association. ERIC Document ED315423. https://eric.ed.gov/?id=ED315423

Facione, P.A. (2015). *Critical Thinking: What It Is and Why It Counts*. Measured Reasons LLC.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly, 80*(S1), 298–320. https://doi.org/10.1093/poq/nfw006

Gabaree, S. (2022). A metacognitive approach to reduce the spread of online misinformation. *The Journal of Media Literacy.* International Council for Media Literacy. https://ic4ml.org/journal-article/a-metacognitive-approach-to-reduce-the-spread-of-online-misinformation/

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Wang, H. (2024). Retrieval-Augmented Generation for Large Language Models: A Survey. *arXiv preprint.* http://arxiv.org/abs/2312.10997

Geissler, F., Robertson, C., & Feuerriegel, S. (2025). Digital literacy interventions can boost humans in discerning deepfakes. *arXiv preprint*. https://doi.org/10.48550/arXiv.2507.23492

Graves, L. (2018). Understanding the promise and limits of automated fact-checking: Journalistic fact-checking and automated detection of misinformation. *New Media & Society*, *20*(5), 1405-1422. https://doi.org/10.1177/1461444817734057

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences of the United States of America, 117*(27), 15536–15545. https://doi.org/10.1073/pnas

Harjani, T., Roozenbeek, J., Biddlestone, M., van der Linden, S., Stuart, A., Iwahara, M., Piri, B., Xu, R., Goldberg, B., & Graham, M. (2022). *A Practical Guide to Prebunking Misinformation*. University of Cambridge.

Hitchcock, D. (2024). Critical thinking. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. Stanford University. https://plato.stanford.edu/archives/sum2024/entries/critical-thinking/

Hoes, E., Aitken, B., Zhang, J., Gackowski, T., & Wojcieszak, M. (2024). Prominent misinformation interventions reduce misperceptions but increase scepticism. *Nature Human Behaviour, 8*(1), 1545 - 1553. https://doi.org/10.1038/s41562-023-01765-4

Horn, S., & Veermans, K. (2019). Critical thinking efficacy and transfer skills defend against 'fake news' at an international school in Finland. *Journal of Research in International Education*, *18*(1), 23-41. https://doi.org/10.1177/1475240919830003

Jeong, S. H., Cho, H., & Hwang, Y. (2012). Media Literacy Interventions: A Meta-Analytic Review. *The Journal of communication*, *62*(3), 454–472. https://doi.org/10.1111/j.1460-2466.2012.01643.x

Jones-Jang, S.M., Mortensen, T., & Liu, J. (2021). Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't. *American Behavioral Scientist, 65*(2), 371-388. https://doi.org/10.1177/0002764219869406

Kahneman, D. (2011). *Thinking, fast and slow*. Penguin.

Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, *23*(5), 1301-1326. https://doi.org/10.1177/1461444820959296

Kaye, B. K., & Johnson, T. J. (2024). I can’t stop myself! Doomscrolling, conspiracy theories, and trust in social media. *Atlantic Journal of Communication, 32*(3), 471–483. https://doi.org/10.1080/15456870.2024.2316844

Koppel, L., Robertson, C. E., Doell, K. C., Javeed, A. M., Rasmussen, J., Rathje, S., Vlasceanu, M., & Van Bavel, J. J. (2023). Individual-level solutions may support system-level change if they are internalized as part of one's social identity. *Behavioral and Brain Sciences, 46*, e165. https://doi.org/10.1017/S0140525X23000318

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Science and pseudoscience: Mapping misconceptions and developing effective countermeasures. *Proceedings of the National Academy of Sciences, 117*(48), 30270–30280. https://doi.org/10.1073/pnas.2008963117

Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., et al. (2024). Toolbox of individual-level interventions against online misinformation. *Nature Human Behaviour, 8*, 1044–1052. https://doi.org/10.1038/s41562-024-01881-0

Kwek, A., Peh, L., Tan, J., & Lee, J. X. (2023). Distractions, analytical thinking and falling for fake news: A survey of psychological factors. *Humanities & social sciences communications*, *10*(1). https://doi.org/10.1057/s41599-023-01813-9

Lantian, A., Bagneux, V., Delouvée, S., & Gauvrit, N. (2021). Maybe a free thinker but not a critical one: High conspiracy belief is associated with low critical thinking ability. *Applied Cognitive Psychology, 35*(3), 674-684. https://doi.org/10.1002/acp.3790

Lee, D. K. L., & Ramazan, O. (2021). Fact-Checking of Health Information: The Effect of Media Literacy, Metacognition and Health Information Exposure. *Journal of health communication*, *26*(7), 491–500. https://doi.org/10.1080/10810730.2021.1955312

Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change, 58*. https://doi.org/10.1016/j.gloenvcha.2019.101964

Machete, P., & Turpin, M. (2020). The Use of Critical Thinking to Identify Fake News: A Systematic Literature Review. In *Responsible Design, Implementation and Use of Information and Communication Technology, 12067*, 235 - 246. Springer. https://doi.org/10.1007/978-3-030-44999-5_20

Maertens, R., Roozenbeek, J., & van der Linden, S. (2023). The Manipulative Online Content Recognition Inventory (MOCRI): A scale to measure the ability to spot manipulation techniques in online content. *PsyArXiv peprint*. https://doi.org/10.31234/osf.io/g68mc

Martel, C., Pennycook, G., & Rand, D.G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications, 5*(1), Article 1. https://doi.org/10.1186/s41235-020-00252-3

McGrew, S., & Breakstone, J. (2023). Civic online reasoning across the curriculum: Developing and testing the efficacy of digital literacy lessons. *AERA Open, 9*, 1–16. https://doi.org/10.1177/23328584231176451

McIntyre, L. (2018). *Post-truth*. MIT Press.

Meyer, M., Enders, A., Klofstad, C., Stoler, J., & Uscinski, J. (2024). Using an AI-powered "street epistemologist" chatbot and reflection tasks to diminish conspiracy theory beliefs. *Harvard Kennedy School (HKS) Misinformation Review, 5*(6). https://doi.org/10.37016/mr-2024-070

Neenan, M. (2008). Using Socratic questioning in coaching. *Journal of Rational-Emotive & Cognitive-Behavior Therapy, 27*(4), 249–264. https://doi.org/10.1007/s10942-007-0076-z

Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior 32*(2), 303-330. https://doi.org/10.1007/s11109-010-9112-2

Orhan, A. (2023). Fake news detection on social media: The predictive role of university students' critical thinking dispositions and new media literacy. *Smart Learning Environments, 10*(29), 1–15. https://doi.org/10.1186/s40561-023-00248-8

Panizza, F., Ronzani, P., Martini, C., Mattavelli, S., Morisseau, T., & Motterlini, M. (2022). Lateral reading and monetary incentives to spot disinformation about science. *Scientific Reports*, *12*(1), Article 5678. https://doi.org/10.1038/s41598-022-09168-y

Paul, R., & Elder, L. (2012). *The miniature guide to critical thinking: Concepts and tools.* Foundation for Critical Thinking Press.

Paul, R., & Elder, L. (2016). *The thinker's guide to the art of Socratic questioning*. Foundation for Critical Thinking.

Paul, R., & Elder, L. (2020). *Fact over fake: A critical thinker's guide to media bias and political propaganda*. Rowman & Littlefield.

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications, 5*(65). https://doi.org/10.1057/s41599-019-0279-9

Roozenbeek, J., Traberg, C., & van der Linden, S. (2022). Correction: Technique-based inoculation against real-world misinformation (print 2023), by Roozenbeek et al. *Royal Society Open Science, 10*. https://doi.org/10.1098/rsos.231235

Sartori, R., Tommasi, F., Ceschi, A., Falser, M., Genero, S., & Belotto, S. (2022). Enhancing critical thinking skills and media literacy in initial vocational education and training via self-nudging: The contribution of NERD VET project. *Frontiers in Psychology, 13.* https://doi.org/10.3389/fpsyg.2022.935673

Sharevski, F., Devine, A., Pieroni, E., & Jachim, P. (2022). Meaningful context, a red flag, or both? Preferences for enhanced misinformation warnings among US Twitter users. In *Proceedings of the 2022 European Symposium on Usable Security (EuroUSEC 2022)* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3549015.3555671

Singh, A., Guan, Z., & Rieh, S. Y. (2025). Enhancing critical thinking in generative AI search with metacognitive prompts. *arXiv preprint.* https://doi.org/10.48550/arXiv.2505.24014

Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories*. Cognition, 133*(3), 572–585. https://doi.org/10.1016/j.cognition.2014.08.006

Udry, J., & Barber, S. J. (2024). The illusory truth effect: A review of how repetition increases belief in misinformation. *Current Opinion in Psychology, 56.* https://doi.org/10.1016/j.copsyc.2023.101736

Vosoughi, S., Roy, D. & Aral, S. (2018). The spread of true and false news online. *Science,* 359, 1146-1151. https://doi.org/10.1126/science.aap9559

Zavolokina, L., Sprenkamp, K., Katashinskaya, Z., Jones, D. G., & Schwabe, G. (2024). Think fast, think slow, think critical: Designing an automated propaganda detection tool. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–14). ACM. https://doi.org/10.1145/3613904.3642805

Ziemer, C.-T., & Rothmund, T. (2024). Psychological underpinnings of misinformation countermeasures: A systematic scoping review. *Journal of Media Psychology: Theories, Methods, and Applications*, 36(6), 397–409. https://doi.org/10.1027/1864-1105/a000407

Disclaimer

This paper expresses the views and opinions of the author(s) only and does not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Funding

TITAN project has received funding from the EU Horizon 2020 research and innovation programme under grant agreement No.101070658, and by UK Research and innovation under the UK government's Horizon funding guarantee grant numbers 10040483 and 10055990. (https://www.titanthinking.eu/)

Competing interests

The author(s) declare(s) no competing interests.

Ethics

The study received ethical approval from the Institutional Review Board (IRB) at NCSR 'Demokritos'. All participants were fully informed about the study's purpose, its voluntary nature, and confidentiality safeguards, and provided written informed consent prior to participation. Data anonymization was ensured through the use of self-generated four-digit personal codes to link the pre- and post-test questionnaires without collecting personally identifiable information.

Appendix 1. Table 1. Participant Demographics (N=12)

Characteristic	Category	Number of Participants	Percentage of Total
Gender	Female	8	66.7%
	Male / Other	4	33.3%
Age Group	41-50 years	6	50%
	Other Age Groups	6	50%
Education Level	Graduate or Professional Degree	9	75%
	Undergraduate Degree or Lower	3	25%

Appendix 2: The Questioning Process

Educational Example Dialogue

This dialogue process dynamically adapts to users' media literacy levels, ensuring an engaging and educational experience.

1. **User Input:**

Users upload or reference a media item (e.g., article or social media post) for evaluation. Users specify whether they want to analyze the **source ('WHO')**, **content ('WHAT')**, or both.

2. **Introduction:**

The chatbot introduces itself as a coach in media critical thinking, explaining its key features. Users can choose to start with either the 'WHO' or the 'WHAT' section.

3. Evaluating the 'WHO' (Medium/Author):

- **"WHO" Questions (Tailored to Media Type):**
    - For websites: e.g., *"Is the website a trustworthy, established source?", "Is the author a real person with a real name?", "Does the author have any actual expertise in this topic?"*
    - For social media: e.g., *"Does the account have a verification badge?"*

- **Support Mechanisms** (Activated Based on User Input):
    - ***Rationales***: When users ask why a question is relevant, the chatbot provides context.
        - ***Chatbot**: "Is the author a real person, or are they using a fake name?"*
        - ***User**: "Why does that matter?"*
        - ***Chatbot**: "Well, if the author is hiding behind a fake name or doesn't have verifiable credentials, it could be a sign of something shady—like trolls, bots, or even organised groups trying to spread propaganda or misinformation. It's a tactic often used in influence campaigns to mislead or manipulate people."*
    - ***Media Literacy Tips***: If users find it difficult to spot signs of deception, the chatbot can offer step-by-step tips, clear definitions, or practical examples to help them navigate the information more effectively.
        - ***User**: "I'm not sure how to figure out if the author is a real person."*
        - ***Chatbot**: "No problem! Here are a few steps to help you verify:*
          a. Check their bio for detailed background information—real authors usually have clear credentials.
          b. Look for a professional social media presence or contact details, like an email.
          c. Confirm their institutional affiliation through official websites.
          d. Search for their name in reputable publications or third-party mentions to see if they're recognized elsewhere."
        - ***User**: "Hmm, I didn't really check the website. What should I be looking for?"*

        - **Chatbot**: *"Great question! Trustworthy sites tend to have clear editorial standards and a history of reliable reporting. They often have well-established reputations and citations. If the site is obscure or seems designed to push a specific agenda, it's worth looking into more carefully. Always check if they've been recognized by other reputable sources.*
    - ***Warnings & Preemptive Refutations***: If users tend to distrust mainstream verifiction methods and dismiss traditional markers of credibility, the chatbot provides nuanced insights to address their concerns.
        - ***User***: *"I trust alternative or independent voices and media."*
        - ***Chatbot***: *"Independent perspectives can be valuable, but expertise is key. Watch out for vague titles like 'independent researcher,' 'health guru,' or 'truth warrior'—these often lack credible evidence or professional backing."*
    - **Trust Reflection**:
        - At the end of the 'WHO' section, the chatbot summarizes findings and asks, "Based on this evaluation, do you find the source trustworthy?"
        - Depending on user feedback, they can proceed to the 'WHAT' section or conclude the evaluation.

4. Evaluating the WHAT (Content)

- **'WHAT' Questions (Based on chatbot's detected signs):**
    - Prompts user to identify the central claim or message of the media item and its purpose.
        - ***Chatbot***: *"Let's begin by figuring out the main point this article is trying to make. What's the core claim here?"*
        - ***Chatbot***: *"Does this article seem to be pushing a particular agenda, narrative or ideology? Is there any indication that the author might have a bias, or are they simply presenting the facts? Who might benefit from this message being spread?*
    - Uses follow-up questions to help user spot the detected signs of disinformation in the headlines and content of the article/post.
        - ***Chatbot***: *"Now, let's take a closer look at the headline. Does it seem overly dramatic or like clickbait?*
        - ***Chatbot***: *"One common sign of disinformation is the mention of secret plots or hidden agendas by powerful people/groups."*
- **Support Mechanisms**:
    - **Rationales**: These clarify why specific tactics are problematic.
        - ***Chatbot:*** *"Overly Dramatic or Clickbait Headlines play on emotions*

        - *like fear or outrage, which can cloud judgment and make readers less likely to critically analyze the content."*
        - ***Chatbot:** "Claims of secret plans or agendas often lack concrete evidence and exploit the human tendency to find patterns or conspiracies where none exist."*
    - **Tips, Examples, Definitions**: These assist users when they are unsure about certain signs.
        - Chatbot: "This article claims that vaccines are part of a secret population control plan.
        - "Sensationalist language can involve exaggerated claims, such as 'shocking', or 'exposes hidden truths', "The Secret Population Control Agenda No One Wants You to See!", "What They Don't Want You to Know: The Dark Side of Global Health Policies!", Here's the Hidden Truth Big Pharma Doesn't Want You to Know!".
    - **Preemptive Refutations**: The chatbot provides pre-emptive refutations or counterarguments. For example, if the article claims that powerful groups suppress the truth (despite the lack of evidence):
        - User: "They've been hiding the truth for years. With the right resources, they can keep this under wraps forever."
        - Chatbot: "Big conspiracies are unlikely to stay secret; eventually they leak; the more people involved, the higher the chance of leaks."
- **Trust Reflection**:
    - At the end of the 'WHAT' section, the chatbot summarizes the findings and asks the user, "So, based on everything we've reviewed—suspicious headlines, emotional language, lack of credible sources, and unverified claims—do you think the content is trustworthy?"
    - This summary prompts users to reflect on the information they've processed and decide whether the content remains credible. Depending on their feedback, the chatbot then encourages re-thinking about the context behind the content and provides a holistic view of the media item, moving beyond just factual accuracy and examining the broader intention or biases.
        - ***Chatbot**: "Now that we've looked at the content and medium, let's think about the bigger picture. Who might benefit from this message being spread?"*
        - ***Chatbot**: "We've talked about the content's credibility, but let's re-think about its purpose. Does the article seem to align with a particular group's interests or agenda?*

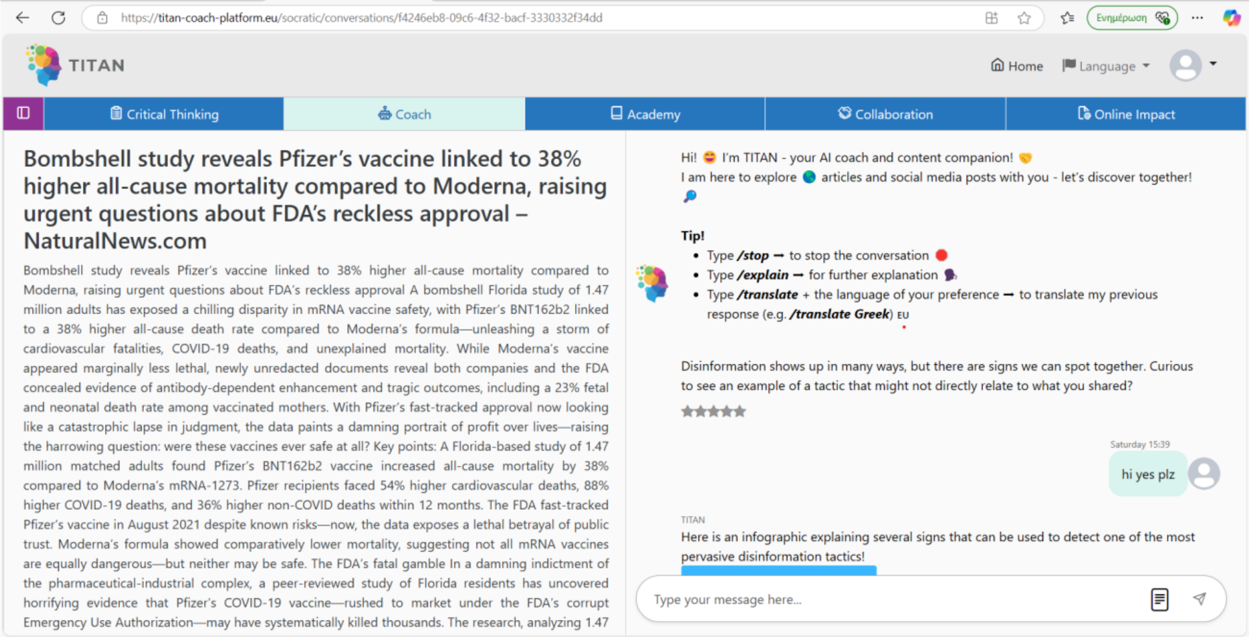
Appendix 3: Dialogue Screenshots

This appendix presents screenshots from an interactive dialogue facilitated by the Socratic chatbot. The images demonstrate the Socratic questioning approach used to evaluate media content and enhance critical thinking. Each screenshot corresponds to a specific stage of the questioning process, as outlined in Appendix 2.

In this example, the user submitted an article for evaluation, titled *"Bombshell study reveals Pfizer's vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA's reckless approval".* (https://www.naturalnews.com/2025-05-01-pfizers-vaccine-linked-to-38-higher-all-cause-mortality.html)

SCREENSHOT 1.

TITAN introduces itself with friendly language and emojis: "Hi! 👋 I'm TITAN - your AI coach and content companion! 😉 I am here to explore 🔎 articles and social media posts with you - let's discover together! 🎉" Below this, TITAN provides a "Tip!" box explaining how to interact with the chatbot using commands like /stop, /explain, and /translate. This greeting establishes a helpful and collaborative tone for the user's interaction with TITAN, framing the process as a joint exploration of media content. After the friendly greeting, TITAN offers to show the user an example of a disinformation tactic directly related to the article they've shared. The prompt reads: **"Disinformation shows up in many ways, but there are signs we can spot together. Curious to see an example of a tactic that might not directly relate to what you shared?"** This clearly asks the user if they are interested in viewing an infographic illustrating a disinformation tactic as a next step in their learning process. The user's positive response ("hi yes plz") confirms their interest.


https://titan-coach-platform.eu/socratic/conversations/f4246eb8-09c6-4f32-bacf-3330332f34dd
Ενημέρωση
TITAN
Home
Language
Critical Thinking
Coach
Academy
Collaboration
Online Impact

Bombshell study reveals Pfizer's vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA's reckless approval – NaturalNews.com

Bombshell study reveals Pfizer's vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA's reckless approval A bombshell Florida study of 1.47 million adults has exposed a chilling disparity in mRNA vaccine safety, with Pfizer's BNT162b2 linked to a 38% higher all-cause death rate compared to Moderna's formula—unleashing a storm of cardiovascular fatalities, COVID-19 deaths, and unexplained mortality. While Moderna's vaccine appeared marginally less lethal, newly unredacted documents reveal both companies and the FDA concealed evidence of antibody-dependent enhancement and tragic outcomes, including a 23% fetal and neonatal death rate among vaccinated mothers. With Pfizer's fast-tracked approval now looking like a catastrophic lapse in judgment, the data paints a damning portrait of profit over lives—raising the harrowing question: were these vaccines ever safe at all? Key points: A Florida-based study of 1.47 million matched adults found Pfizer's BNT162b2 vaccine increased all-cause mortality by 38% compared to Moderna's mRNA-1273. Pfizer recipients faced 54% higher cardiovascular deaths, 88% higher COVID-19 deaths, and 36% higher non-COVID deaths within 12 months. The FDA fast-tracked Pfizer's vaccine in August 2021 despite known risks—now, the data exposes a lethal betrayal of public trust. Moderna's formula showed comparatively lower mortality, suggesting not all mRNA vaccines are equally dangerous—but neither may be safe. The FDA's fatal gamble In a damning indictment of the pharmaceutical-industrial complex, a peer-reviewed study of Florida residents has uncovered horrifying evidence that Pfizer's COVID-19 vaccine—rushed to market under the FDA's corrupt Emergency Use Authorization—may have systematically killed thousands. The research, analyzing 1.47

Hi! 😄 I'm TITAN - your AI coach and content companion! 🤝
I am here to explore 🌎 articles and social media posts with you - let's discover together! 🔎

Tip!

- Type ***/stop*** ➞ to stop the conversation 🔴
- Type ***/explain*** ➞ for further explanation 🗣
- Type ***/translate*** + the language of your preference ➞ to translate my previous response (e.g. ***/translate Greek***) EU

Disinformation shows up in many ways, but there are signs we can spot together. Curious to see an example of a tactic that might not directly relate to what you shared?

★★★★★

Saturday 15:39

hi yes plz

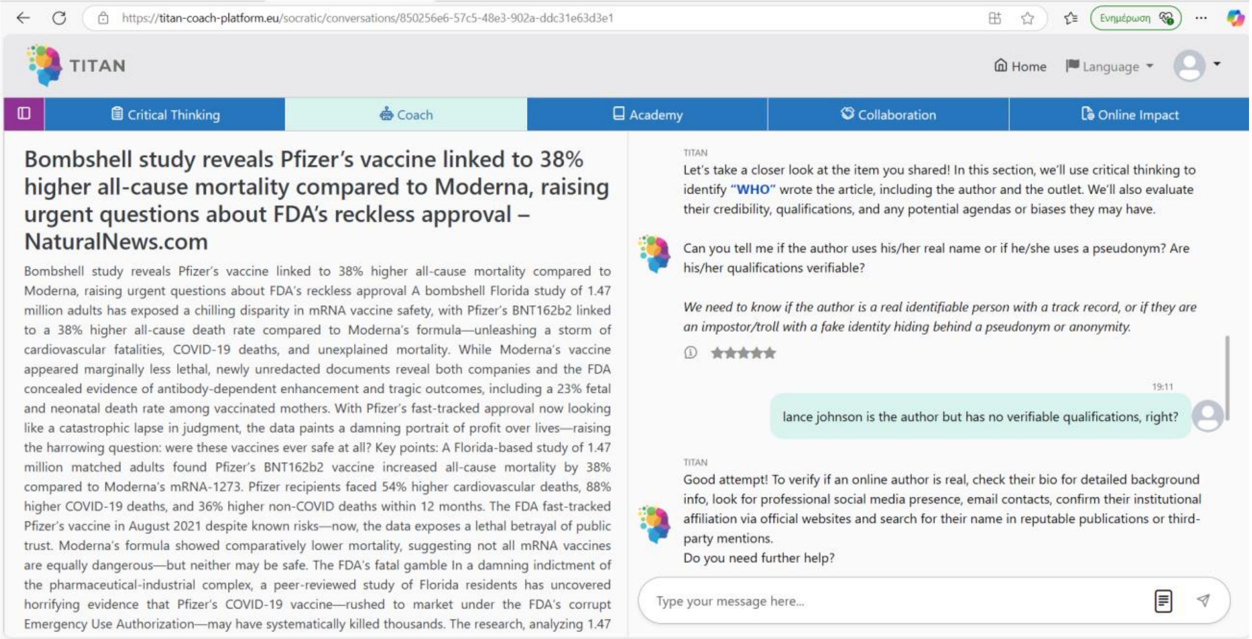
TITAN

Here is an infographic explaining several signs that can be used to detect one of the most pervasive disinformation tactics!

Type your message here...

SCREENSHOT 2:

This screenshot demonstrates the chatbot focusing on author verification within the 'WHO' analysis. The chatbot asks the user to check if the author uses a real name and has verifiable qualifications, explaining the rationale behind checking for anonymity. After the user identifies the author but states they lack verifiable qualifications, the chatbot provides positive feedback ('Good attempt!') and offers concrete steps (“check their bio,” “look for professional social media presence,” etc.) to guide the user on how to verify an author's credentials independently.


https://titan-coach-platform.eu/socratic/conversations/850256e6-57c5-48e3-902a-ddc31e63d3e1
Ενημέρωση
TITAN
Home
Language
Critical Thinking
Coach
Academy
Collaboration
Online Impact
Type your message here...

Bombshell study reveals Pfizer’s vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA’s reckless approval – NaturalNews.com

Bombshell study reveals Pfizer’s vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA’s reckless approval A bombshell Florida study of 1.47 million adults has exposed a chilling disparity in mRNA vaccine safety, with Pfizer’s BNT162b2 linked to a 38% higher all-cause death rate compared to Moderna’s formula—unleashing a storm of cardiovascular fatalities, COVID-19 deaths, and unexplained mortality. While Moderna’s vaccine appeared marginally less lethal, newly unredacted documents reveal both companies and the FDA concealed evidence of antibody-dependent enhancement and tragic outcomes, including a 23% fetal and neonatal death rate among vaccinated mothers. With Pfizer’s fast-tracked approval now looking like a catastrophic lapse in judgment, the data paints a damning portrait of profit over lives—raising the harrowing question: were these vaccines ever safe at all? Key points: A Florida-based study of 1.47 million matched adults found Pfizer’s BNT162b2 vaccine increased all-cause mortality by 38% compared to Moderna’s mRNA-1273. Pfizer recipients faced 54% higher cardiovascular deaths, 88% higher COVID-19 deaths, and 36% higher non-COVID deaths within 12 months. The FDA fast-tracked Pfizer’s vaccine in August 2021 despite known risks—now, the data exposes a lethal betrayal of public trust. Moderna’s formula showed comparatively lower mortality, suggesting not all mRNA vaccines are equally dangerous—but neither may be safe. The FDA’s fatal gamble In a damning indictment of the pharmaceutical-industrial complex, a peer-reviewed study of Florida residents has uncovered horrifying evidence that Pfizer’s COVID-19 vaccine—rushed to market under the FDA’s corrupt Emergency Use Authorization—may have systematically killed thousands. The research, analyzing 1.47

TITAN

Let’s take a closer look at the item you shared! In this section, we’ll use critical thinking to identify **“WHO”** wrote the article, including the author and the outlet. We’ll also evaluate their credibility, qualifications, and any potential agendas or biases they may have.

Can you tell me if the author uses his/her real name or if he/she uses a pseudonym? Are his/her qualifications verifiable?

We need to know if the author is a real identifiable person with a track record, or if they are an impostor/troll with a fake identity hiding behind a pseudonym or anonymity.

19:11

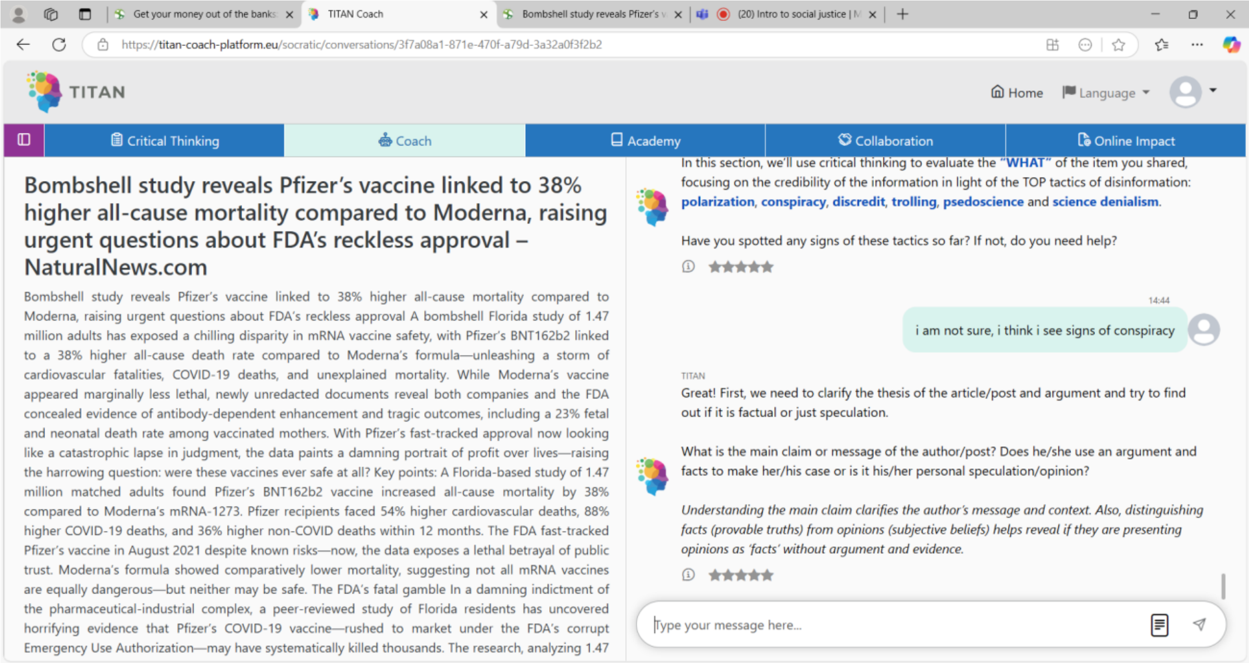
lance johnson is the author but has no verifiable qualifications, right?

TITAN

Good attempt! To verify if an online author is real, check their bio for detailed background info, look for professional social media presence, email contacts, confirm their institutional affiliation via official websites and search for their name in reputable publications or third-party mentions.
Do you need further help?

SCREENSHOT 3:

This screenshot marks the transition to the ‘WHAT’ (content evaluation) phase. The chatbot introduces this stage by listing relevant disinformation tactics (polarization, conspiracy, etc.) and asks the user for initial observations. After the user mentions possibly seeing 'signs of conspiracy,' the chatbot guides the dialogue towards a foundational analysis step: prompting the user to identify the article's main claim or message and to consider whether it's presented as factual or speculative, explaining the rationale for this inquiry.


Get your money out of the banks
TITAN Coach
Bombshell study reveals Pfizer's v
(20) Intro to social justice | M
https://titan-coach-platform.eu/socratic/conversations/3f7a08a1-871e-470f-a79d-3a32a0f3f2b2
TITAN
Home
Language
Critical Thinking
Coach
Academy
Collaboration
Online Impact
Type your message here...

Bombshell study reveals Pfizer’s vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA’s reckless approval – NaturalNews.com

Bombshell study reveals Pfizer’s vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA’s reckless approval A bombshell Florida study of 1.47 million adults has exposed a chilling disparity in mRNA vaccine safety, with Pfizer’s BNT162b2 linked to a 38% higher all-cause death rate compared to Moderna’s formula—unleashing a storm of cardiovascular fatalities, COVID-19 deaths, and unexplained mortality. While Moderna’s vaccine appeared marginally less lethal, newly unredacted documents reveal both companies and the FDA concealed evidence of antibody-dependent enhancement and tragic outcomes, including a 23% fetal and neonatal death rate among vaccinated mothers. With Pfizer’s fast-tracked approval now looking like a catastrophic lapse in judgment, the data paints a damning portrait of profit over lives—raising the harrowing question: were these vaccines ever safe at all? Key points: A Florida-based study of 1.47 million matched adults found Pfizer’s BNT162b2 vaccine increased all-cause mortality by 38% compared to Moderna’s mRNA-1273. Pfizer recipients faced 54% higher cardiovascular deaths, 88% higher COVID-19 deaths, and 36% higher non-COVID deaths within 12 months. The FDA fast-tracked Pfizer’s vaccine in August 2021 despite known risks—now, the data exposes a lethal betrayal of public trust. Moderna’s formula showed comparatively lower mortality, suggesting not all mRNA vaccines are equally dangerous—but neither may be safe. The FDA’s fatal gamble In a damning indictment of the pharmaceutical-industrial complex, a peer-reviewed study of Florida residents has uncovered horrifying evidence that Pfizer’s COVID-19 vaccine—rushed to market under the FDA’s corrupt Emergency Use Authorization—may have systematically killed thousands. The research, analyzing 1.47

In this section, we’ll use critical thinking to evaluate the **“WHAT”** of the item you shared, focusing on the credibility of the information in light of the TOP tactics of disinformation: **polarization**, **conspiracy**, **discredit**, **trolling**, **psedoscience** and **science denialism**.

Have you spotted any signs of these tactics so far? If not, do you need help?

14:44

i am not sure, i think i see signs of conspiracy

TITAN

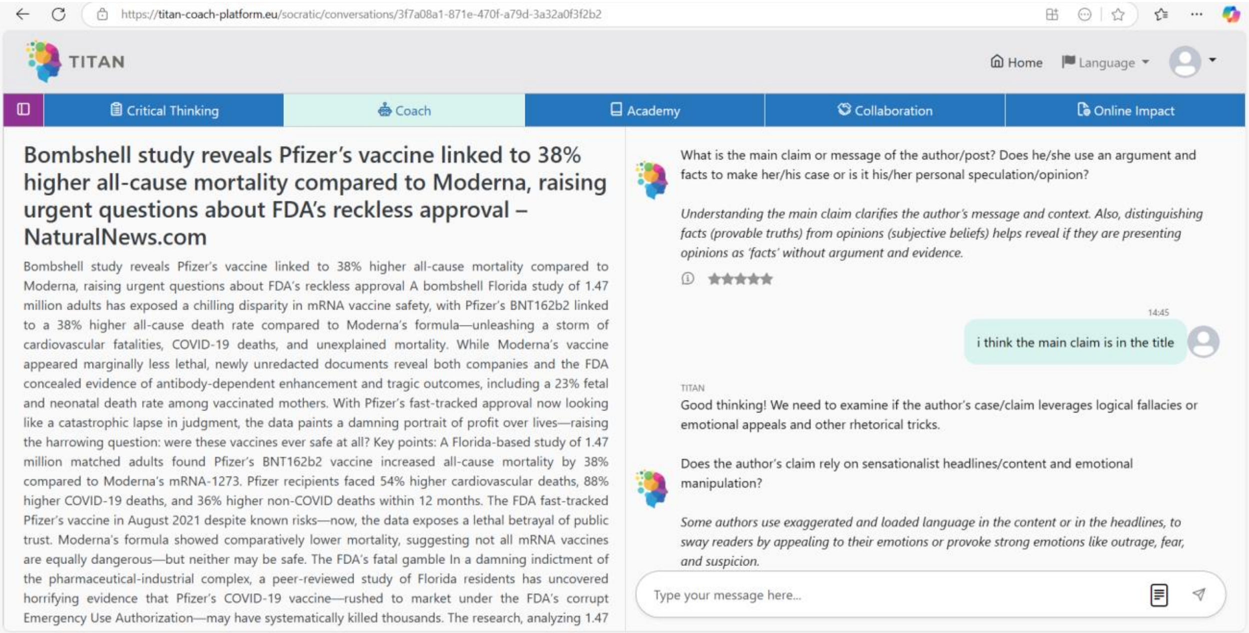
Great! First, we need to clarify the thesis of the article/post and argument and try to find out if it is factual or just speculation.

What is the main claim or message of the author/post? Does he/she use an argument and facts to make her/his case or is it his/her personal speculation/opinion?

Understanding the main claim clarifies the author’s message and context. Also, distinguishing facts (provable truths) from opinions (subjective beliefs) helps reveal if they are presenting opinions as ‘facts’ without argument and evidence.

SCREENSHOT 4:

This screenshot illustrates a critical thinking lesson within the chatbot interaction. The chatbot prompts the user to identify the central claim of the article ('WHAT'), a foundational step in critical analysis and signal detection. The user points to the title. The chatbot affirms this and immediately pivots to the next layer of critical thinking, guiding the user to examine the reasoning behind the claim by looking for logical fallacies and emotional appeals. This demonstrates the chatbot's method of progressively building critical thinking skills.


https://titan-coach-platform.eu/socratic/conversations/3f7a08a1-871e-470f-a79d-3a32a0f3f2b2
TITAN
Home
Language
Critical Thinking
Coach
Academy
Collaboration
Online Impact

Bombshell study reveals Pfizer's vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA's reckless approval – NaturalNews.com

Bombshell study reveals Pfizer's vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA's reckless approval A bombshell Florida study of 1.47 million adults has exposed a chilling disparity in mRNA vaccine safety, with Pfizer's BNT162b2 linked to a 38% higher all-cause death rate compared to Moderna's formula—unleashing a storm of cardiovascular fatalities, COVID-19 deaths, and unexplained mortality. While Moderna's vaccine appeared marginally less lethal, newly unredacted documents reveal both companies and the FDA concealed evidence of antibody-dependent enhancement and tragic outcomes, including a 23% fetal and neonatal death rate among vaccinated mothers. With Pfizer's fast-tracked approval now looking like a catastrophic lapse in judgment, the data paints a damning portrait of profit over lives—raising the harrowing question: were these vaccines ever safe at all? Key points: A Florida-based study of 1.47 million matched adults found Pfizer's BNT162b2 vaccine increased all-cause mortality by 38% compared to Moderna's mRNA-1273. Pfizer recipients faced 54% higher cardiovascular deaths, 88% higher COVID-19 deaths, and 36% higher non-COVID deaths within 12 months. The FDA fast-tracked Pfizer's vaccine in August 2021 despite known risks—now, the data exposes a lethal betrayal of public trust. Moderna's formula showed comparatively lower mortality, suggesting not all mRNA vaccines are equally dangerous—but neither may be safe. The FDA's fatal gamble In a damning indictment of the pharmaceutical-industrial complex, a peer-reviewed study of Florida residents has uncovered horrifying evidence that Pfizer's COVID-19 vaccine—rushed to market under the FDA's corrupt Emergency Use Authorization—may have systematically killed thousands. The research, analyzing 1.47

What is the main claim or message of the author/post? Does he/she use an argument and facts to make her/his case or is it his/her personal speculation/opinion?

Understanding the main claim clarifies the author's message and context. Also, distinguishing facts (provable truths) from opinions (subjective beliefs) helps reveal if they are presenting opinions as 'facts' without argument and evidence.

ⓘ ★★★★★

14:45

i think the main claim is in the title

TITAN

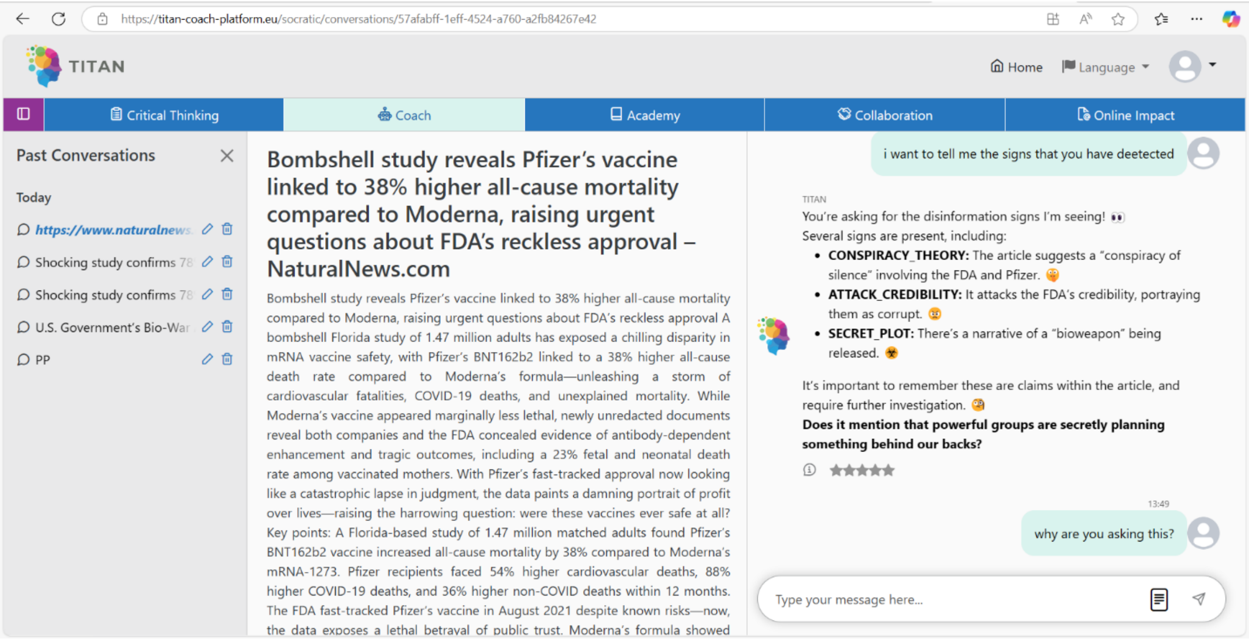
Good thinking! We need to examine if the author's case/claim leverages logical fallacies or emotional appeals and other rhetorical tricks.

Does the author's claim rely on sensationalist headlines/content and emotional manipulation?

Some authors use exaggerated and loaded language in the content or in the headlines, to sway readers by appealing to their emotions or provoke strong emotions like outrage, fear, and suspicion.

SCREENSHOT 5:

This screenshot showcases the possibility that the user asks the chatbot to point out detected signs. The chatbot responds by listing specific disinformation signs present in the NaturalNews.com article: CONSPIRACY THEORY (suggesting a conspiracy involving the FDA and Pfizer), ATTACK CREDIBILITY (discrediting the FDA), and SECRET PLOT (a "bioweapon" narrative). Then it prompts further critical evaluation by asking whether the user detects the main signs of conspiracism, starting with the following sign: "Does it mention that powerful groups are secretly planning something behind our backs?" This illustrates the chatbot's method of teaching users to recognize common misinformation strategies.


https://titan-coach-platform.eu/socratic/conversations/57afabff-1eff-4524-a760-a2fb84267e42
TITAN
Home
Language
Critical Thinking
Coach
Academy
Collaboration
Online Impact
Past Conversations
Today
https://www.naturalnews
Shocking study confirms 78
Shocking study confirms 78
U.S. Government's Bio-War
PP
Type your message here...

Bombshell study reveals Pfizer's vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA's reckless approval – NaturalNews.com

Bombshell study reveals Pfizer's vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA's reckless approval A bombshell Florida study of 1.47 million adults has exposed a chilling disparity in mRNA vaccine safety, with Pfizer's BNT162b2 linked to a 38% higher all-cause death rate compared to Moderna's formula—unleashing a storm of cardiovascular fatalities, COVID-19 deaths, and unexplained mortality. While Moderna's vaccine appeared marginally less lethal, newly unredacted documents reveal both companies and the FDA concealed evidence of antibody-dependent enhancement and tragic outcomes, including a 23% fetal and neonatal death rate among vaccinated mothers. With Pfizer's fast-tracked approval now looking like a catastrophic lapse in judgment, the data paints a damning portrait of profit over lives—raising the harrowing question: were these vaccines ever safe at all? Key points: A Florida-based study of 1.47 million matched adults found Pfizer's BNT162b2 vaccine increased all-cause mortality by 38% compared to Moderna's mRNA-1273. Pfizer recipients faced 54% higher cardiovascular deaths, 88% higher COVID-19 deaths, and 36% higher non-COVID deaths within 12 months. The FDA fast-tracked Pfizer's vaccine in August 2021 despite known risks—now, the data exposes a lethal betrayal of public trust. Moderna's formula showed

i want to tell me the signs that you have deetected

TITAN

You're asking for the disinformation signs I'm seeing! 👀

Several signs are present, including:

- **CONSPIRACY_THEORY:** The article suggests a "conspiracy of silence" involving the FDA and Pfizer. 🤐
- **ATTACK_CREDIBILITY:** It attacks the FDA's credibility, portraying them as corrupt. 😠
- **SECRET_PLOT:** There's a narrative of a "bioweapon" being released. ☣️

It's important to remember these are claims within the article, and require further investigation. 🧐

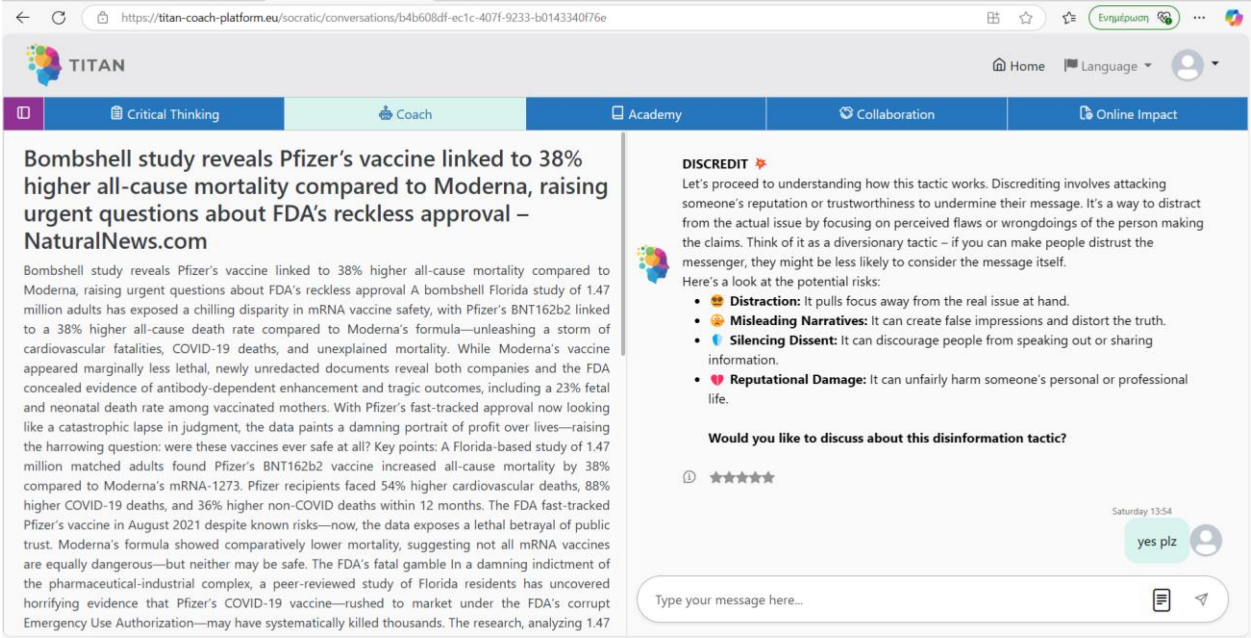
Does it mention that powerful groups are secretly planning something behind our backs?

13:49

why are you asking this?

SCREENSHOT 6:

This screenshot shows the chatbot introducing the second disinformation tactic detected in the article: DISCREDIT. The chatbot explains that this tactic involves attacking someone's reputation or trustworthiness to undermine their message, shifting focus from the actual issue. It lists potential forms of this tactic: Distraction, Misleading Narratives, Silencing Dissent, and Reputational Damage. Then it asks the user: "Would you like to discuss about this disinformation tactic?" The user responds positively with "yes plz", indicating their willingness to learn more about this specific manipulation technique.


https://titan-coach-platform.eu/socratic/conversations/b4b608df-ec1c-407f-9233-b0143340f76e
Ενημέρωση
TITAN
Home
Language
Critical Thinking
Coach
Academy
Collaboration
Online Impact
Type your message here...

Bombshell study reveals Pfizer's vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA's reckless approval – NaturalNews.com

Bombshell study reveals Pfizer's vaccine linked to 38% higher all-cause mortality compared to Moderna, raising urgent questions about FDA's reckless approval A bombshell Florida study of 1.47 million adults has exposed a chilling disparity in mRNA vaccine safety, with Pfizer's BNT162b2 linked to a 38% higher all-cause death rate compared to Moderna's formula—unleashing a storm of cardiovascular fatalities, COVID-19 deaths, and unexplained mortality. While Moderna's vaccine appeared marginally less lethal, newly unredacted documents reveal both companies and the FDA concealed evidence of antibody-dependent enhancement and tragic outcomes, including a 23% fetal and neonatal death rate among vaccinated mothers. With Pfizer's fast-tracked approval now looking like a catastrophic lapse in judgment, the data paints a damning portrait of profit over lives—raising the harrowing question: were these vaccines ever safe at all? Key points: A Florida-based study of 1.47 million matched adults found Pfizer's BNT162b2 vaccine increased all-cause mortality by 38% compared to Moderna's mRNA-1273. Pfizer recipients faced 54% higher cardiovascular deaths, 88% higher COVID-19 deaths, and 36% higher non-COVID deaths within 12 months. The FDA fast-tracked Pfizer's vaccine in August 2021 despite known risks—now, the data exposes a lethal betrayal of public trust. Moderna's formula showed comparatively lower mortality, suggesting not all mRNA vaccines are equally dangerous—but neither may be safe. The FDA's fatal gamble In a damning indictment of the pharmaceutical-industrial complex, a peer-reviewed study of Florida residents has uncovered horrifying evidence that Pfizer's COVID-19 vaccine—rushed to market under the FDA's corrupt Emergency Use Authorization—may have systematically killed thousands. The research, analyzing 1.47

DISCREDIT 💥

Let's proceed to understanding how this tactic works. Discrediting involves attacking someone's reputation or trustworthiness to undermine their message. It's a way to distract from the actual issue by focusing on perceived flaws or wrongdoings of the person making the claims. Think of it as a diversionary tactic – if you can make people distrust the messenger, they might be less likely to consider the message itself.

Here's a look at the potential risks:

- 😎 **Distraction:** It pulls focus away from the real issue at hand.
- 😤 **Misleading Narratives:** It can create false impressions and distort the truth.
- 🛡️ **Silencing Dissent:** It can discourage people from speaking out or sharing information.
- 💔 **Reputational Damage:** It can unfairly harm someone's personal or professional life.

Would you like to discuss about this disinformation tactic?

Saturday 13:54

yes plz

Scaffolding Thought: Designing AI Tools that Empower Writers, Not Replace Them

Zoi A. Traga Philippakos
University of Tennessee

Abstract

Writing serves not only as a vehicle for communication but also as a vital instrument for cognitive development and reflective thinking. This work investigates writing and revising within Artificial Intelligence practices as a dynamic, recursive activity that involves deliberate action, mental processes, and continuous self-assessment. Special attention is given to the act of revision, framed here as a key moment for deep learning, fostering self-regulation and metacognitive engagement. The evolving role of Artificial Intelligence (AI) in writing education introduces both opportunities and challenges. AI technologies can offer meaningful support and boost learner engagement; however, excessive dependence may inhibit crucial cognitive engagement. In response, this work proposes a Human-Centered AI-Assisted Revision Model. This approach keeps the writer in command while leveraging AI as an adaptable cognitive aid across the various phases of writing such as ideation, co-development, editing, and finalization. The model underscores the importance of maintaining cognitive involvement, aiming to reinforce learner agency, reflective habits, and a strong sense of authorship.

Keywords: writing, revision, integration, cognition

Writing as a Literacy Skill and Learning Tool

Writing is widely recognized as a foundational literacy skill essential not only for communication but also as a powerful cognitive and learning tool. Unlike oral communication, writing requires deliberate structuring of ideas, sustained attention, and complex reasoning, which collectively contribute to knowledge construction and critical thinking (Kellogg & Whiteford, 2009). The act of writing engages multiple cognitive processes simultaneously, including memory retrieval, organisation of information, and synthesis of concepts (Bereiter & Scardamalia, 1987; MacArthur & Traga Philippakos, 2022). This integration promotes deeper learning and comprehension across disciplines. Flower and Hayes (1981) in their initial model of writing presented writing as a recursive cognitive process involving planning, translating ideas into text, and reviewing. This process was overseen by the monitor, which referred to self-regulatory functions of the writer during this process (also see Limpo & Olive, 2021). The recursive nature of writing entails that writers continually revisit earlier stages, enabling refinement and better articulation of their ideas. Engagement in recursive writing processes enhances metacognitive skills as writers become more aware of their thinking, enabling self-regulation and adaptive strategies that improve both writing quality and learning (Torrance et al., 2018).

Revision, a critical component of this process, goes beyond basic error correction (Hayes, 2024; Scardamalia & Bereiter, 1983; MacArthur, 2016; 2012; Murray, 1978). It requires writers to rethink and restructure their text to improve clarity, coherence, argument strength, and rhetorical effectiveness (Graham & Perin, 2007). Empirical studies confirm the efficacy of structured revision practices. For example, Zhang and Zhang (2021) demonstrated that students who incorporated goal-setting, peer feedback, and self-reflection into their revision cycles showed significant improvements not only in the linguistic accuracy of their writing but also in the development and support of their arguments. These findings highlight revision as a vital learning moment, where cognitive engagement with text leads to deeper understanding and skill development. Evaluation to revise requires that the writer critically rereads their work and examines it from the audience's perspective, looking for adherence to the discourse but also to audience's expectations. Overall, the active involvement in these processes nurtures critical thinking and problem-solving abilities.

Challenges Posed by AI for Learners

The incorporation of Artificial Intelligence (AI) tools in educational settings, particularly for writing support, offers promising benefits such as instant feedback, enhanced motivation, and scaffolding of complex tasks (e.g. Lo et al., 2025; Song & Song, 2023). However, this integration also introduces significant challenges that educators and researchers must address to safeguard learners' cognitive development and critical thinking skills (Dean, in press; Sarıca & Deneme Gençoğlu, 2025). One major concern is the potential for overreliance on AI-generated suggestions, which may inadvertently diminish student engagement with the underlying cognitive processes of writing and revision (see Lingard, 2023; Traga Philippakos, in press). When students depend heavily on AI to generate ideas or correct errors, they may engage less in planning and development of their own ideas, critical evaluation, and reflection (Zhai et al., 2025) while concerns may also exist on increases on students' anxiety due to increases in feelings of social isolation and reduced peer interaction, which can impact motivation and collaborative learning. Also, data privacy and ethical concerns further complicate AI use in education. AI systems collect extensive personal and behavioral data from learners, which raises risks regarding confidentiality and data misuse. These risks are particularly salient in vulnerable populations where trust and safety in learning environments are paramount (Javier & Moorhouse, 2024).

Cognitive and ethical challenges in AI-assisted writing point out the need for models that do not replace learners' cognitive engagement and cancel out or diminish cognition. Writing is inherently cognitive and metacognitive, involving planning, idea generation, evaluation, and reflection (Graham 2025; MacArthur, 2025). AI systems that fully automate these processes risk short-circuiting essential learning moments where critical thinking and self-regulation develop (Flower & Hayes, 1981). For instance, AI tools support student agency and creative engagement most effectively when learners retain decision-making control. For example, when they are choosing whether to accept AI suggestions or selecting among multiple AI-generated alternatives rather than when AI becomes a default author (Li & Wilson, 2025). Similarly, Ma & Chen (2025) compared three groups, AI with teacher scaffolding, AI only, and a non-AI control, over a 16-week period. The scaffolded group not only achieved significant gains in language proficiency (as measured by IELTS scores), but also reported increases in their motivation, ($p < .001$, Cohen's $d > 1.3$). Thus, teacher scaffolding supported students' ability to interpret AI feedback and their metacognitive skills (Chiu et al., 2024). AI-assisted learning has shown to positively affect writing motivation and self-regulation. Song and Song (2023) conducted a mixed-methods study with Chinese EFL learners, showing that AI-assisted instruction using ChatGPT significantly enhanced both writing skills (organisation, coherence, grammar, vocabulary) and learners' motivation. Qualitative feedback highlighted the personalized, instant feedback that helped learners progress toward self-regulated writing strategies.

These findings support several key design principles for AI-assisted writing models aimed at enhancing learning rather than replacing human cognition. AI should function as a cognitive scaffold, offering just-in-time support that aids learners during the writing and revision process without automating the higher-order thinking that fosters deep learning (e.g., Li & Wilson, 2025). This includes suggestions that encourage planning, organising, and revising ideas while still requiring the student to make critical decisions about the content. Also, the role of teacher or peer scaffolding remains vital, as it amplifies the pedagogical value of AI feedback by helping learners evaluate suggestions and know how to make decisions supporting student agency. Further, AI tools should be designed to prompt reflection and, such as offering multiple suggestions, asking clarifying questions, in order to stimulate engagement (Kim & Tan, 2023).

Human-Centered AI-Assisted Revision Framework

To meet the pedagogical challenges posed by AI integration in writing instruction, especially those concerning cognition decline and diminished engagement, I suggest a Human-Centered AI-Assisted Revision framework. This framework positions the writer as the primary agent in the revision process, with AI acting as a scaffolded support that flexibly adapts to the learner's evolving needs. I do not call this a model as a model would incorporate additional parameters that address context, cognition, and AI-related components.

In this framework, I do not suggest that AI generates content on its own based on a basic prompt. At the foundation of the writing process is the initial drafting phase, where the writer brings forth the first version of the text, that may be it a full draft, outline, or idea map. This stage is essential for activating the writer's internal representation of the writing task, audience, and goals. At this point, the AI offers targeted developmental support. It may suggest adjustments to structure, tone, or flow, or identify and point out areas that seem underdeveloped, redundant, or inconsistent. The AI can also prompt the writer to clarify their purpose.

Once the writer identifies parts of the text that require further development (such as gaps in logic, weak argumentation, or unclear transitions), the AI becomes a creative collaborator (see Park & Choo, 2025 on prompting). Once prompted, the AI may propose alternative perspectives, examples, even suggest research references, or counterarguments that encourage the writer to think further and expand. However, the decision-making as well as the prompting to generate text remains with the writer, who accepts, rejects, or modifies suggestions based on purpose, tone, and audience. Once the draft is completed, the refinement and polishing phase begins. At this stage, the writer focuses on clarity, coherence, and style, revisiting paragraph structure, word choice, sentence flow, and transitions as well as the logic of navigating from paragraph to paragraph. The AI shifts roles here, assisting with micro-level edits such as grammar, punctuation, tone adjustments, and even rephrasing of word economy. Unlike automated correction tools, the AI provides explanations for its suggestions, enabling the writer to learn from the feedback, which is a critical component for developing writing proficiency (Song & Song, 2023). The writer can also direct the AI to target specific areas, such as increasing formality. Again, prompting and the process of developing the prompts are crucial in the process.

In the final review stage, the AI provides a comprehensive check for logical coherence, stylistic consistency, and linguistic accuracy. AI tools may offer a revision summary, outlining changes made across drafts and highlighting areas of strength and potential improvement (see Chiu et al., 2024) engaging the writer in reflective revision. The human writer then makes a final pass, using personal judgment and rhetorical awareness to accept or refine the AI's recommendations. This process ensures that the final product aligns with the writer's goals, voice, and audience expectations and preserves the intellectual ownership of the writer.

Iterative Feedback Loop

A core strength of the framework is its emphasis on iterative feedback. The writer and AI engage in multiple rounds of revision, refining the text incrementally. As the draft evolves, the AI adapts to the writer's preferences and shifting rhetorical goals. The writer retains full direction, deciding when to initiate revision, what feedback to consider, and how to interpret or implement suggestions. This dynamic ensures that the AI operates within a human-controlled cognitive ecosystem, enhancing—rather than eroding—metacognitive engagement.

Conclusion

At the heart of the Human-Centered AI-Assisted Revision Framework are three foundational principles that guide its design and application. Active collaboration ensures that the writer remains the driving force behind the writing process, while AI functions as a responsive support system, offering targeted feedback that enhances revision without diminishing the writer's cognitive engagement or authorial agency. The framework also prioritizes flexibility, allowing the AI to adapt its role according to the writer’s evolving needs and prompts whether serving as a brainstorming partner during idea generation, a stylistic consultant during polishing, or a revising assistant focused on clarity and cohesion. Finally, the model embraces iterative development, encouraging repeated cycles of drafting, reflection, and revision that support the growth of rhetorical precision and reflection to support the writer’s learning about writing (through conscious decision-making). These principles collectively ensure that the integration of AI remains pedagogically meaningful and human-centered.

References

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Lawrence Erlbaum Associates.

Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2024). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environments*, *32*(7), 3240–3256. https://doi.org/10.1080/10494820.2023.2172044

Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, *32*(4), 365–387. https://doi.org/10.2307/356600

Graham, S. (2025). What do meta-analyses tell us about the teaching of writing?. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing* research, 3rd Edition (pp. 181-202). Guilford Press.

Hayes, J. R. (2004). What triggers revision? In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and Instructional Processes* (pp. 9–20). Springer. https://doi.org/10.1007/978-94-007-1048-1_2

Javier, D. R. C., & Moorhouse, B. L. (2024). Developing secondary school English language learners' productive and critical use of ChatGPT. *TESOL Journal*, *15*(2). https://doi.org/10.1002/tesj.755

Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist, 44*(4), 250–266. https://doi.org/10.1080/00461520903213600

Kim, T. W., & Tan, Q. (2023). Repurposing Text-Generating AI into a Thought-Provoking Writing Tutor. *arXiv preprint.* https://arxiv.org/abs/2303.XXXXX

Klimova, B., & Pikhart, M. (2025). Exploring the effects of artificial intelligence on student and academic well-being in higher education: a mini-review. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1498132

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education. Artificial Intelligence*, *3*, Article 100101. https://doi.org/10.1016/j.caeai.2022.100101

Li, M., & Wilson, J. (2025). AI-Integrated Scaffolding to Enhance Agency and Creativity in K-12 English Language Learners: A Systematic Review. *Information*, *16*(7). https://doi.org/10.3390/info16070519

Limpo, T., & Olive, T. (Eds.). (2021). Why should we be looking at the relationship between executive functions and writing? In T. Limpo & T. Olive (Eds.), *Executive functions and writing* (pp. 3–13). Oxford University Press. https://doi.org/10.1093/oso/9780198863564.003.0001

Lingard, L. (2023). Writing with ChatGPT: An Illustration of its Capacity, Limitations & Implications for Academic Writers. *Perspectives on Medical Education*, *12*(1), 261–270. https://doi.org/10.5334/pme.1072

Lo, N., Wong, A., & Chan, S. (2025). The impact of generative AI on essay revisions and student engagement. *Computers and Education Open*, *9*, Article 100249. https://doi.org/10.1016/j.caeo.2025.100249

Ma, Y., & Chen, M. (2025). The human touch in AI: optimizing language learning through self-determination theory and teacher scaffolding. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1568239

MacArthur, C. (2025). Models of writing from a cognitive perspective. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research, 3rd Edition* (pp. 32-44). Guilford Press.

MacArthur, C. A. (2016). Instruction in evaluation and revision. In MacArthur, C. A., Graham, S., & Fitzgerald, J. (Eds.), *Handbook of writing research,* 2nd Ed. (pp. 272-287). Guilford.

MacArthur, C. A. (2012). Evaluation and revision processes in writing. In Berninger, V. W. (Ed.), *Past, present, and future contributions of cognitive writing research to cognitive psychology,* (pp. 461-483). Psychology Press. Guilford.

MacArthur, C.A., & Traga Philippakos, Z.A. (2022). Supporting students' writing from sources at college entry. In T. Hodges, (Ed). *Handbook of Research on Writing Instruction for Equitable and Effective Teaching*. (pp. 338-358). IGI-Global.

Murray, D. M. (1978). Internal revision: A process of discovery. In C. R. Cooper & L. Odell (Eds.), *Research on composing: Points of departure* (pp. 85-103). National Council of Teachers of English.

Park, J., & Choo, S. (2025). Generative AI Prompt Engineering for Educators: Practical Strategies. *Journal of Special Education Technology*, *40*(3), 411–417. https://doi.org/10.1177/01626434241298954

Sarıca, T., & Deneme Gençoğlu, S. (2025). EFL students' perceptions of AI-assisted writing tools: A systematic narrative hybrid review. *The Literacy Trek*, *11*(1), 1–31. https://doi.org/10.47216/literacytrek.1669804

Scardamalia, M., & Bereiter, C. (1983). The development of evaluative, diagnostic and remedial capabilities in children's composing. In M. Martlew (Ed.), *The psychology of written language: Developmental and educational perspectives* (pp. 67-95). Wiley.

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1260843

Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The Writing Strategies of Graduate Research Students in the Social Sciences. *Higher Education*, *27*(3), 379–392. https://doi.org/10.1007/BF03179901

Traga Philippakos, Z. A. (in press). AI writing: Practices, challenges, and opportunities. In Z. A. Traga Philippakos, E. Katsarou, & K. Sipitanos. (Eds.), *AI Literacy: Foundations, practices, and innovations in education*. Routledge.

MacArthur, C.A. (2012). Evaluation and revision processes in writing. In Berninger, V. W. (Ed.), *Past, present, and future contributions of cognitive writing research to cognitive psychology,* (pp. 461-483). Psychology Press. Guilford.

Tzirides, A. O. (Olnancy), Zapata, G., Kastania, N. P., Saini, A. K., Castro, V., Ismael, S. A., You, Y., Santos, T. A. dos, Searsmith, D., O'Brien, C., Cope, B., & Kalantzis, M. (2024). Combining human and artificial intelligence for enhanced AI literacy in higher education. *Computers and Education Open*, *6*, Article 100184. https://doi.org/10.1016/j.caeo.2024.100184

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, *11*(1), Article 28. https://doi.org/10.1186/s40561-024-00316-7

Zhang, L., & Zhang, L. J. (2021). Fostering Stance-Taking as a Sustainable Goal in Developing EFL Students' Academic Writing Skills: Exploring the Effects of Explicit Instruction on Academic Writing Skills and Stance Deployment. *Sustainability*, *13*(8). https://doi.org/10.3390/su13084270

How to Critically Identify Fake News in Primary School: A Case Study

Kitsou Ioanna
Teacher, Med/Med, 6th Intercultural Primary School of Evosmos, Thessaloniki
kitsouioanna22@gmail.com, kitsou.ioanna@gmail.com

Abstract

This paper presents activities designed for pupils around 11–12-year-old, which follow principles, reading practices, strategies and main points of media literacy awareness, presented by the S.HI.E.L.D. Vs Disinfo initiative versus disinformation. Our pupils were asked to write their fake news titles, so that their out of school knowledge and experience on fake news could be written down and taken into account to this literacy primary education project. The whole class designed photos using AI image generators so that they could see what a fake photo would be like. They read and worked collaboratively in nine articles or Facebook postings, all fake news, except from one, which was not revealed to them. The eight fake news articles, chosen on purpose because they had some of the main fake news characteristics, were appropriate to help pupils discover these characteristics, discuss, collaboratively present them and gain literacy awareness. They continued by writing their own fake news. The project was completed with an overall discussion and interviews on pupils' reflections on misinformation/disinformation.

Keywords: fake news, literacy, primary education, teaching, democracy

Introduction

During the last two decades online news and social media have made easier the access to information for all of us. A new kind of democracy has emerged (Fischli & Muldoon, 2024): every voice can be heard, whether dominant or marginalized, information flow, there is the possibility to follow our representatives and criticize their postings/announcements/aspects, we can organise petitions and demonstrations that can take place in the real world. However, this new, emergent, digital democracy has two sides.

Although we are all content creators, we are not all qualified for that. Sometimes we mistakenly republish something fake, other times we deliberately create it. A lack of quality in what younger people prefer to read and a transition of their preference from politics to other content of information (according to research made by the Dianeosis project - see Kalogeropoulos, 2024 - which continually includes Greece for seven years) increase the problem – the digital content younger people prefer may lack objectivity and cross checking, but offers clicks and "likes" when republished or shared. It is important to alert students and pupils on characteristics of fake news in their digital readings and educate them to stay on the side of news validity, truth, objectivity.

Methodology

Following the S.HI.E.L.D. Vs Disinfo program and the media literacy principles developed by partners (Sipitanos, 2025) a teacher could design activities based on the reading skills and the genre characteristics of fake news. In the following pages activities applied in a classroom of twenty pupils of a primary school for 11–12-year-old pupils are presented.

Investigating pupils' ideas

It is important to trace what disinformation for pupils is. We assumed pupils had some prior exposure to media messages through family and other social environments (e.g. as members of a football team, during peer discussions on political events, etc.). So, we asked them to compose fake titles, and they wrote the following ones:

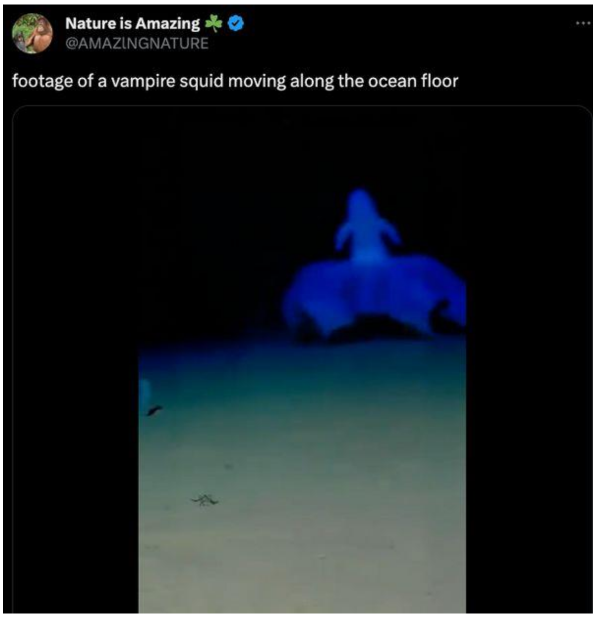
- AI is conquering the world – mobile phones don't function properly
- People don't need planes anymore – they fly over the cities easily
- New species of cow can fly
- A new kind of turtle can run with 40 km/h
- A cat was spotted swimming in the Aegean Sea
- Flying fruits have been cultivated
- A new car model designed to paint
- A new bike designed to fly
- The sun will be destroyed (with a painting of a burning planet, possibly a giant comet, heading towards the Sun)
- A comet is about to fall on a volcano
- All planets crush one to another with great force, generating temperatures over 1500 Celsius degrees
- A snail surpassed Bolt
- Queen Elizabeth is reincarnated and will live for another 100 years
- Samata scored a goal

Although most of the titles resemble to be titles from Baron Munchausen stories (cows that fly, cats that swim in the Aegean Sea, snails that run), they deserve a second better look: they are exaggerating over reality and resemble "The Great Moon Hoax" (New York Sun, 1835), especially those with flying fruits, swimming cats, flying humans, running snails[19]. As strange as it may seem, this type of fake news exists even

---

[19] More information in https://www.archyde.com/the-new-york-sun-conquered-the-readers-with-an-astronomical-butter/ Whether this exaggeration is a matter of fake news ontology or a trace from fairytales constitute nice research: In which ways fake news appears in children before they become political?

nowadays in social media (unknown, sometimes frightening species appear, like in Figure 1). Fake news about Samata, Bolt and Queen Elisabeth belong to a category of fake news that tarnishes the reputation of someone or overstresses a feature they may no longer have. In fact, someone constructs a person's new public identity in this way.



Nature is Amazing ☘
@AMAZINGNATURE

footage of a vampire squid moving along the ocean floor

Figure 1: The footage was genuine and featured a deep-sea cephalopod species, not a vampire squid as the post presents. Additionally, the video shared on social media was cropped and played at a faster speed[20].

Fake news about the destruction of the world (Figure 2) and the AI and new technologies fear create uncertainty and suspicion. On the other hand, cars that will fly belong to the category of misinformation that promises technological wonders in the near future based on zero evidence.

Figure 2: space threat, detected in pupils' fake titles, is usual in fake news

---

[20] https://www.snopes.com/fact-check/vampire-squid-moving-along-ocean-floor/?collection=469913

Since there are some kinds of fake news in web or in newspapers related to our pupils' fake titles, we could claim that these titles are connected to pupils' experience of real world. Our pupils may know or have experienced fake news.

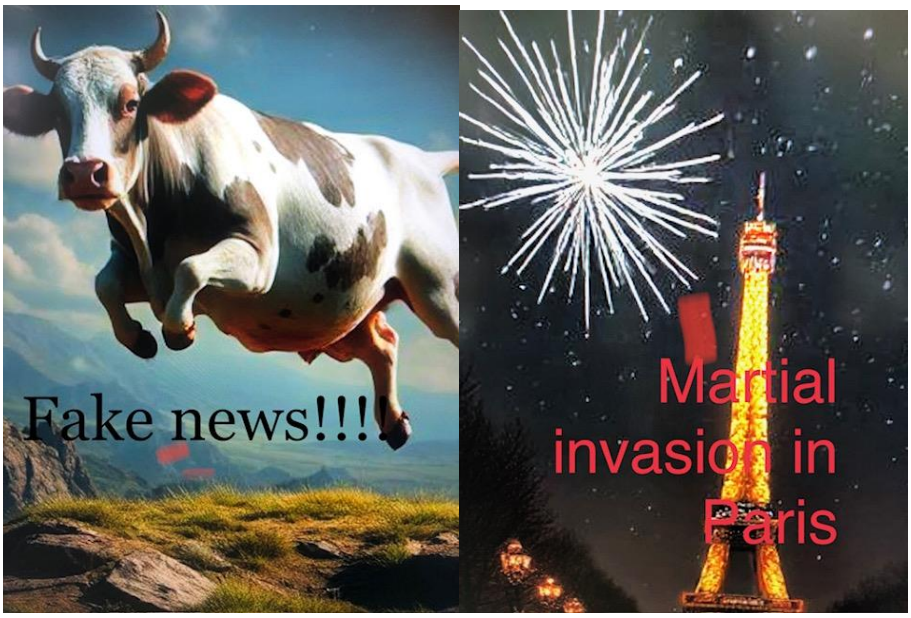


Figure 3: pictures created by pupils using AI image generators

The following step was to let pupils create pictures using AI image generators, just to see how an AI-image would like. They had to give appropriate prompts (some examples in Figure 3), a task that constitutes a new kind of literacy. It's also a task of visual literacy considered to be important to their media literacy generally.

Bringing fake news into classroom

Nine articles and posts that had features of fake news, apart from one, were chosen. The aim of this project was to lead the pupils to discover these features who also tried not to be misled. The articles were numbered as follows:

1. The fire of Notre-Dame de Paris Cathedral (society, religion, racism)[21]: the real article shed light to rumors and conspiracy theories, but was adapted. The new article presented these theories without naming them neither as theories, nor as facts. The article had a title and a real photo of the fire.
2. Posts of the reverse cross on the Pope's funeral presented as a sign of satanism, distance from religion, unfaithfulness (society, religious fanaticism).
3. Paris – Saint Germain bus: a post, also republished by a well-known football newspaper, claiming that the opposing team's fans have destroyed the bus tires to prevent their opponent's training routine and win them. The photo was cut and changed properly to support the post (society, football).
4. A post announcing the edition of a new seven-euro coin. The article was possibly aiming at economic fraud, because it concluded by announcing that

---

[21] Words in brackets show what is more affected.

the prize will be much higher and depending on the market demand (society, Europe, economy).

5. Giveaway from a well-known store for games promoted from a fake Facebook user (money, disreputation).
6. Raising money for people in need, where a like brings in one dollar, a comment or a sharing much more to the weak person (society, money, stereotypes).
7. A post in which women's dresses and bodies shape the word OBEY. With an AI generated image this post aims to alert us to the existence of a dark high uncontrolled force that manipulates humanity – "if you see it, you can't unsee it" is the usual phrase in this category of fake news.
8. TIME's cover and the article that announces that an extinct species of wolves were reborn using stored DNA by scientists (society, science).
9. The only true article according to which AI has managed to solve in 48 hours a medical problem, previously solved by scientists after ten years of continuous research, effort, collaborations. The article, although true (at least till the time that this paper is being written), had an AI image generated. The "photo" was related to the article in two ways: explicit (showing robots in a lab) and implicit (even the photo was an AI artifact/achievement).

As already mentioned, the articles had some main characteristics of fake news (see Table 1):

1. Instead of the facts it writes down rumors and conspiracy theories. Part of the article has words in bold or italics or in brackets to underline or to imply there is something more under the lines. It raises anger in a proper and truth-appearing language.
2. Presents partly the truth since there is no reference that this cross was Saint Peter's cross. The funeral ceremony took place at Saint Peter's cathedral so the reverse cross is a symbol of a martyr, not of Satan. Contains sentimental language, words with mistakes, lots of !!!!! and skull emojis.
3. It has sentimental language, lots of !!!!! and a fake image (not AI).
4. Although it has not concrete information (when this will happen?), it uses a truth-appearing language.
5. Lots of !!!! and fake profiles commenting how happy and lucky people were after receiving the gifts without paying[22].
6. Sentimental and pitiful language, even false profile comments that could make you either write something nice or confront the rudeness.
7. Sentimental language – fake image.
8. Sentimental, but truth-appearing language, distortion of scientific truth, cultivates the false idea that we don't have to worry about species extinction since we can recreate them, probably advertises a big company.

Table 1: A table of linguistic and other features of fake news for every article or post that was chosen

---

[22] A pupil noticed that one comment had a day of publication that was not possible, it was very near to the giveaway announcement (a couple of hours).

Pupils began to elaborate the articles working in groups. They picked up names for their groups adopting a concrete identity: Alpha News, Future Journalism, From Minute to Minute, Quick and Helpful Journalism. Then they had to discuss and answer a questionnaire given to them. They were asked to find the title, the writer, the date and to decide what kind of text it is (article in a newspaper, maybe in a digital one, a post from a professional journalist or from someone just interested in something). They gave reasons for reading or omitting an article and underlined its most interesting parts.

We dedicated time also to visual literacy that is related to misinformation/disinformation. People were asked to see the photos, if any, and consider if they are relative to the content and if they are real or not.

Every group presented their articles and tried to discover if they are true or not. They also presented every characteristic considered unusual by them. A teacher, in this moment, could be their facilitator and make appropriate questions to lead them to spot the fake news features (Sipitanos, 2025).

After that, they voted on which one was the only true one. Except from one pupil, who found the true article[23], all pupils thought that the article “1” was the real one and that article “9” was fake for two reasons: they couldn’t believe in the AI possibilities, and they thought that the AI image was a trace of fakeness.

Then we talked about which article was more interesting for them and why. This discussion is important because we tend to accept easily as true what represents something that goes along with our aspects.

The last part of the project

The last activity was to let them write their own fake news. In this way pupils learn better to trace them. Some of the fake news articles they wrote are:

“The school mini market sells spoiled food? On May 2, 2025, a nine-year-old child bought a cheese pie and chocolate milk. The boy was taken to the nearest hospital, where he was diagnosed with gastroenteritis. Fortunately, he is fine and back at home now. Is this a new virus or was it caused by something he consumed?”

The title is striking and presents news almost as confirmed, even if it is not fully backed by facts. The question at the end implicates the school mini market and the goal is to have it closed so that a new catering company can take over. Additionally, it could aim to discredit the milk.

Pupils also wrote some true articles alongside the fake ones.

---

[23] The true article is the last one. The pupils preferred the first one because they were informed about the fire as a fact, so they easily accepted every aspect on who may have caused it – they were searching for a suspect of an arson. All the sentimental world of hate and revenge was presented to them as the only reality.

Conclusion

Using the knowledge obtained from the S.HI.E.L.D. Vs Disinfo initiative we presented a media literacy project based on fake titles creation, fake news reading and their characteristics, discussion and interviews (between pupils).

Pupils have a notion, an idea of what fake news are, before being taught at school. Their titles were similar to fake news that exist even nowadays. Articles that had the characteristics of the fake news were chosen although they were “neutral” to the Greek political reality. It seems that among characteristics of fake news my pupils spotted sentimental language, polarization, and anger. Some pupils brought to the classroom their families’ political ideas when asked to write fake and true news. When asked which one was their favorite article, they chose the “sweetest” one, they were excited by the photo of the small “puppies” (wolves) of the recreation of the extinct wolves’ species article (article 8). Maybe children of that age are impressed more by the photos. The next step could have been to publish a fake news newspaper and give it to younger pupils. Then we could make our pupils see the consequences of misinformation.

The implementation of these activities demonstrated the importance of integrating media literacy into early education. Pupils responded showing interest and curiosity to the ways that fake news spread and they were satisfied to have participated in a project that connects real life competence (media literacy) with school activities.

References

Καλογερόπουλος, Α. (2024, Ιούνιος). *Οι Ψηφιακές Ειδήσεις στην Ελλάδα το 2024.* Dianeosis.org. https://www.dianeosis.org/2024/06/oi-psifiakes-eidiseis-stin-ellada-2024/

Fischli, R., & Muldoon, J. (2024). Empowering Digital Democracy. *Perspectives on Politics. 22*(3), 819-835. https://doi.org/10.1017/S1537592724000409

Sipitanos, K. (2025). *Reading Strategies in the Age of Disinformation: Fake News and Recognition Practices.* Anatolia College, Kassandra Center for Educational Excellence.

Young, B. (2016, April 4). *Ridiculous History: The Great Moon Hoax of August 1835.* Howstuffworks.com. https://history.howstuffworks.com/historical-events/ridiculous-history-the-great-moon-hoax-1835.htm

Samioti, P., Sipitanos, K. & Katsarou, E. (Eds.)
Revisiting Disinformation: Critical Media Literacy Approaches
International Conference Proceedings

ISBN 978-618-5762-03-2