Preparing teachers for the AI Development in Education as an Innovative Asset

PAIDEIA

Co-funded by the European Union

UNIVERSITÀ
DEGLI STUDI
DI MILANO

CENTRO STUDI VILLA MONTESCA
FONDAZIONE

Preparing teachers for the AI Development in Education as an Innovative Asset

Report about how to use AI in education in spreading disinformation/misinformation. Evaluating the potential positive impact of ai in contrasting fake news

Authors: Giovanni Ziccardi, Arianna Arini, Giulia Pesci, Malwina Wójcik-Suffia

ISBN: 9788894149593

Edited in April 2025

This project has received funding from the European Education and Culture Executive Agency Erasmus Plus Programme Under Grant Agreement no 101132955.
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.

www.paideiaproject.eu

The PAIDEIA Project is a groundbreaking initiative funded by the European Commission through the European Agency for Education and Culture (EACEA). Our mission is to revolutionize education by integrating Artificial Intelligence (AI) into teaching and learning, empowering educators, and enhancing student outcomes.

CONSORTIUM

FONDAZIONE HALLGARTEN FRANCHETTI CENTRO STUDI VILLA MONTESCA, an Italian high education research center, coordinates PAIDEIA project. It has been historically founded in 1902 (then institutionally renewed) in the place where the Montessori method was experimented and published for the first time in 1909. It has an international vocation in educative and pedagogic research activities and has a very relevant experience in EU projects management and development. FCSVM has strong partnership with schools and other educational organizations, as well as a solid background in research-actions activities, fostering critical thinking, diversity as a value and innovation from the pedagogical and didactic point of view.

DUBLIN CITY UNIVERSITY

UCLL UNIVERSITY OF APPLIED SCIENCES

FUNDACIÓN UNIVERSIDAD SAN JORGE

NATIONAL INSPECTORATE OF EDUCATION - BULGARIA

GOVERNMENT OF MALTA - MINISTRY OF EDUCATION, SPORT, YOUTH, RESEARCH AND INNOVATION

UFFICIO SCOLASTICO REGIONALE TOSCANA

DEPARTAMENTO DE EDUCACIÓN, CULTURA Y DEPORTE DE GOBIERNO DE ARAGÓN

CENTER FOR EDUCATIONAL INITIATIVES

UNIVERSITÀ DEGLI STUDI DI MILANO STATALE

GOVERNORSHIP OF INSTABUL - T.C. İSTANBUL VALILIĞI

TABLE OF CONTENTS

1. INTRODUCTION

This report analyses the relationship between artificial intelligence and the phenomena of disinformation, misinformation, and fake news. The deliverable D2.5 is part of the tasks assigned by the PAIDEIA project to work package number 2.

Within the report, the themes related to the use of artificial intelligence in education and in spreading disinformation/misinformation and the potential positive impact of AI in counteracting fake news will be explored in depth.

Thanks to the analysis of the literature, research activities, and the study of responses provided to practical case studies, the report will provide a comparative evaluation of the impact that AI use in education can have in fighting misinformation and disinformation with an indication of potential responding strategies.

In the first section, the connection between artificial intelligence and the mentioned phenomena will be analysed.

The second section will examine the responses provided by the partner countries of the project regarding the discussion of practical case studies.

Finally, the third section will focus on potential response strategies.

2. METHODOLOGY

The search strategy employed in the present report, which refers to deliverable D2.5 (WP2), aimed at 1) gathering a comprehensive collection of existing literature and at 2) collecting data from semi-structured interviews and focus groups, about the use of AI in education in spreading disinformation/misinformation and in counteracting fake news. The extent of the research required a systematic approach to be adopted, consisting of different steps for each of the two sections (four steps for section i. and five steps for section ii.) and driven by the need to gain a broad understanding of the multifaceted nature of the phenomenon.

1) Section one: gathering of a comprehensive collection of existing literature about the use of AI in education in spreading disinformation/misinformation and in counteracting fake news

Step #1 has been the identification of the following research questions:

1. What is the definition of artificial intelligence?
2. What is the definition of disinformation, misinformation and fake news?
3. What are the specific risks of the spread of disinformation/misinformation for minors (interested target)?
4. What is the impact of AI in spreading disinformation/misinformation for the interested target? Is there a difference, in terms of spreading disinformation/misinformation, between AI as a means of such spread and AI as misleading content?
5. Why is disinformation/misinformation so engaging and attractive?
6. How can AI contrast fake news in education? How may AI be used in schools to contrast disinformation/misinformation? What AI tools are capable of detecting disinformation/misinformation?
7. What are the possible effects (either positive or negative) of the use of AI for minors' development? Can AI be an ally to the development of critical thinking? How can minors learn to question AI?

Step #2 consisted of the collection of a wide interdisciplinary spectrum of literature material about the use of AI in education in spreading disinformation/misinformation and in counteracting fake news. This first gathering outlined the boundaries of the review and was facilitated by using a set of generic keywords and more specific long tale keywords: as the research delved deeper into the specific literature, keywords were adapted to include long-tail keywords, as outlined in Table 1.

Criteria	Terms
Keywords	"Artificial Intelligence" OR "AI" or "Generative Artificial Intelligence" OR "GenAI" OR "Disinformation" OR "Misinformation" OR "Education" OR "School" OR "Teaching" OR "Learning" OR "Training" OR "Instruction" OR "Assessment" OR "Teacher" OR "Student" OR "Pupil" OR "Instructor" OR "LLM" OR "Large language model" OR "Deep fake" OR "Fake news" OR "Digital skills" OR "Sociologist" OR "Pedagogues" OR "Anthropologist" OR "Chatbot" OR "Infodemic" OR "Media" OR "Clickbait" OR "Troll" OR "Eco-chambers" OR "Filter-bubbles" OR "Viral content" OR "Trustworthy sources" OR "Education levels"
Long tale keywords	"Positive impact of AI in education" OR "Negative impact of AI in education" OR "Civic edu cation and AI" OR "Detecting disinformation" OR "Detecting misinformation" OR "Counter measures to disinformation/misinformation" OR "Best practices for the use of AI" OR "Extensive use of AI"

To achieve the goal of step #2, multiple academic databases, and search engines (such as Minerva – the access point to the bibliographic resources of the University of Milan –, Scopus, Google Scholar) have been leveraged to ensure coverage of the available literature, encompassing a diverse range of sources (scholarly articles, academic papers, newspapers, books, journals) from various disciplines. In addition, manual searches were conducted within specific selected journals and archives, many of which under paywall. Search limiters, such as data range and language were applied at the point of searching and are outlined in Table 2 below.

Criteria	Inclusion	Exclusion
Education Level	Primary and/or Secondary/Post-Primary Levels; Teacher Education	Early Childhood, Further Education and Training, Higher Education, Adult Education, non-education settings
Language	Published in English or in Italian	Non-English or Non-Italian
Publication type	Books; Peer-Reviewed Journal Articles; Book chapters; Published Conference Proceedings;	Non-Peer-Reviewed Articles, Blog Posts, Grey literature
Access	Full text available	Full text not available

Step #3 has been the selection of **step #2** findings using the criteria in Table 2.
Step #4 consisted of the cataloguing of **step #2** findings as selected under step #3 in Table 3.

The final number of papers included in the report is 47. The full bibliographical details are available in the report bibliography.

SID	Author(s)	Paper Title	Year	Country
S01	Abiola Gilbert et al.	The Role of Artificial Intelligence (AI) in Combatting Deepfakes and Digital Misinformation	2024	India
S02	Benanti, P.	Human in the loop. Decisioni umane e intelligenze artificiali	2022	Italy
S03	Bharathi Mohan, G. Et al.	An analysis of large language models: their impact and potential applications	2024	India
S04	Buitrago López, A. Et al.	Frameworks, Modeling and Simulations of Misinformation and Disinformation: A Systematic Literature Review	2024	Spain
S05	Castells, M.	The Rise of the Network Society, The Information Age: Economy, Society and Culture	1996-1998	Spain
S06	Cavus, N., Goksu, M., Oktekin, B.	Real-time fake news detection in online social networks: FANDC Cloud-based system	2024	International
S07	UE Commission	Libro verde vivere e lavorare nella società dell'informazione: priorità alla dimensione umana	1996	Belgium
S08	Corazza, P.	Filter bubbles e echo chambers: origini pre-digitali e elementi di novità. Riflessioni dalla prospettiva della media education	2022	Italy
S09	Crawford, K.	Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence	2021	USA
S10	Cugler de Moraes Silva, E., Carlos Vaz, J.	What characteristics define disinformation and fake news?: review of taxonomies and definitions	2024	Brasil
S11	Deibert, R. J.	Parchment, Printing and Hypermedia. Communication and World Order Transformation	1997	Canada
S12	Deepak, P. Et al.	Geo-Political Bias in Fake News Detection AI: The Case of Affect	2024	U.K.
S13	Dutta, S. et al.	Enhancing Educational Adaptability: A Review and Analysis of AI-Driven Adaptive Learning Platforms	2024	India

S14	European Commission. Directorate General for Education, Youth, Sport and Culture	Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators	2022	UE
S15	European Commission. Directorate General for Education, Youth, Sport and Culture	Guidelines for teachers and educators on tackling disinformation and promoting digital literacy through education and training	2022	UE
S16	Floridi, L.	Il verde e il blu. Idee ingenue per migliorare la politica	2020	Italy
S17	Goshevski, D., Veljanoska, J. and Hatziapostolou, T.	Review of Gamification Platforms for Higher Education	2017	Greece
S18	Greenwald A.G., Hamilton Krieger, L.	Implicit Bias: Scientific Foundations	2006	USA
S19	Ichino, A	Che cos'è una teoria del complotto? (E perché questa è una domanda importante)	2024	Italy
S20	Janiesch, C., Zschech, P., Heinrich, K.	Machine learning and deep learning	2021	Swiss
S21	Johnston, N.	Living in the World of Fake News: High School Students' Evaluation of Information from Social Media Sites	2020	Australia
S22	Jordan, M. I., Mitchell, T. M.	Machine learning: Trends, perspectives, and prospects	2015	USA
S23	Kahneman, D.	Thinking, fast and slow	2011	USA
S24	Kumar, Ajay, James W. Taylor	Feature Importance in the Age of Explainable AI: Case Study of Detecting Fake News & Misinformation via a Multi-Modal Framework	2024	U.K.
S25	Kurni, M., Mohammed, M.S. and Srinivasa, K.G.	AI-Enabled Gamification in Education	2023	India

S26	Lai, Vivian, Chenhao Tan	On Human Predictions with Explanations and Predictions of Machine Learning Models: A Case Study on Deception Detection	2024	USA
S27	Machlup, F.	The production and distribution of knowledge in the United States	1962	USA
S28	Mienye, I.D., Swart, T.G.	A Comprehensive Review of Deep Learning: Architectures, Recent Advances, and Applications	2024	International
S29	Molenaar, I. et al.	Artificial Intelligence and Education: Different Perceptions and Ethical Directions	2025	U.K.
S30	Montoro Montarroso, A. Et al.	Fighting Disinformation with Artificial Intelligence: Fundamentals, Advances and Challenges	2023	Spain
S31	Pawlicka, A. Et al.	AI vs Linguistic-Based Human Judgement: Bridging the Gap in Pursuit of Truth for Fake News Detection	2024	Poland
S32	Pariser, E.	The filter bubble: what the Internet is hiding from you	2011	USA
S33	Pérez-Escoda, A.	Infodemic and Fake News Turning Shift for Media: Distrust among University Students	2022	International
S34	Sarker, I.H.	Machine Learning: Algorithms, Real-World Applications and Research Directions	2021	India
S35	Saxena, D., Yasobant, S.	Information Overload	2022	India
S36	Selnes, F. N.	Fake News on Social Media: Understanding Teens' (Dis)Engagement with News	2024	Norway
S37	Skumanich, A., Kim, H. K.	Modes of Analyzing Disinformation Narratives With AI/ML/Text Mining to Assist in Mitigating the Weaponization of Social Media	2024	USA
S38	Spitale, G., Biller-Andorno, N., Germani, F.	AI model GPT-3 (dis)informs us better than humans	2023	Swiss
S39	Suffia, G.	Pulire l'infosfera. Intelligenza artificiale e contrasto alla disinformazione	2022	Italy
S40	Turing, A. M.	Computing Machinery and Intelligence	1950	U.K.
S41	Tversky, A., Kahneman, D.	Judgment under Uncertainty: Heuristics and Biases	1974	USA

S42	Valencia-Arias, A. Et al.	Understanding the Spread of Fake News: An Approach from the Perspective of Young People	2023	Colombia
S43	Vartiainen, H. Et al.	More than Fabricated News Reports: Children's Perspectives and Experiences of Fake News	2023	Finland
S44	Wurman, R. S.	Information Anxiety	1989	USA
S45	World Health Organization	Disinformation and Public Health	2024	International
S46	Yankoski, M.	Meme Warfare: AI Countermeasures to Disinformation Should Focus on Popular, Not Perfect, Fakes	2021	USA
S47	Ziccardi, G.	Diritti digitali. Informatica giuridica per le nuove professioni	2022	Italy

2) Section two: collecting data from semi-structured interviews and focus groups, about the use of AI in education in spreading disinformation/misinformation and in counteracting fake news

Step #1 has been the research designing with the aim to explore the role of AI in education regarding its potential in spreading disinformation/misinformation and its efficacy in counteracting fake news. First, the issue of the use of artificial intelligence in education was explored, and then this use in countering misinformation and disinformation was further explored. Finally, potential strategies for responding to issues that may arise from the use of AI in the area of false, distorted or manipulated information were reflected upon. The selected approach was a qualitative research design, using semi-structured interviews and focus groups as primary data collection methods.

Step #2 consisted of the selection of participants through purposive sampling (on a voluntary basis) to ensure that individuals with relevant experiences and expertise are included. The target participants will include sociologists, anthropologists and pedagogues for the semi-structured interviews; teachers for the focus groups.

First, four experts were identified among sociologists, anthropologists and pedagogues to be interviewed (the study target of the questionnaire concerned the school age group between 6 and 18 years old) and then each PAIDEIA partner identified at least 10 teachers (5 for the 6-11 age group and 5 for the 11-18 age group) to participate to subsequent focus groups. Teachers were contacted independently by each partner country, also involving regional school offices.

Step #3 required the development of semi-structured interviews' and focus groups' guidelines, comprising open-ended questions that allow flexibility and in-depth exploration. Such guidelines are outlined in Table 4.

Criteria	Semi-structured interviews	Focus groups
Process	Questionnaires	Questionnaires and case study
Language	Submitted in Italian.	Conducted in the native language of each PAIDEIA partner. Results sent back in English.
Duration	-	2 hours (approximately)
Recording	No	No
Ethical considerations	Informed consent, confidentiality, ethical approval by the institutional ethic committee.	Informed consent, confidentiality, possible ethical approval by the institutional ethic committee (if required).
Questions	Session 1: Artificial intelligence 1)From your perspective, what impact has the advent of artificial intelligence had in today's society and educational context? 2) In your experience, what should ethical use of artificial intelligence in the classroom look like? 3) From your perspective, what is the teachers' perception of artificial intelligence? 4) From your perspective, what is the students' perception of artificial intelligence? 5) What is the possible role of artificial intelligence in the development of a child's critical thinking? 6) Are there any artificial intelligence tools that can have a positive impact on a child's growth and education? 7) Are there any artificial intelligence tools that can negatively impact a child's growth and education? Session 2: Misinformation, disinformation and fake news 1) Can a cultural context enable the spread of misinformation? If so, how?	General Information: 1) How many participants teach at primary, post-primary, etc. Level? 2) How long have participants been teaching? Introduction: AI and Fake news 1) From your perspective, what impact has the advent of AI had in today's society and educational context? 2) Are there AI tools that can have a positive/negative impact on a child's growth and education? 3) From your perspective, what is the students' perception of misinformation, disinformation and fake news? 4) Do any artificial intelligence tools exist/have you used in your professional experience that contribute, from your point of view, to the spread of disinformation? 5) Do any artificial intelligence tools exist/have you used in your professional experience that contribute, from your point of view to counter disinformation? Case Study 1: primary level school You are a teacher of a class of children between 8 and 10 years old. Just before a break from lessons aket h holidays, you set your students an assignment. The assignment asked your students to write a research paper. The

	2) What are the elements that make fake news attractive, engaging and widespread for minors? 3) Can misinformation and disinformation have an impact in raising a child? If so, how? 4) What are the major risks of using artificial intelligence systems in the context of disinformation/misinformation? 5) Are there effective tools for recognizing disinformation and misinformation? 6) What are the major benefits of using artificial intelligence systems to counteract misinformation/disinformation? Session 3: Information required for the creation of "case studies" With the results of this questionnaire, UMIL will need to develop case studies to be later presented to teachers. These case studies will focus on AI, misinformation, disinformation, and fake news. The case studies must be realistic and reflect situations that could genuinely occur in a classroom context. To such purpose, what do you think are the topics to be included in the case studies described above? What do you think are the elements to consider in identifying the positive impact of AI in recognizing and countering disinformation?	essay was about a historical period. When the students returned from holiday, they give you their homework. Correcting the homework, you realise that 80 % delivered an essay with the same repeating mistakes. The essays are very well written, but contain precisely the same inaccuracies. The remaining 20 %, on the other hand, submit different homework, some written better than others, but with no errors from a historical perspective. - Is there any evidence that could make you think that the tasks containing errors were processed with the help of AI (e.g. ChatGPT)? - How do you decide to address this issue? Would you discuss AI directly with the students? - Considering that your students use AI tools, on which aspects would you focus your work with the class in the long term? Case Study 2: secondary level school You are a teacher of a class of students aged between 16 and 18. In your country, the electoral campaign for the upcoming general elections is underway. During your lesson, talking about current affairs, one of your students intervenes aket hatg a candidate and claiming to have seen a video in which the candidate declares war on Italy. Class discussion begins: many students support the point of view of the student who intervened, others remain silent, one intervenes saying that the video is fake. - How do you decide to handle the debate? - What tools do you advise your students to use to aket hat true from false information? - Given the situation, what are the aspects you think you should intervene on in the long term with the class? Closing remarks and areas not covered in questions provided

		1) Ask the participants if they have any other comments to aket hat were not raised by the questions asked. 2) Ask the participants if they have any questions for you (the interviewer) about the interview process or the project in general. 3) Inform the participants of the next steps – data processing, timeline for publication of the report. 4) Thank the participants for their time.
Participants	4 among sociologists, anthropologists and pedagogues	Almost 10 teachers (5 from primary school and 5 from secondary school)
Used platforms	Survio	Zoom or another online platform

Step #4 has been the thematic analysis of the collected data, aimed at developing a comprehensive understanding of it. Themes were refined through iterative readings and discussions among the research team, and data was triangulated to ensure the validity and reliability of findings. Divergent views were explored to gain a nuanced understanding of the issues.

Step #5 consisted of the reporting of the collected data in a manner that highlights key themes and insights. Quotations from participants were included with the insurance of anonymity.

3. THE CONNECTION BETWEEN AI AND MISINFORMATION/DISINFORMATION

3.1 Introduction

Before analysing the concrete connection between artificial intelligence and the phenomenon dissemination of false, distorted or manipulated content, it is necessary to provide definitions that help better understand the phenomenon itself.

«Misinformation is the spread of false information without the intent to mislead» (World Health Organization, 2024). «Disinformation is designed or spread will full knowledge of it being false (information has been manipulated), as part of an intention to deceive and cause harm» (World Health Organization, 2024).

The term "fake news" is also often used. It is a more generic term that refers to false news circulated both online and offline through traditional media.

For the Cambridge Dictionary "fake news" means "false stories that appear to be news, spread on the internet or using other media, usually created to influence political views or as a joke".

The spread of false, distorted, or manipulated news can have significant impacts on society. This phenomenon can lead to consequences, sometimes severe, in terms of security, justice, and political, social, and economic stability. Disinformation and misinformation can cause harm to public and individual health and can greatly undermine the right to education and learning.

For these reasons, it is important to emphasise the difference between misinformation and disinformation.

People who spread misinformation may genuinely believe that the information in question is true, interesting, or useful to know. Therefore, these individuals do not necessarily have harmful or malicious intentions toward the recipients of the information.

Disinformation, on the other hand, is deliberately created and disseminated to deceive its recipients. By intentionally designing a false piece of information, the senders may have various objectives: economic gain, promoting a specific ideology, creating social disorder, or garnering political support.

Disinformation has ancient origins, closely tied to the history of communication (Deibert, 1997). It should be noted, however, that recent technological developments have contributed to the growth of the phenomenon, fostering unique characteristics and consequences. Over the years, information has achieved high levels of pervasiveness, strength, and dissemination (Suffia, 2022), which are characteristic of the information society.

The term "information society" (Machlup 1962, Bell 1973, Castells 1996) refers to the contemporary society, characterized by information and communication technologies, and specifically by the Internet.

The European Union, in fact, wrote in 1996 that we were experiencing a historical period of technological transformations, driven by the development and the increasingly widespread application of information and communication technologies (ICT); a period that brought with it significant potential for wealth creation, higher standards of living, and improved services (Commission of the European Communities, 1996).

In the 1990s, ICT had already become an integral part of society – albeit with inequalities in terms of accessibility among citizens and geographic areas – providing tools and services useful in households, workplaces, and many other circumstances. The information society thus became a reality of everyday life for many people, transforming the structure of society as it was traditionally known. The production of goods and services increasingly relies on knowledge and information.

Today, we speak of a society where life takes place "onlife", within a digital and analog space that has been renamed the "Infosphere" (Floridi, 2020).

The phenomena of disinformation and misinformation today are based on the characteristics of the information society, new technologies, and people's cognitive biases.

The term "filter bubble" (Pariser, 2012) is used to describe how search engines and their algorithms can personalise the content provided to users, creating a "bubble" that filters and determines the types of content a user can view.

Information is thus filtered, allowing into the bubble only what aligns with the user's interests and beliefs, while excluding everything else.

This significantly limits dialogue and greatly reduces the exchange of ideas.

The "filter bubble" finds its fullest expression online, in search engines and social networks. However, it is worth noting that this dynamic can also occur offline, when an individual surrounds themselves with specific types of information while excluding others.

In recent years, information has been facing a new challenge brought about by the advent of artificial intelligence; a complex technology that is increasingly becoming accessible to everyone. Artificial intelligence plays a significant role today in both the spread and the mitigation of misinformation and disinformation. On one hand, AI-based tools can generate false or misleading content with high levels of realism, thereby facilitating the creation of fake news, fabricated content or "deepfakes". As a result, it can be challenging for individuals to distinguish between content generated by AI systems and that created by humans (Spitale, Andorno, Germani, 2023). Additionally, issues related to AI-mediated information may involve freedom of expression, privacy rights, and copyright protection. Defining AI's responsibilities in cases where damages or

problems occur also proves to be a complex matter. On the other hand, however, as will be seen in the present analysis, artificial intelligence can also be utilised to identify the spread of false information, thereby mitigating the impact of the phenomenon. This process can be achieved through the creation and development of AI tools which, when combined with human intervention and critical thinking, can help individuals discern reality from fiction.

3.2 Literature Review

3.2.1 Artificial intelligence

Attempts to define artificial intelligence date back many years, when the digital revolution was still far off.

In 1950, Alan M. Turing posed the famous question "Can machines think?" and devised the "Turing test", also known as "the imitation game" (Turing, 1950).

The test involves determining whether the responses provided by a machine can be distinguished from those given by a human, effectively differentiating the "reasoning" of the machine from that of a person.

In 1956, the famous Dartmouth Conference was held, organised by John McCarthy and several other pioneers of computing at the time. It was during this conference that the field of research in artificial intelligence was formally identified.

John McCarthy, in fact, defined artificial intelligence as the science and engineering of creating intelligent machines.

From that moment onward, the study of artificial intelligence has primarily focused on the creation of a sort of replica of human intelligence within machines.

Artificial intelligence remains, to this day, a concept difficult to explain and define. When considering the idea of "embedding" a brain within a machine, ethical, practical, computational, legal, and accountability issues immediately arise, along with significant concerns and fears. The notion of a concept housed within a machine is highly evocative, inspiring fascinating and frightening stories and legends.

Today, rather than imagining a thinking machine's brain as similar or identical to that of humans, we speak of machine learning. Artificial intelligence is now capable of acquiring and processing information, then reworking it to produce outputs.

Machine learning is a branch of artificial intelligence that operates through the development of algorithms and techniques that enable machines to learn automatically.

This process occurs by improving the experience of computer programs with respect to a certain class of tasks and performance measures (Jordan, Mitchell 2015).
The task of building analytical models to perform specific activities and produce specific outputs, such as object detection or natural language translation, is thus automated. This goal is achieved by applying algorithms that learn from specific training data without being explicitly programmed. Machine learning learns from previous activities and computations, extracting data and information from massive databases. Additionally, machine learning systems can produce reliable and repeatable decisions (Janiesch, Zschech, Heinrich, 2020).
Machine learning, therefore, consists of the ability of artificial intelligence systems to learn and improve automatically through experience.
Training methods of machine learning algorithms can be divided into four categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning.
Supervised learning involves the task of learning a function that takes inputs to produce outputs, using labelled training data and a collection of training examples. The most common tasks in supervised learning are data classification and regression.

Unsupervised learning analyses unlabelled datasets without human intervention. Unlike supervised learning, it does not follow a data-guided process. It is used to extract generative features, identify trends, explore data, and group results.

Semi-supervised learning operates on both labelled and unlabelled data. This model can provide better predictions compared to those using only labelled data.
Reinforcement learning, on the other hand, is a type of machine learning that allows software and machines to automatically evaluate behaviour in a specific context or environment to improve efficiency. This system is based on rewards or penalties, and its goal is to take actions that maximize rewards or minimize risks (Sarker, 2021).
Nowadays, machine learning is widely present and utilized in people's daily lives. It can be found in home automation (the IoT devices), entertainment and leisure (such as video games or video and photo editing applications), virtual assistants and chatbots, within smartphones and applications (like facial recognition to unlock devices, predictive keyboards, or enhanced photography), in transportation (maps, vehicle-sharing apps, and other means), in finance, health and wellness, e-commerce, online search and content filtering (as seen with spam and security filters), and social media (personalized content, facial recognition and automatic tagging in photos, content moderation).
Another branch of artificial intelligence needs to be briefly outlined.
Deep learning is a subset of machine learning and utilizes architecture structures with numerous layers of nodes or neurons, where each layer is designed to model patterns using increasingly

complex data. Today, deep learning has evolved to include highly sophisticated neural networks capable of performing a variety of tasks, even complex ones, such as image recognition and natural language processing (Mienye, Swart, 2024).

In recent years, there has been increasing discussion about "large language models", highly advanced artificial intelligence models designed to understand and generate natural language. These models are continuously trained using vast datasets of text to learn the structures, rules, and nuances of human language.

Large language models (LLMs) are based on deep learning techniques and are meticulously trained on massive textual datasets. Today, LLMs can provide logical responses by first interpreting highly complex verbal patterns and then effectively reusing them in a wide range of real-world scenarios (Bharathi Mohan, Prasanna Kumar, Vishal Krishh 2024).

These artificial intelligence systems, though sophisticated, do not yet understand human natural language. They combine words, decipher results, process training data (input), and return a relevant outcome (output). LLMs calculate probabilities within datasets and texts to find and generate the correct responses. Increasingly, the extreme levels of precision these systems are achieving are becoming evident.

Despite being highly sophisticated and providing exceptionally accurate results, human oversight of the outcomes remains essential.

The recent European regulation on artificial intelligence has also established the importance of human oversight.

Article 14 outlines the requirements for human oversight of high-risk AI systems to ensure they are used safely and responsibly while protecting health, safety, and fundamental rights. It mandates that these systems must be designed and developed with tools that allow natural persons to effectively oversee their operation during use. The primary goal of this oversight is to mitigate risks that may arise from the intended use or foreseeable misuse of the system, even when other safety measures are in place.

The measures for oversight must be proportional to the risks, level of autonomy, and context in which the system operates. These measures can either be built into the system by the provider before its deployment or implemented by the deployer, as identified by the provider. High-risk AI systems must be provided in a way that allows individuals responsible for their oversight to fully understand their capabilities and limitations, monitor their operation for anomalies or malfunctions, and remain vigilant against over-reliance on system outputs (automation bias). These individuals must also be able to interpret system outputs correctly, override or disregard them when necessary, and safely intervene or halt the system using appropriate mechanisms, such as a "stop" button.

For certain high-risk systems, particularly those specified in Annex III, point 1(a) of the regulation,

additional safeguards are required. Actions or decisions based on system outputs must be verified and confirmed by at least two competent and authorized individuals. However, this requirement does not apply to systems used in law enforcement, migration, border control, or asylum when deemed disproportionate under Union or national law. Overall, the article underscores the importance of human oversight as a critical element in the responsible deployment and operation of high-risk AI systems.

People are therefore increasingly engaging in interactions with these artificial intelligence models. "Prompts", in fact, are gaining more and more importance, as the results directly depend on their effective use.

One aspect that emerges and requires particular attention is the fact that, today, interacting with powerful and sophisticated technological and artificial intelligence systems no longer requires learning and practicing a specific programming language. Natural language is now used in communication and interaction between humans and machines.

3.2.2 Misinformation and disinformation

Before proceeding with the analysis of the interaction between artificial intelligence and disinformation and misinformation, it is necessary to briefly outline the two phenomena in question.

First and foremost, it is important to make an initial and fundamental distinction.

Disinformation and misinformation are distinct phenomena. They share certain characteristics, but the reasons behind the creation and dissemination of their content are significantly different. Today, various terms are used to refer to information that can be classified as false, untrue, partially true, or manipulated. Terms like "fake news", "disinformation", and "misinformation". Disinformation specifically refers to false information deliberately designed to mislead those who meet it.

Misinformation, instead, refers to the spread of false information that is mostly disseminated unintentionally. Misinformation, therefore, refers to false, inaccurate, or imprecise information that is spread unintentionally. Disinformation, on the other hand, involves the deliberate dissemination of such information with the intent to deceive or manipulate (Buitrago López, Pastor-Galindo, Ruipérez-Valiente, 2024).

Disinformation is a multifaceted and polysemic phenomenon, involving information that is entirely or partially false, created by one or more individuals with the aim of misleading or manipulating reality through deliberately crafted content. The phenomenon is amplified by the individual

characteristics of the receiver in their interpretation of the facts and the contextual factors in which it arises. The term "fake news", although somewhat generic, is also used to refer to information that is entirely or partially false, created and disseminated by a sender with the aim of misleading the "public", using content designed to manipulate reality (Silva, Vaz, 2024).
It is possible to speak of "malinformation" when information, even if true, is deliberately disseminated in a different context with the intent to confuse, manipulate, or cause harm. Malinformation, therefore, shares with disinformation the intention to harm individuals or the community, but it is distinct in that the information used is true, yet presented out of context.
Other terms are essential to understand when considering the phenomena of disinformation and misinformation.
The term "information overload" is used to describe an excessive or disproportionate amount of information, often complex, that exceeds an individual's capacity to process, understand, and effectively use it. This phenomenon compromises the quality of decision-making, the understanding of events, and the dissemination of information itself.
Information overload occurs when the volume or complexity of accessible information surpasses an individual's ability to process it within a given timeframe (Saxena, Yasobant, 2022). The concept of "information anxiety" is also discussed, referring to the sense of frustration, stress, incapacity, or anxiety that individuals experience when they are unable to process or understand information effectively, often due to the complexity or excessive volume of data and information (Wurman, 1989).

"Post-truth", named the word of the year 2016 by Oxford Dictionaires, indicates the circumstance in which people tend to respond more to emotions and beliefs than to facts (Oxford Learner's Dictionaires). The term has been widely used in historical and political events of recent years and refers to the possibility that appeals to emotions and beliefs may be more influential in shaping public opinion than objective facts and reality.

The term "clickbait" refers to the practice of crafting titles designed to encourage readers to click on articles. These titles often use sensationalistic or alarming language and provide minimal information to capture attention and spark curiosity. Occasionally, this practice can be intentionally misleading (Thirumala, Ferracane, 2023).

"Conspiracy theories" constitute the phenomenon of conspiracism (the proliferation of conspiracy theories about various events and aspects of reality). Conspiracy theories are characterized by their opposition to the official version of events supported by epistemic authorities and accepted by most people. These theories propose alternative "counter-stories", typically supported by weak or insufficient evidence. They are often persistently upheld by individuals who claim that conspirators are actively working to cover up the conspiracy itself, which is allegedly created to deceive.

Conspiracy theories, therefore, are theories that: refer to a conspiracy as the cause of the events or facts in question; oppose an official version of the facts; and are rooted in a conspiratorial mindset that asserts powerful conspirators are acting to mislead others (Ichino, 2024).

Another element of essential importance is represented by "cognitive biases".

In the context of disinformation and misinformation, beliefs, biases, and judgments play a crucial role. People often rely on heuristics, leading to systematic errors in judgment. Biases often result in errors of judgment or misinterpretation of data and information. Biases, in fact, are embedded in everyday decision-making processes, many of which can have a significant impact on society.

Biases can be different and relate to distinct responses of human beings.

Biases in question are being studied, and a precise and timely classification of them would be very useful placed in the context of the information society.

Beliefs and biases inherent in human minds can relate to how people remember information, how they interact with others, how they explain or justify their own behaviours, their self-esteem and regard for themselves and others, and the difficulty in changing their minds or questioning their opinions.

It is interesting to note that these biases can also affect the operation of digital technology and artificial intelligence.

Biases in artificial intelligence, for example, may refer to errors that can occur during a predictive process of generalization, that is, a misclassification that is repeated systematically or consistently. (Crawford, 2021)

Bias can, therefore, result in a technical error, or connect in human beliefs, stereotypes or discrimination. Cognitive biases, therefore, constitute the ways in which human judgments determine the systematic deviation of probabilistic expectations (Tversky, Kahneman, 1974). Unconscious attitudes, beliefs, prejudices and stereotypes can produce behaviours that do not correspond to the beliefs or principles explicitly stated or endorsed by the subject in question (Greenwald, Hamilton Krieger, 2006).

From these reasonings, it emerges that emotions and emotional states play essential roles in the context of disinformation and misinformation. Understanding the interaction between emotions and biases is crucial for improving decision-making processes and their outcomes. Emotional states can influence attention, risk perception, and confidence.

"Echo chambers" are environments where individuals primarily interact with people who share similar opinions, excluding opposing viewpoints. They create a form of isolation that reinforces preexisting beliefs, increases polarization, and excludes differing opinions (Corazza, 2022).

"Filter bubbles" (Pariser, 2012) refer more specifically to environments filtered by algorithms used online and on social media. In these environments, algorithms personalize the content presented to the user based on their past behaviour (clicks, search history, "likes", and shares).

This process isolates individuals from information that might contradict or challenge their beliefs, creating a bubble that excludes anything outside their sphere of interests or convictions.

The digital revolution first, and artificial intelligence later, are driving significant changes and innovations in the realm of disinformation and misinformation. Deepfakes, specifically, but also the relatively simple and accessible use of applications and systems capable of manipulating or distorting reality, present and will continue to present new challenges in combating false or manipulated information.

3.3 Conclusions and considerations

Artificial intelligence is becoming increasingly powerful and sophisticated, but it is important to remember that it is not infallible and, in many ways, still differs significantly from human intelligence.

An example of this statement can be illustrated by artificial intelligence hallucinations. These hallucinations occur particularly in generative AI systems and involve the generation of outputs that are inaccurate or entirely incorrect, failing to fully address the initial request.

The interaction between artificial intelligence, misinformation, and disinformation is becoming increasingly complex and can have significant repercussions in education and training. In this regard, it is considered essential to invest in digital literacy, AI literacy, and particularly the ethical use of artificial intelligence.

Generative AI is increasingly present in the daily lives of students and teachers in various parts of the world. It is therefore crucial to understand AI and the phenomena of disinformation and misinformation.

Understanding AI involves basic knowledge of how these systems operate and awareness that AI primarily relies on the collection and processing of data; data that may embed human biases. It is equally important to understand the phenomena of misinformation and disinformation both from a technical perspective and regarding their repercussions and consequences, including psychological and emotional effects.

In this context, critical thinking remains arguably the most effective tool to cultivate and leverage for navigating an environment characterized by the presence of AI and phenomena associated with the spread of false information.

Technology and AI can play a pivotal role, either positively or negatively, in the dissemination of disinformation, misinformation, and malinformation. As discussed, algorithms that power social media and networks today can promote personalized (user-profile-based) and often polarized

content, reinforcing users' pre-existing beliefs or fueling violence and aggression toward differing opinions.

Additionally, "deepfake" technology has revolutionized the creation and dissemination of false information.

A "deepfake" can be a photograph, video, or audio created using artificial intelligence software ("deep" as in "deep learning") that, by utilizing real content, can modify or recreate, in an extremely realistic way, the features, movements, or voice of a person (Italian Data Protection Authority).

Deepfakes represent a technique and a phenomenon that must be closely monitored and understood as much as possible, enabling individuals to critically assess the content they may encounter.

AI is enabling the creation of increasingly realistic content, making it sometimes very challenging to distinguish reality from fiction.

This phenomenon is highly significant: deepfakes can be used for political campaigns, to commit crimes, or to harm other people. Moreover, deepfakes can recreate archaeological sites that have been gone for centuries or represent historical events for which there is limited evidence.

Generative artificial intelligence can also produce disinformation. Human oversight and critical thinking play a crucial role in addressing this challenge. The correlation between artificial intelligence and misinformation or disinformation is therefore quite significant. For this reason, as previously noted, it is essential to understand all the elements under consideration.

Technology further amplifies spatial and temporal distances: disinformation can reach users who are, in every sense, far removed from the content they meet. Moreover, the reduction in empathy that can occur when interacting online and the false perception of anonymity can make online attacks even more aggressive and dangerous (Ziccardi, 2022).

Regarding the risks that artificial intelligence can pose to individuals, it is also necessary to briefly address the issue of personal data protection and privacy. As noted, artificial intelligence is trained and enhanced using data. In many cases, this data is personal and sensitive, often collected without the user being fully aware of it.

An additional essential consideration for this analysis concerns the digital divide. Particularly in the field of education, it is important to reflect on the disparities that may arise, geographically and socially, for instance, in relation to the use of technology and artificial intelligence.

The presence of increasingly sophisticated and powerful systems, yet inaccessible to most people, can lead to significant digital inequalities.

AI literacy and training in recognizing and countering misinformation and disinformation (for young people and students, but also for teachers and educators) must aim to start from the basics, striving to ensure that no one is left behind.

At the conclusion of this study, a more detailed analysis will be conducted on the functions and uses of artificial intelligence that enable the countering of disinformation and misinformation.
In the context of disinformation, and beyond, two concepts of "systems" used for decision-making have been developed.
"System 1" refers to the automatic, involuntary, and unconscious fast-thinking processes people apply daily (fast and intuitive thinking). "System 2", on the other hand, involves the slow, deliberate mechanism people use when reasoning is required to analyse and solve problems (slow and deliberate thinking) (Kahneman, 2011).
With artificial intelligence, the situation becomes even more complex, introducing what has been defined as a new cognitive system, "System 0" (Benanti, 2022). This system emerges from the interaction between humans and artificial intelligence and precedes System 1, incorporating AI biases and correlations.
Artificial intelligence is capable of processing vast amounts of information that the human mind would struggle to handle, particularly in a short amount of time. The computational and data-processing power of AI is truly remarkable.
On this point, it is crucial to reflect on the potential reduction in human reflection and critical thinking because of machines increasingly providing information.
The cognitive "System 2" is based on the outcome of slower reasoning, "System 1" on what has already been seen or heard, and "System 0" on results produced by artificial intelligence. Both System 1 and System 0 can, however, rely on data and information considered accurate. For this reason, critical thinking and reasoning remain indispensable.
Teachers have significant responsibilities regarding the education of students on the use of artificial intelligence and in combating disinformation and misinformation.
For this reason, promoting an ethical use and understanding of artificial intelligence is of fundamental importance.
The ethical use of artificial intelligence and its promotion should consider, at the very least, the following points: teachers' competence obligations; system transparency and their use; the protection of personal data and privacy; basic cybersecurity principles; risk assessment; the presence of inclusive environments and contexts; and the safeguarding of copyright.
Today, people of all ages and educational backgrounds, to varying degrees, are exposed to disinformation. In education and training, it is essential to promote awareness of these issues while encouraging a curious and positive approach to technology, which has become an integral part of everyday life for many.
Artificial intelligence (generative and otherwise) is now used in various areas of education and teaching, which will be explored in greater depth in the following pages. For this reason, it is useful to divide the interaction between artificial intelligence, misinformation, and disinformation

into two intersecting macro areas.

On one hand, there is the use of artificial intelligence by teachers for educational activities and personal development. On the other hand, there is its use by students for educational activities (such as completing assignments or conducting research) and outside the school context. Both teachers and students, in fact, can use artificial intelligence outside the classroom and gather information through AI or AI-generated or modified content.

These two areas converge in the use of artificial intelligence in shared educational activities between teachers and students. This shared use requires particular attention to understand how it can be effectively utilized and enhanced. Using AI for classroom debates, discussions about its uses, risks, opportunities, and ethical implications can prove extremely valuable and important.

4 CASE-STUDIES ANALYSIS AND EXPECTED IMPACTS

4.1 Introduction and methodology

The phenomena of misinformation, disinformation and artificial intelligence, as noted above, are complex and multifaceted.

An attempt has been made to find answers regarding their relationship. First, the issue of the use of artificial intelligence in education was explored, and then this use in countering misinformation and disinformation was further explored. Finally, potential strategies for responding to issues that may arise from the use of AI in the area of false, distorted or manipulated information were reflected upon.

In order to better address and understand these issues, Italian experts were first interviewed by submitting open-ended questionnaires. The experts contacted deal with subjects such as sociology, anthropology and legal sciences. They are people who work or have worked in research for the University of Milan and who have delved into the topics of misinformation, disinformation and the use of artificial intelligence.

The questionnaires were divided into two distinct purposes: the first concerned an in-depth discursive study of misinformation and the possible positive impact of AI in combating the phenomenon; the second, on the other hand, was aimed at gathering ideas for structuring subsequent focus groups on the topic.

The study target of the questionnaire concerned the school age group (between 6 and 18 years old).

UMIL sent the partners the same documents used to conduct the focus groups in Italy. UMIL sent two documents, one for the first age group of interest (6-11) and one for the second group (11-18). The documents included the areas to be filled in (with who conducted the focus group, the number of participants and the partner country) and the questions to be asked to the teachers, concluding with the case study to be discussed with the participants.

As for Italy, teachers were contacted by the project partner USR Toscana, which sent e-mails through the office contacts to identify teachers who could attend the meeting and who were aware of the topics of the meeting (misinformation, disinformation and fake news). The other partner countries were also asked to proceed independently in identifying teachers to be invited to participate in the project.

After receiving the answers to the questionnaires, we reflected on the themes that emerged and elaborated on the focus groups.

The focus groups addressed the following themes: 1) Current Usage of AI in Education; 2)

Opportunities of AI in Education; 3) The phenomenon of misinformation and disinformation; 4) The impact of the use of AI in education and in spreading misinformation/disinformation; 5) The impact of AI for students at all education levels in countering fake news.
Teachers participated in the focus groups, divided into two groups: primary-level teachers and second-level teachers.
Only four Italian experts were contacted for the questionnaires. All PAIDEIA project partner countries participated in the focus groups, interviewing teachers from each country.
The results of the questionnaires and focus groups were useful and allowed for stimulating discussions on the project topics.

4.2 Data and Information collected

Four Italian experts on sociology, anthropology, disinformation and misinformation, and artificial intelligence were contacted for the questionnaires.
The experts were first told about the project, its aims and objectives and then answered fourteen questions.
The first seven were encapsulated in the first session on artificial intelligence.
The next six in the second session on misinformation, disinformation and fake news. Finally, the last broader question included the information required for the creation of "case studies" to be included in subsequent focus groups.
Thus, four questionnaire responses were obtained from the Italian experts and fourteen focus group responses from the project partner countries, including Italy.

4.3 Findings

4.3.1. Questionnaires

For session number one on artificial intelligence, the Italian experts answered the following questions.

1. **From your perspective, what impact has the advent of artificial intelligence had in today's society and educational context?**

The answers reveal, first of all, the difficulty in framing a definition of artificial intelligence. "Artificial intelligence", in fact, is considered a very broad label, ranging from search engines and

their operation to tools such as ChatGPT. In short, very different tools with different benefits, risks and spin-offs.
What emerges is that artificial intelligence can have both positive and negative aspects. The positive ones concern research activities, making it possible to obtain an immense amount of data, the negative ones, for example, the possibility of plagiarism or copyright infringement.
Another reflection concerns the impact of "marketing" around AI: an impact that can be positive, negative, improving or worsening, liberating or constructive.
Another response suggests reflection on the "aggressive" entry of artificial intelligence into society, worsening a process of "cultural devastation" that had already begun with the advent of social networks.
Finally, artificial intelligence is held responsible for "revolutionizing" today's society, speeding up and optimizing processes and posing ethical challenges. In the educational context, on the other hand, artificial intelligence is deemed capable of facilitating the adaptation of learning methods, but the effects on children's and young people's attention and creativity are being tested.

2. **In your experience, what should ethical use of artificial intelligence in the classroom look like?**

The answers suggest balancing the two opposites: prohibition and free use. The importance of bringing programming into school curricula is underlined. The importance of not using AI uncritically is emphasised.
The ethical use of artificial intelligence, then, should aim to educate young people in critical thinking about conscious use (e.g. in relation to privacy, data processing and the risk of incorrect information). Finally, it is proposed that the process of reworking by the human being should never be missing.

3. **From your perspective, what is the teachers' perception of artificial intelligence?**

The answers reveal the difficulty in generalising in order to provide an answer. In fact, the answer may vary depending on many factors: age of teachers, training on the subject, IT skills.
Artificial intelligence is considered to make it easier and faster to perform the usual behaviours and actions of teaching. It emerges that some teachers have a negative view of the use of AI by students: they want to "do less", "cheat", as if it were the "degenerative upgrade" of "copy and paste" from the Internet.

4. **From your perspective, what is the students' perception of artificial**

intelligence?

The responses reveal both a great deal of curiosity and interest, and a consideration of AI as a tool that allows one to "do less". Creative experimentation and also the use of AI to make certain actions easier and faster is emphasised. Finally, the profound impact of AI on learning dynamics and timing is also reported; an impact that is still being studied.

5. **What is the possible role of artificial intelligence in the development of a child's critical thinking?**

Experts believe that if used as a stimulating medium and interlocutor, AI can activate interesting cognitive and metacognitive processes. It is also reported that artificial intelligence could make experiences, such as the one in Finland, that place critical thinking and countering misinformation as central to a child's curriculum, scalable and replicable.

Finally, scholars believe that students can be encouraged to ask questions, explore complex concepts and solve problems creatively. To enable this, it is crucial to foster an attitude of research and critical evaluation of information.

6. **Are there artificial intelligence tools that can have a positive impact on a child's growth and education?**

From the answers, it appears that each tool could have a positive or negative impact. To give positive examples, AI used as a self-assessment tool may have an interesting impact (e.g. by generating multiple "simulations" of the task/assessment cycle, the student may arrive more prepared for the real test/exam).

7. **Are there any artificial intelligence tools that can negatively impact a child's growth and education?**

Experts believe that the negative impact of the use of AI can result from all students being used passively without critical thinking. The risk of copyright infringement is reiterated.

Finally, the risk of addiction-like effects of technology use, isolation and reduced social interactions, passivity of use and exposure to age-inappropriate or incorrect content is mentioned.

For session number two on misinformation, disinformation and fake news, the Italian experts answered the following questions.

1. **Can a cultural context enable the spread of misinformation? If so, how?**

For experts, a cultural context can allow the spread of disinformation through the tools used for

information such as social networks, or by following certain situations and mechanisms that make it easy to spread disinformation. These include so-called "epistemic bubbles" (communities in which voices that contradict a certain opinion have been more or less deliberately excluded - think of a themed social group, in which dissenters are not allowed), or even worse "echo chambers" (communities in which voices that contradict a certain opinion are not excluded, but are systematically discredited on the basis of a radical manipulation of trust relations between those who belong to those communities and those who do not).
Poor media literacy or a lack of critical thinking when verifying information can also create the basis for disinformation.

2. **What are the elements that make fake news attractive, engaging and widespread for minors?**

For experts in general, fake news seduces us because it promises to satisfy some important psychological needs that we all have. «In the case of minors, I am thinking in particular of two needs that are particularly relevant. On the one hand, the "need for uniqueness" (to feel "a little bit special" - feeling like the guardians of important truths hidden from most). On the other, the need to belong, to feel part of a group (the community of those who have "opened their eyes" and do not go along with the "official versions"). Several studies have indeed underlined the central role of these two needs in determining the adherence to fake news of various kinds (in particular, of a conspiracy nature)».
Hence, reflection also emerges on the type of fake news, emphasising how, in general, any information that "speaks" to people's emotions, including children, can fascinate and capture attention. And fake news a fortiori plays on emotions and communicative effectiveness. The inexperience of the child is emphasised, which, in many cases, can play to the child's disadvantage.

3. **Can misinformation and disinformation have an impact in a child's growth? If so, how?**

The experts answer in the affirmative. Language and cognitive processes are part of a subject's biological transformation and "information bubbles" can determine an individual's life choices. On minors, misinformation can have a considerable impact. Adherence to fake news can be linked to a multiplicity of even anti-social behaviour. In extreme cases, then, adherence to certain disinformation theories can be linked to fanaticism and violence.
Disinformation and misinformation then can contribute to a distorted knowledge of reality: increasing fears, stereotyped beliefs or erroneous elements on which to base decisions.

4. **What are the major risks of using artificial intelligence systems in the context of disinformation/misinformation?**

Among the risks reported by the experts interviewed are the possibilities of imitation and falsification that AI opens up (such as deepfake), the opacity of the interests behind disinformation and that AI multiplies the effects of this conflict, and the difficulty in distinguishing between content based on true or historical facts and content that is not based on facts even if it appears realistic. Furthermore, algorithms can amplify and speed up the dissemination of misinformation and polarised views. Making disinformation through AI, finally, is now easier and cheaper than countering the phenomenon itself.

5. **Are there effective tools for recognizing disinformation and misinformation?**

For experts, there are now tools that can be used to recognise disinformation or misinformation and to initiate fact-checking procedures. In any case, it appears from the answers that critical thinking and an understanding of the psychological and social mechanisms governing the phenomena under scrutiny can help human beings recognise and counteract them.

6. **What are the major benefits of using artificial intelligence systems to counteract misinformation/disinformation?**

It is clear from the answers that artificial intelligence can be used to identify and report disinformation and misinformation. Indeed, AI is capable of analysing huge amounts of data in a short time, automating the verification of sources. AI to counter these phenomena could, therefore, keep pace with the use of AI itself for misinformation. In any case, the cognitive effort between disinformation and its countering is highlighted.

Finally, the interviewees were asked to provide ideas for structuring the case study to be submitted to the teachers of the PAIDEIA project partner countries in the context of the focus groups.

Regarding the topics to be included in the case studies, the following points emerged from the experts' answers:

1. the spatial and temporal distance of the uninformed from the place or time of the action/image being provided;
2. the distinction between satire and humour (in communication mainly on social media);
3. the verification of sources of fake news;
4. the distinction between manipulated videos and images and original ones;
5. using topics that closely concern young people;

6. raising awareness of cognitive bias.

In relation to the other central theme of this report, the experts were asked to report on the elements to be considered in identifying the positive impact of AI in recognizing and countering disinformation.
The responses first of all concerned the framing of artificial intelligence as a tool: a tool that should be evaluated in its being more or less able to support the user's critical thinking. Furthermore, it is important for experts to keep in mind the cognitive and social mechanisms that determine the spread of misinformation in all contexts, even those not related to AI. Finally, the identification of patterns of speed in the detection of fake content to enhance automatic moderation systems.

4.3.2. Focus groups and case studies

Focus group number one: primary-level teachers.

The case study in the first focus group addressed is as follows:
You are a teacher of a class of children between 8 and 10 years old.
Just before a break from lessons for the holidays, you set your students an assignment. The assignment asked your students to write a research paper. The essay was about a historical period.
When the students returned from holiday, they give you their homework. Correcting the homework, you realise that 80 % delivered an essay with the same repeating mistakes. The essays are very well written, but contain precisely the same inaccuracies.
The remaining 20 %, on the other hand, submit different homework, some written better than others, but with no errors from a historical perspective.

1. Is there any evidence that could make you think that the tasks containing errors were processed with the help of AI (e.g. ChatGPT)?
2. How do you decide to address this issue? Would you discuss AI directly with the students?
3. Considering that your students use AI tools, on which aspects would you focus your work with the class in the long term?

Belgium

8 participants, 7 teachers with different years of experience (33 - 3).
Regarding the impact of AI in society and in the educational context, the teachers interviewed noted the rapid adoption of AI tools (especially among young people), the reduced critical thinking skills ("students search for answers directly via AI instead of thinking for themselves"), increasing naivety when assessing online information, and difficulty distinguishing between real and false information.
In relation, on the other hand, to AI's positive/negative impact on a child growth and education, teachers proposed, as positive applications, learning support (use of AI for speaking engagements, help with homework when parents are not available, support for students who have less help at home), while, as negative aspects, the "abuse in education" (taking photos of exercise and using AI for solutions, lack of understanding of the subject matter, wrong learning strategies due to AI-generated explanations, distorted picture of student performance).
Speaking of misinformation and disinformation, according to the teachers who participated in the focus groups, many children accept "TikTok videos" as truth, or consider manipulated videos (e.g. of politicians) as real. According to them, students also have difficulties recognising edited content.
As strategies to counter misinformation, teachers framed the following: online search strategies; consult multiple sources for verification apart from output AI; using scientific sources; approach Wikipedia critically; open conversation about AI use in the classroom.

Finally, in relation to the case study brought to their attention, the teachers asked themselves these two questions: "Use GPT detectors?"; "Use plagiarism detectors?".
As final indications, then, they noted the importance of making the assignments in class with and without AI, and letting students check their outcomes in other sources than AI.
In conclusion, the teachers discussed today's challenges for teachers regarding misinformation and artificial intelligence, identifying the following critical points:
1. finding a balance between traditional and AI-assisted education;
2. determine when AI use is or is not allowed;
3. recognizing and preventing plagiarism.
4. keeping up with technological development.

Bulgaria

9 participants, 7 teachers.
The teachers interviewed teach a variety of subjects and have between 13 and 5 years' experience.

For the teachers interviewed, AI is inappropriate for young learners as it can undermine basic skills such as reading and writing, but it could have a positive effect on teachers by making it easier to create materials. In addition, respondents see potential in artificial intelligence for enriching learning through visualizations and data analysis and to develop analytical skills if used appropriately, but also express concerns that students may rely too much on AI.
Some argue that in the primary stage traditional methods are more effective and that AI is valuable for administrative tasks, but in the other hand they think that AI should be used sparingly and is inappropriate for early years pupils.
For the focus group participants, there are AI tools that are useful both in teaching and in everyday life. Useful tools are, for example, "ChatGPT" for text generation or "KAHOOT" for interactive quizzes. Excessive use, however, according to the interviewees can reduce basic skills such as writing and analysis. Teachers are concerned that AI may reduce students' cognitive skills. The factor of incorrect answers by AI, according to the interviewees, is also something to watch out for.
Regarding online information, for the teachers interviewed, primary students cannot distinguish true content from misleading information and tend to assume that everything they see online is true. Primary school children are more vulnerable, easily manipulated and less aware of the risks.

In relation to the case study, the teachers in Bulgaria saw the case analysed not only as a problem, but also as a learning opportunity. The incident, according to the teachers, could be used to discuss the credibility of information and the ability to recognise errors and misinformation. Recommended actions include discussion of the incident and group work; the use of the case study to develop critical thinking and revision techniques; and the possibility of creating assignments requiring working with multiple sources.
Key differences concern that elementary teachers emphasise basic skills and a controlled environment, and they see potential for more complex tasks that require analytical thinking and comparison of sources.

Ireland

6 participants, 5 teachers with several years of experience (25+ - 10).
Regarding the advent of AI in society, Irish teachers highlighted the following points: 1. While there is a wider mistrust of AI in education, this is rooted in a lack of professional learning for teachers and policies for schools. However, as AI is quickly and ever-evolving, teachers will have to be proactive in regularly upskilling themselves; 2. AI provides potential advantages for both children and their teachers. While its use is still limited, new ways of applying it to education of

it are being discussed more than before; 3. Children to be taught how to effectively and ethnically use AI; that it not just a tool to help do schoolwork faster; 4. In order to recognise misinformation/disinformation, children need to be better taught how to develop critical thinking skills, especially younger children.
About misinformation created with artificial intelligence, it appears that, according to some teachers, children do not ask as many questions about the veracity or otherwise of a piece of content as they do about entertainment.
Regarding teacher training, on the other hand, the speed with which AI develops and the difficulties in keeping up to date on the subject are also highlighted. It is highlighted that younger teachers are more motivated to learn and know how AI works.

The responses to the case study outlined key points: the teachers were aware that children are still learning how to use AI effectively and ethically and were all reluctant to apply sanctions but rather use such an event as a teachable moment.
The teachers, empathizing with the case, responded that they would talk to the students about the task, leaving them free to explain, if necessary, how much they engaged with the work themselves. Furthermore, they would show the students how AI often does not give the best possible answers and that the work they create themselves can be much more individual and probably better than the AI response.

Italy

10 participants, 6 teachers.
Teachers with several years of experience (34 to 10), dealing with different subjects. Two participants play the role of "digital animators".
For the teachers interviewed, the advent of artificial intelligence in today's society and educational environment is not particularly evident in primary schools. Only children in the last classes use artificial intelligence (especially generative AI systems such as ChatGPT).
The discussion then shifted to children's use of social networks as a means of information. The focus group revealed that some children seem to trust what they find on social networks: "for example, if an influencer says something, they believe it".
Another element that emerged concerns the difficulty in distinguishing the true from the false. Children are fascinated by what is on TV, social media or YouTube: what they see for them is reality. This becomes much more complicated with the advent of artificial intelligence.
Therefore, the importance of being able to support and train children in the use of technology emerged, but even before that, to provide adequate training for teachers.

Finally, as far as the case study is concerned, the teachers found that if the children use the same question (prompt), it is easy to understand that they used generative AI and that they all made the same kind of mistake. Other clues to recognise the use of artificial intelligence may be the repetitiveness of vocabulary and the lack of personalization in the text. Another element could be that the child then fails to explain the text produced.

As a reaction to the incident, the teachers interviewed promote dialogue in the classroom, also addressing the issues of ethics and fairness in action. In order to be able to comprehensively address the topic in the classroom, however, the teachers consider it essential to first train them on the subject, also in order to be able to provide the correct answers to the students. A final or aspect that they would address together with the class, even in the long run, concerns the correct search for sources, exploiting their curiosity and desire to learn.

Malta

10 participants, 8 teachers ranged from 8 to 22 years.

Participants discussed the significant impact AI has had on both society and education, noting that schools do not exist in isolation from technological advancements. One participant remarked, "Once society adopts AI, it will inevitably be adopted in schools".

Participants had mixed views on AI's role in education. Some felt it makes tasks easier for both teachers and students. "It makes the work of the teacher much easier," one participant said, while another added, "It's easier for students to research and easier for teachers to plan." However, others pointed out the challenges it brings, particularly in critical thinking and assessment. One participant noted, "It's harder for students to critically think about the answers and come up with a solution," while another highlighted the difficulty teachers face in identifying AI-generated student work.

Several participants acknowledged that educators are still adapting to AI's rapid development.

Participants felt that the impact of AI tools on a child's growth and education depends largely on how they are used.

There was agreement that AI can have both positive and negative effects. Concerns were raised about its potential to encourage over-reliance and hinder critical thinking.

A discussion emerged about the broader implications of AI on children's cognitive and social development. Some participants highlighted concerns about prolonged screen time and AI shaping students' thinking. One participant referenced studies indicating that "longer exposure to screens, tablets, and AI-powered devices, especially at a young age, is having an impact on cognitive development".

Several participants stressed the need for responsible AI education.
Participants noted that younger students, particularly those in primary school, often struggle to identify misinformation and tend to rely on online sources without questioning their credibility. There was general agreement that social media plays a significant role in shaping students' perceptions of information.
A specific example mentioned was a fake headline generator that allows users to create false news stories. Participants felt that while such tools can serve educational purposes, they also demonstrate how easily misinformation can be fabricated and spread. This discussion underscored the need for stronger media literacy education, ensuring that students develop the ability to question sources and cross-check information rather than assuming that all content, whether from social media or AI, is accurate.
Participants discussed AI tools that can help counteract disinformation, with some sharing their experiences of using different verification methods. One participant mentioned the practice of cross-referencing responses from multiple AI platforms, saying, "For example, cross-reference between ChatGPT and Copilot".
When asked if this approach was effective, the participant confirmed that it helped verify information, explaining, "Both platforms don't hallucinate in the same way, so the information, although different in detail, kept the same overall narrative." This suggests that while AI-generated responses may vary, comparing outputs from different tools can help assess consistency and reduce the risk of misinformation.
The discussion highlighted that while AI can sometimes contribute to misinformation, it can also be used strategically as a fact-checking tool when combined with critical thinking and verification from multiple sources.

Finally, in relation to the case study participants identified several indicators that AI may have been used in completing the assignments. The most telling sign would be the identical mistakes found in 80% of the essays, as well as the lack of grammatical or other writing mistakes. One participant suggested that the best way to confirm this would be to ask students directly how they completed the work: "I would try to verify this with them one-to-one, asking for their opinion and explaining the facts they wrote." This approach would help determine whether they relied on AI or conducted their own research.
Another participant pointed out that AI-generated content often follows recognizable patterns, which could explain the similarities in phrasing and structure across multiple submissions. The fact that a smaller portion of the class produced unique responses without errors further suggests that the rest may have depended on AI without critically assessing the accuracy of the information.

About the strategy to be adopted in the classroom in the short and long term, participants agreed that AI use should not be discouraged outright but should instead be openly discussed with students. One approach suggested was shifting the focus from written assignments to verbal discussions, allowing teachers to assess student understanding beyond what was submitted. "Instead of asking for typed work like a document, ask them to discuss the topic".

Some teachers actively ask students to use large language models and then assess the prompts rather than the final output by itself. Instead of evaluating only the written work, these teachers focus on how students structure their questions and interact with AI to refine their answers. This method helps ensure that students are engaging with the tool thoughtfully rather than simply copying and pasting text.

Teachers emphasized that teaching responsible AI use should be a long-term priority. The focus should be on helping students develop critical thinking skills, ensuring that they engage with AI-generated content thoughtfully rather than passively accepting it.

Spain

7 participants, 5 teachers with several years of experience (23-3).

For the Spanish primary school teachers interviewed, AI has generally made life easier. In primary education facilitates administrative tasks: course programming, scheduling, sequencing, statistics…

Other considerations include the fact that, according to the participants, AI foments immediate gratification – a worrying aspect of primary education – children have little patience or ability to cope with frustration. Additionally, "AI makes children want instant gratification. It makes things faster, but quality is lost in the process".

Regarding the fact that students may use AI to copy or not do the required homework and activities, some Spanish teachers believe that "children have always copied, whether it be from Wikipedia, Encarta, the family encyclopedia or a book from the library" and that, for this reason, it is important to focus on the positive aspects and risks that artificial intelligence offers, always bearing in mind the fact that AI often has many mistakes.

It also emerges from the focus group that AI is a complement to other aspects of children's education. It can help improve attention span, memory, critical thinking and other skills, but it cannot enhance or take the place of physical activity, nature, sociability.

For some participants, students tend to think that all they see and hear on a screen (is "true" and that teachers are completely "out of touch". In addition, children are completely manipulated by fake images and stories – how to act, how to look, how to be and this generates permanent frustration.

Family education, work on scepticism and critical thinking are deemed necessary.

About the case study submitted to the Spanish teachers, they believe that children use the first thing they find and therefore it is important to teach them how to search for more information, compile bibliographies and question information. The teachers think it is necessary to have the children read and verify the information, even finding errors, then asking them how they found the information, questioning what they found also in groups and creating something new.

Türkiye

6 participants, 6 teachers with different years of experience (20 - 4).
The discussion highlighted AI's transformative impact on education, emphasizing its potential to personalize learning, motivate students, and support progress in areas like language acquisition. Participants acknowledged its advantages, such as faster feedback and tailored educational programs, while also recognizing challenges like exacerbating inequalities and discouraging low-achieving students. Generative AI tools like ChatGPT were noted for their practicality in providing quick information, despite limitations in originality. Overall, the conversation underscored AI's dual role as a tool for innovation and a source of new challenges in education.

It is noted that concerns about children's prolonged use of conversational AI tools, emphasizing the need for monitoring such interactions. Participants noted that AI-generated emotions and thoughts directed by children could distort their sense of reality. While tools like "ChatGPT", "Quillbot", and "Socratic" are recognised for consolidating educational tasks and saving time, their practical application in public education systems remains limited. There is concern that students who misuse such tools may hinder the development of critical thinking and inquiry skills. Additionally, photo-editing applications were flagged for their potential negative impact on children, and chatbot responses were noted as sometimes inappropriate, underscoring the importance of oversight and responsible usage.
Participants highlighted how fake news and speculation undermine students' trust in the world. Younger children tend to believe information unconditionally, though this tendency decreases with age as critical thinking develops. To address this, activities and projects on disinformation are being implemented in classrooms. However, students may still act based on misinformation they believe, with AI-generated content like videos often being perceived more as humorous than misleading.
About artificial intelligence tools that contribute to counter disinformation, the conversation focused on raising awareness of platforms like "Teyit" and "Doğruluk Payı" (Turkish Platforms), as well as fact-checking programs like "Full Fact", "ClaimBuster", and "Factmata". Although

several participants acknowledged the existence of these tools, they stated that they had not personally felt the need to use them. Others mentioned the possibility of creating systems that validate news. A few individuals said they knew very little about the subject.
About the case study, regarding the case study, regarding the possibility of recognizing an AI product, participants had not encountered many AI-related errors, they observed issues in emotionally nuanced topics. Concerns were raised about AI contributing to misinformation, especially through deepfake technology. Some noted that students often copy-paste content without understanding it, highlighting the need for controlled use of AI tools. Others remarked they might initially suspect students of copying from peers rather than using AI but would still advise cautious and responsible AI use if identified.
For the handling of the issue with the class, the participants responded that they would discuss the functional use of AI tools, particularly in emotion-related tasks, and encourage students to rely on credible sources for information. Examples like social media algorithms—how preferences are tracked and influence content recommendations—were suggested as relatable ways to explain AI's impact. All participants agreed on the responsibility of teachers to educate students about AI and its implications.
On the other hand, regarding the management of the issue in the long term, participants emphasised creating a more controlled working environment in class, where students could focus on producing work with AI-supported controls and feedback to improve their outputs. They suggested teaching students how to use AI as a tool to assist with assignments and tasks, highlighting that it could make life easier when used correctly. While AI can help with research and information gathering, it is crucial for students to verify the information. In language learning, students should be encouraged to use AI tools that align with their individual learning styles. Participants also stressed promoting design-focused approaches and acknowledged that while AI can be useful for basic or quick information, they personally do not recommend or use it extensively.

Focus group number two: second-level teachers.
The case study in the first focus group addressed is as follows:
You are a teacher of a class of students aged between 16 and 18.
In your country, the electoral campaign for the upcoming general elections is underway. During your lesson, talking about current affairs, one of your students intervenes criticising a candidate and claiming to have seen a video in which the candidate declares war on Italy. Class discussion begins: many students support the point of view of the student who intervened, others remain silent, one intervenes saying that the video is a fake.
1) How do you decide to handle the debate?

2) What tools do you advise your students to use to recognise true from false information?
3) Given the situation, what are the aspects you think you should intervene on in the long term with the class?

Belgium

4 participants, 3 teachers with 8 and 18 years of experience in teaching.
Participants generally understand AI as a powerful tool that significantly influences both society and education. AI tools can offer personalized learning experiences, adapting to the unique needs of each student. However, there is a concern that over-reliance on AI might reduce opportunities for developing critical thinking and problem-solving skills. As one noted, "students often use AI like Google or ChatGPT but lack the skills to verify the reliability of the sources". Another teacher emphasised, "students use AI to find information but do not always understand what they are saying". They agree that the impact of society is just as big as the impact on education. The ecological impact of AI seems to be a big concern for the participants.
For participants, in addition, AI tools can have both positive and negative impacts on a child's growth and education. On the positive side, advanced AI tools can enhance learning by selecting reliable sources and providing critical insights and AI can be a valuable educational resource when used correctly. However, there are concerns about students using advanced vocabulary without understanding it, which can hinder genuine learning. Additionally, the use of free, less reliable AI tools can lead to misinformation and another concern relates that AI may hinder creativity within students who tend to follow and AI tool instead of their own ideas. Teachers involved feel that students have become more insecure about their own competencies since the rise of AI.
Participants think that students are constantly exposed to misinformation, and they need to be taught to double-check information. The teachers, however, do explain that fake news is not a new phenomenon, but they agree that fake news has become more prevalent since AI.
The participants note that they do not know any AI tools that help counteract disinformation. However, they tend to use fact-checking websites that provide reliable sources and critical insights. They believe that AI could help counteract misinformation, but those tools need to be used appropriately.

In response to the case study all teachers agree on the importance of critical thinking and the need to teach students how to verify information. They emphasise the role of educators in guiding students to use reliable sources and critically evaluate the content they encounter. They agree on the need for reliable sources and fact-checking tools to help students discern true from false information. In the long term, they highlight the importance of developing students' self-regulation

skills and their ability to independently verify information, ensuring they are well-equipped to navigate the complexities of the digital age.

All teachers agree on the importance of encouraging open discussion and critical thinking. One teacher would encourage an open discussion, allowing students to express their views while guiding them to question the validity of the information; other teacher would facilitate a structured debate, encouraging students to present evidence for their claims and critically analyse the video in question and the last one would encourage students to critically evaluate the video and discuss the potential for misinformation.

About tools and strategies to advise students to recognise false information from true information all teachers emphasise the use of reliable sources and fact-checking tools. Two teachers also stress the need to analyse the information and see if the spread of misinformation benefits someone and why one would spread that information, while the other teacher would try to "dissect" the video in class with a scientifical approach.

In conclusion, teachers agree on the importance of developing critical thinking skills and the ability to verify information.

Bulgaria

9 participants, 7 teachers of different subjects and with several years of experience (28 - 1).

For the part dedicated to exploring participants understanding and knowledge of AI and disinformation/misinformation, Bulgarian teachers expressed different points of view: AI transforms the role of the teacher into a mentor; AI enriches language learning, but students often take the information at face value;): AI is a powerful tool for structuring information, but requires contextualization; AI is useful for visualizations but risks reducing motivation for independent work.

Regarding the positive and negative uses of artificial intelligence, it was found that platforms such as "ChatGPT" or "KAHOOT" can be useful in creating materials and engaging students. The risks, on the other hand, concern hallucinations, misinformation, and the reduction of students' analytical and creative skills.

Teachers' answers also reveal a concern that for students all information created by AI may be credible, without considering it necessary to verify answers and content.

About countering disinformation, AI can be useful if used for fact-checking, data analysis, and teaching students how to recognise fake news (for example text and image verification tools).

In response to the focus group, four common positions emerged: 1) Debate and teamwork: divide students into groups to present opposing viewpoints and arguments; 2) Creating fake news: learning tasks that aim to create fake news in order to develop recognition and verification skills;

3) Using AI with frames: all teachers emphasise the need to frame the use of AI so as to avoid its uncritical acceptance; 4) Checking sources: encourage students to seek information from a variety of reliable sources.
Among the proposed strategies is the importance of creating a classroom debate to verify information and seek alternative resources, working in groups. A history teacher also suggests creating fake news to discuss it together, also creating alternative scenarios in history to develop critical thinking. Other teachers also advocate the importance of presenting the results of misinformation or misinformation created by AI to the children so that they can discuss it together.

Ireland

6 participants, 5 teachers with between 25 and 9 years of teaching experience.
The themes that emerged from the focus group with Irish teachers can initially be summarized as follows: 1. The teachers felt that the students are more informed about AI tools then they are. Number of years teaching did not impact upon this; 2. There was no clear understanding of any AI tools that contribute to countering disinformation; 3. Concerns around AI in education extended beyond academic achievement and were rooted in its potential wider societal impact; 4. Being able to recognise when algorithms are "pushing" content is something that students need to understand, especially in light of global political trends; 5. The teachers agreed that they have a role in encouraging their students to develop critical thinking skills and developing their sense of community.
The teachers who participated in the focus group also reflected on some aspects concerning society more broadly reflecting on the awareness or otherwise of how algorithms and social networks work, including the negative consequences and the spread of disinformation. The difficulty for young people, but not only for them, of being able to recognise the false from the true also emerged, and the consequences of the functioning of algorithms for the content proposed to users were also discussed.

In response to the case study brought to their attention, the Irish teachers found that using the moment in question as a teaching moment is perfect because you can then bring up the topic of respectful debate and teaching the students the difference between arguments.
For another teacher artificial intelligence is a tool, which one must therefore learn to know and deal with.
The importance of considering everyone's responsibilities also emerged, including in the way adults relate to artificial intelligence and the information created by it. Some teachers felt it was

important to emphasise that there is still a "sense of society" today: "there is still an understanding that communities matter, clubs matter, societies matter. There is still trust in the teaching profession, despite what some elements of the media would wish to portray".

Italy

8 participants, 4 teachers with different teaching paths and years of experience.
For the participants, artificial intelligence has had a huge impact on society and questions were raised as to how aware people are of it. Moreover, the impact is regulated in different cultural contexts. In schools, too, AI has had a very strong impact and many teaching colleagues express fear of the tool. Training on the subject is undoubtedly considered important. Teachers need to be accompanied in AI training and schools often fail to meet these needs. AI saves a lot of time when it comes to repetitive tasks, suggesting solutions and methodologies.
One concern that emerged was that AI is used first by the children and then by the teachers, who may not be ready. Critical thinking is also considered crucial. It is the school's task, in fact, to train critical thinking.
One teacher recounted discovering an assignment made with AI because she did not recognise the student's style, finding the text very "cold", as if "a robot had made it".
Finally, about misinformation, the importance of verifying sources and consulting other reliable sources to compare information was emphasised.

In response to the submitted case, teachers propose an activity to search for primary sources, also in other countries, showing the use of reliable sources. Another teacher promotes confrontation, regardless of the truthfulness of the content: "we are interested in developing civil dialectical skills in school".
Analysing sources is therefore essential: either by using a plurality of resources, consulting official organs and sources, or by cross-referencing sources to check substantiation.
In the long run, on the other hand, it is important for teachers to develop digital skills on the part of both teachers and students.

Malta

11 participants, 9 teachers working as teachers for between 10 and 30 years.
Participants noted that AI has been making headlines and is influencing various aspects of society. Some felt that it is changing.
There were concerns among some participants that students are using AI to make their work easier, with one stating. This led to discussions about the need for teachers to stay updated in

their professional practice and become more knowledgeable about AI.
Some also highlighted the role that digital literacy teachers should take in ensuring that other teachers are informed about new developments and supported in understanding how AI can be used effectively in education.
Participants shared mixed views on the impact of AI tools on a child's growth and education. Some felt that large language models, such as Copilot, could be used both positively and negatively, and are not inherently positive or negative. One participant noted, "It's not the applications themselves, but it is the role of the teacher to balance the use".
Participants discussed also the need for teachers to adapt to this new reality and reconsider how assessments are designed.
About disinformation and misinformation, participants expressed concerns that students are not always fully aware of misinformation, disinformation, and fake news. Some felt that students tend to accept information from large language models at face value without questioning its accuracy. There was also discussion about the difficulty in distinguishing between real and manipulated content, especially with AI-generated images.
Some participants felt that large language models contribute to the spread of disinformation. They noted that if someone relies on a chatbot for finding information, it may not always be accurate. One participant pointed out, "It's not the right tool; it can hallucinate".

Some participants mentioned that certain AI tools can help identify fake images, providing a way to counteract disinformation. They noted that "a good chatbot can identify fake images", which could be useful for verifying content.
Additionally, there was discussion about the importance of developing students' ability to critically assess information.

The responses to the case study outlined the importance of verifying sources and encouraging students to question the reliability of the information they encounter, and participants discussed various strategies for handling a classroom debate on misinformation, particularly in the context of an election campaign.
Participants also felt that teachers should guide students through a structured approach to evaluating claims, encouraging them to cross-check information rather than immediately accepting or dismissing it. Bias was another key concern, with participants pointing out that "some social media channels quote tabloids that are not always verified.
In terms of tools, some mentioned fact-checking websites like "Snopes" and other hoax verification platforms. Others brought up reverse image search tools such as "Google Reverse Image Search", which can help determine whether an image has been manipulated or taken out of context. AI-generated content detectors were also mentioned as a way to spot deepfakes.

Participants stressed that ideally, students should always compare information across multiple sources before making up their minds.
Beyond addressing the immediate discussion, participants felt that teachers should take a long-term approach to developing students' media literacy and critical thinking skills. Many agreed that teaching students how to assess information critically, rather than just giving them answers, is essential. Participants mentioned the importance of integrating AI literacy into the curriculum, so students understand both the risks and benefits of AI-generated content.
The discussion, finally, reinforced the idea that students need to develop the habit of questioning and verifying information, rather than assuming that everything they see online is true.

Spain

8 participants, 6 teachers with experience from 33 years to one year.
For Spanish teachers, artificial intelligence has also had an important impact on the life of society. For example, some claim that "students now use ChatGPT instead of google when they search for things". For others, there is no knowledge of the impact of AI, it can be used to simplify tasks, like academic work ("In schools it is very unknown still").
For the teachers who participated in the focus group, AI also helps them save time and learn new things.
During the focus group, it emerged that even according to Spanish teachers, many students believe everything they read and see: they do not read newspapers and believe everything they find on social media. The importance of teaching and training critical thinking also emerged. Teachers also point out that "fake news is everywhere" ("Wikipedia can be changed").

The responses to the case study focused on the possibility of counteracting information, using media deemed reliable, verifying the date of publication, understanding how algorithms work and what is behind them, inviting political experts to talk about control, and leaving room for classroom debate.

Türkiye

6 participants, 5 teachers with between 23 and 2 years of teaching experience.
concerning the impact of artificial intelligence in society, it was noted in this focus group that when technology is used correctly, it can be highly productive and beneficial, but in countries where it is misused, it can become an uncontrollable force. Proper application of technology was seen as useful, making certain tasks easier, but it also led to a tendency toward shortcuts and laziness. On the positive side, it allowed individuals to learn at their own pace and according to

their preferences, which was seen as a significant benefit. However, students who were not motivated to engage in educational activities could have their tasks done for them, which ultimately hindered their progress. Overall, while technology has made basic information more accessible, it has both positive and negative impacts.

Speaking of disinformation, however, the discussion highlighted that many students tend to believe information that aligns with what they want to be true, even if they are aware of its inaccuracies. There is a lack of sufficient knowledge, with some students showing little concern about verifying information and believing everything they encounter. A notable issue is the very low level of trust, although some students are aware of this and the need for critical evaluation. The discussion pointed out that social media platforms like Instagram, Facebook, and TikTok, which are widely used, often feature AI algorithms that promote false information and harmful content, and tools like ChatGPT don't always provide accurate results.

In relation to the case study, regarding the management of classroom debate, teachers focused on ensuring access to verified information, with participants emphasizing the importance of consulting experts or official sources like government or university-authorized bodies. If expert advice is not available, they would encourage students to research reliable internet sources collectively in class. They also stressed the value of using multiple sources, especially official ones, to ensure accuracy. One participant proposed turning this process into a learning opportunity by organizing a workshop where students could research a historical topic, debate different perspectives, and evaluate the credibility of sources, aiming to foster critical thinking and epistemological methods.

About tools and strategies to advise students to recognise false information from true information, the conversation emphasised the importance of recommending verified, official sources from governments and universities over social media platforms for accurate information. Participants advised students to evaluate the credibility of sources, suggesting that they use reliable websites, expert articles, and reverse image searches for verification. They also highlighted the need for cross-checking information from multiple sources and being critical of inconsistencies or illogical content. Overall, the focus was on helping students develop skills for identifying reliable information and avoiding misinformation by relying on trustworthy and authoritative sources.

Finally, participants reflected on the use of AI as a tool to enhance creativity, particularly in areas like language learning, where AI can provide tailored and practical support.

Participants emphasised also the need for students to be critical and sceptical, especially in the post-truth era. They advised fostering an epistemological approach where students question information, cross-check sources, and develop skills in scientific literacy. Encouraging students to seek guidance from experts and use multiple sources is seen as key in avoiding misinformation.

4.3.3 Answers and literature

Analysing the answers obtained from experts and teachers, an initial difficulty in defining artificial intelligence emerges.

Already with the earliest studies on the subject (think back to 1950, when Alan M. Turing questioned whether machines could think), the definition of artificial intelligence was not simple. Today, in addition to the difficulty in defining artificial intelligence, there is the added complexity of circumscribing the definition to systems and tools.

Experts and teachers have replied to questions that now concern the increasingly close relationship between human beings and technology, so much so that it has merged into a concept defined as "onlife".

At a time when human beings find themselves living in the "information society", disinformation is an extremely complex phenomenon that also concerns schools and the world of education.

Speaking of disinformation, the interviewees highlighted the role of social networks at various times. Increasingly sophisticated algorithms can influence the actions and sometimes the thoughts of users. In the literature study, reference was made to "filter bubbles", and many teachers were concerned about the idea that their students might believe everything they find on the Internet to be true.

The problem is that, often, what they find online is part of the "bubble" of their beliefs.

In several responses, a dualism emerged between the responses towards artificial intelligence: those who are sceptical (and afraid) and those who promote its freer use. The teachers also acknowledged at various times that artificial intelligence could help (especially in terms of time) their work.

Critical thinking and the ethical use of these tools, however, are always emphasised, regardless of the approach towards AI.

Even in the recent European regulation on artificial intelligence, as has been reported, human supervision plays a key role.

Teachers and experts, then, highlighted the opacity of the interests behind disinformation and the possibility of fake news being used for dangerous or violent purposes. These characteristics emerged in the definitions provided of the various types of misinformation that are known today and it is clear that, in a reality also characterised by "information overload", these aspects must gain the attention of teachers and educators.

The deepfake and hallucinations of artificial intelligence further complicate the situation. The teachers and experts interviewed, in fact, often referred to these two issues.

In the study of the literature, then, the topic of the digital divide emerged. In order to manage

these issues and to be able to exploit the advantages and potential of artificial intelligence, it is necessary to possess the appropriate tools and digital skills. In the absence of these, it is inevitable that some people may be left behind and may, also, find it more difficult to recognise true information from false and manipulated information, even through artificial intelligence.

4.4 Conclusions and Considerations

The questionnaires and focus groups revealed an important awareness of artificial intelligence. Both in terms of the benefits and the problems and risks.

Artificial intelligence is a rapidly evolving technology and, as was discussed, presents various challenges and problems in the world of education and, even before that, in the training of young minds.

Artificial intelligence affects – with important cultural, geographical and socio-economic differences – people's lives.

The teachers who participated in the focus groups showed curiosity about this technology and also awareness of the fact that AI can help in performing repetitive and schematic tasks. It was also noted that AI can be a valuable aid in creativity and the development of new and useful forms of expression.

On the other hand, concerns have been raised on several occasions about artificial intelligence: a tool that can lead to a reduction in the skills of students and teachers, that can "lag" in the demand to perform more or less complex tasks, that can present hallucinations or incorrect results. Artificial intelligence can also be used to create disinformation and misinformation, making an already dangerous phenomenon even more complex and insidious, and one that has already been enhanced by the advent and growth of social media.

As far as schools and education are concerned, the people who participated in the activities of this report often mentioned "ChatGPT", emphasising how the advent of generative AI has had a major impact in schools and education, in many cases taking over the teaching and creative activities of teachers and learners.

Almost all the experts interviewed and teachers participating in the focus groups emphasised the importance of critical thinking and reiterated how essential human intervention still is.

While recognising and emphasising the positive aspects and strengths of artificial intelligence, what emerges is the importance of being able to help and support students (children and young people) in developing and building critical thinking. Critical thinking that is needed both to be able to use technology consciously and positively and to be able to manage the continuous flow of

information. Information that, as we have seen, can be false, distorted or manipulated.
In fact, the importance of accompanying human intervention in the use of artificial intelligence has been repeatedly promoted. Intervention that can verify whether or not the results produced by the AI are correct or relevant or even enhance the result obtained, personalising it and adapting it to the concrete context.
Critical thinking is also essential when it comes to disinformation and misinformation. Concerns emerged regarding the use of sources by young people. Many teachers believe that they only inform themselves through social networks and do not consult authoritative and verified sources. For this reason, the intervention of adults and teachers was proposed as a response to show young people the correct and safe way to inform themselves and to possibly question ideas and beliefs.
In conclusion, the importance of teacher training was emphasised. In order to manage a hyper-connected world inhabited by artificial intelligence, adults, and especially teachers, must be able to possess first-hand the skills, abilities and tools necessary to use artificial intelligence in a profitable, ethical, creative and useful way.Educators, therefore, must be able to manage, deal with and explain to students the risks and problems that AI can bring. In addition to training, finally, it was also proposed to increase concrete (e.g. governmental) guidelines regarding the use of AI in schools.
Teacher training, the development of critical thinking and the promotion of open dialogues are, therefore, key elements in the use of artificial intelligence today.

5 POTENTIAL RESPONDING STRATEGIES

5.1 State of the art in countering disinformation with AI

The unprecedented speed with which information travels in today's infosphere, coupled with the rise of AI-powered fake news, presents a scalability challenge for effectively detecting and eliminating misinformation and disinformation. Disinformation campaigns, often automated and highly sophisticated, demand equally automated countermeasures.

This has opened the way to automated, ML-based approaches to fighting fake news. They especially address the limitations of manual fact-checking processes, which are resource-intensive and incapable of keeping pace with the sheer volume of content generated online. AI tools streamline the labour-intensive process of fact-checking, making it feasible to verify information at scale.

All the ML techniques are currently adopted for the task (Montoro-Montarroso et al., 2023):

1. Supervised Learning: Models are trained on labelled datasets to classify content as truthful or disinformative. Popular algorithms include Decision Trees, Logistic Regression, Support Vector Machines, and Deep Learning models. Supervised learning is effective when high-quality, labelled data is available, but models trained on specific domains often fail to generalise across diverse contexts.
2. Unsupervised Learning: Used when labelled data is unavailable. Techniques like clustering and anomaly detection identify patterns that deviate from the norm, signalling potential disinformation.
3. Semi-Supervised Learning: Combines labelled and unlabelled data to improve model performance while reducing the need for extensive manual labelling. It's most effective in scenarios with limited labelled datasets.
4. Deep Learning (DL): Employs neural networks, such as Convolutional Neural Networks (CNNs) for image analysis and Transformers for text-based disinformation detection. DL models like BERT and GPT excel in extracting complex relationships in textual data.

Showing comparable performance to properly trained human detectors, these ML-based tools offer a clear advantage of efficiency in dealing with huge amounts of online content. They excel in identifying patterns (pattern recognition) within vast datasets, enabling them to detect disinformation based on linguistic, stylistic, or content-based features.

For a deeper analysis, key methodologies include syntactic analysis (which detects unnatural

sentence structures) and semantic analysis (which assesses coherence and logical flow). Automated classification models, meticulously trained on annotated datasets, exhibit increasing proficiency in distinguishing authentic material from fabricated information.

The adoption of ML is driven by its efficiency in processing large datasets. These datasets, often sourced from real-world examples, can be enhanced with synthetic cases to improve model robustness and diversity. Moreover, AI facilitates the extraction of defining features of disinformation —including lexical, syntactic, and stylistic markers— while enabling real-time social network analysis (SNA). SNA elucidates the complex dynamics of coordinated activities, identifying patterns of bot behavior, the propagation pathways of deceptive campaigns, and the relationships between key disseminators of false narratives (Montoro-Montarroso et al., 2023).

Key Concepts and Methods are:

1. Capture, Track, Respond (CTR) Framework:

a) Capture: Tools focus on identifying disinformation elements through text mining, keyword extraction, and data preprocessing.

b) Track: Monitoring changes in disinformation narratives over time, identifying emerging threats.

c) Respond: Providing actionable insights to design counter-messaging strategies.

2. Data Analysis Techniques:

a) TF-IDF (Term Frequency-Inverse Document Frequency): Used to weight the significance of terms across datasets, highlighting keywords in disinformation.

b) Keyness Analysis: Examines word frequency differences between target and reference corpora, revealing unique narrative elements.

c) Network Analysis: Maps relationships between terms, hashtags, or entities to uncover clusters and echo chambers of disinformation.

These systems leverage Natural Language Processing (NLP) to identify key indicators of disinformation, such as manipulative language, stylistic inconsistencies, and coordinated network behaviour. For example, NLP tools analyse text for emotional tone, sensationalist headlines, and contradictory claims, while SNA maps the propagation of disinformation within digital networks, identifying clusters of coordinated activity or inauthentic amplification. By modelling the behaviour of disinformation campaigns, AI systems can predict the trajectories of false narratives, enabling proactive interventions that curb the spread of harmful content before it reaches a broader audience. Techniques such as supervised learning, which relies on labelled datasets to train models, and unsupervised learning, which identifies anomalies in unlabelled data, are instrumental in these efforts.

By continuously enhancing the capabilities of these models, researchers aim to refine disinformation detection, enabling faster and more precise responses to emerging threats in the

digital landscape.
Case studies (as in Skumanich, Han Kyul, 2022) illustrate the effectiveness of AI/ML in detecting specific disinformation narratives, such as anti-Semitic language or conspiracy theories. Visualization tools like network graphs and timeline analysis enhance interpretability, making insights actionable for policymakers and researchers.
For instance, TF-IDF analysis identified dominant conspiracy theories and hate speech narratives. Keyness analysis provided insights into narrative shifts, such as reactions to significant events like mass shootings or political debates.
Against deepfakes, AI technologies deploy sophisticated image and video analysis techniques, focusing on anomalies such as textural inconsistencies, unnatural lighting conditions, or audio-visual desynchronization. Advanced detection frameworks leverage convolutional neural networks (CNNs) and other deep learning architectures that dissect content at a granular level, identifying artefacts imperceptible to the human eye. Additionally, these models are supported by iterative refinement, whereby feedback loops enhance their precision and adaptability to new forms of synthetic media.
Nonetheless, the evolving sophistication of generative models perpetuates an enduring arms race in this domain, highlighting significant limitations in current AI approaches (Gilbert & Gilbert, 2024). The transferability and robustness of AI systems across varying domains and contexts remain considerable challenges. Hybrid approaches that synergize automated analytical methods with human expertise demonstrate substantial promise, but even these integrated systems face inherent constraints. By augmenting the precision and expediency of fact-checking processes, these systems empower journalists and subject-matter experts with AI-enhanced tools for comprehensive claim verification and source validation, addressing only some of the limitations of purely automated mechanisms (Montoro-Montarroso et al., 2023).
However, ML models are often not able to analyse the information in a relevant context and thus should not be used without human supervision. On the ability of analysing information in an ever-effective way many studies point out the possible shortcomings and difficulties.
A second layer of complexity is represented by their adoption in the educational environment.

5.2 Challenges in countering disinformation with AI

Literature is concordant in saying that ML-based systems can exhibit false positive or false negative results, alongside correct outcomes.

On the one hand, AI can perpetuate biases, wrongly flagging legitimate speech as fake news. For instance, the leading sentiment analysis tools are trained on data coming from the US and Europe and designed without consideration for cultural differences in expressing emotions, leading to geo-political bias in disinformation detection (Bhadra et al. 2024). On the other hand, ML-based tools can also fail to detect disinformation, particularly when it requires a contextual and holistic analysis of multiple features, often in different data modalities, such as text and image. This makes AI systems perform poorly in the detection of non-evident fake news, for instance, those presented in the form of a meme (Yankoski et al. 2021). While the growing interest in multimodal foundation models provides hope for the future development of more resilient tools, fake news detection models operating simultaneously on different data modalities are currently in their infancy (Kumar and Taylor 2024). Moreover, ML models still struggle to effectively counter the so-called "Adversarial Content" (such as deepfakes). The speed of generation outpaces the development of detection systems.

Language Models (LLMs) are increasingly considered a potential tool for debunking misinformation. However, current evaluations suggest that they are not yet reliable for this purpose. One major issue is the tendency of LLMs to generate "hallucinations" —fabricated or misleading content— which research has shown to be mathematically unavoidable (Ziwei Xu, Sanjay Jain, Mohan Kankanhalli, 2024).

For example, recent audits indicate that leading AI chatbots frequently fail in their responses to misinformation prompts. An audit by NewsGuard, a leading disinformation-detecting platform, ("December 2024 — AI Misinformation Monitor of Leading AI Chatbots", available at link: https://www.newsguardtech.com/ai-monitor/december-2024-ai-misinformation-monitor/) revealed that the top 10 generative AI models collectively repeated false claims 40.33% of the time, failed to provide a response in 21.67% of cases, and successfully debunked misinformation only 38% of the time, leading to a total failure rate of 62%. This marked a significant drop in performance compared to previous months, with the decline likely attributed to the rapid rollout of new features and updates, which may have outpaced efforts to enhance safeguards against misinformation, suggesting that safety measures are lagging behind technological advancements.

The findings suggest that while LLMs have the potential to counteract misinformation, their current limitations —including systemic biases, hallucinations, and vulnerability to manipulation— undermine their effectiveness in debunking false narratives.

In their monthly report, launched in July 2024, NewsGuard evaluates AI chatbots by testing their responses to misinformation-related prompts. The structured approach assesses accuracy, reliability, and vulnerability to false claims.

NewsGuard's methodology provides a structured and repeatable framework for assessing AI chatbots. By using real-world misinformation, varied prompt types, and a clear scoring system, it highlights both strengths and weaknesses in generative AI. While AI has the potential to counteract misinformation, its current limitations pose a significant risk in the spread of falsehoods.

The methodology relies on "Misinformation Fingerprints," a proprietary database cataloguing false claims circulating online. Each audit selects 10-15 claims based on prominence in current news cycles, potential harm, and widespread dissemination.

To test AI models, three types of prompts are used. "Innocent User Prompts" are neutral questions that reflect typical user inquiries, such as asking about a widely circulated but false news story. "Leading Prompts" assume a false claim is true and request further details, testing whether the chatbot passively accepts misinformation. "Malign Actor Prompts" are crafted to deliberately manipulate the chatbot into generating disinformation, revealing whether the AI has safeguards against bad-faith use.

Each response is categorized into three possible outcomes. A "Debunk" means the AI correctly identifies the false claim and refutes it with evidence. A "Non-Response" occurs when the AI declines to answer or provides ambiguous responses, failing to correct misinformation. "Misinformation" happens when the chatbot repeats or reinforces the false claim, which is the most serious failure mode.

For data collection, each chatbot is tested with 300 prompts covering 10 false claims. The failure rate is calculated as the percentage of responses that either contain misinformation or fail to provide a corrective response. While individual chatbot results are not publicly disclosed, companies can request their specific performance data.

Furthermore, a specific examination of DeepSeek, a Chinese AI chatbot, found that it advanced government-aligned narratives in 60% of responses concerning Chinese, Russian, and Iranian disinformation. The chatbot performed particularly poorly, with an 83% fail rate, making it one of the least reliable AI tools tested. Like other chatbots, DeepSeek struggles to resist manipulation by misleading prompts, making it vulnerable to disinformation campaigns.

The poor results by the LLMs can be partially solved with more transparency in their training datasets. To combat these challenges, researchers emphasize:

1. Developing advanced detection models tailored to LLM-generated disinformation and deepfakes.
2. Promoting Explainable AI (XAI) to ensure that detection systems are interpretable and

trustworthy.

3. Enhancing public media literacy to recognize and question suspicious content.

By addressing these challenges, AI can be better aligned to counteract its misuse, ensuring it supports, rather than undermines, information integrity. The challenges of applying AI against disinformation highlight the complexity of the issue and the limitations of relying solely on technology. Over-reliance on AI could lead to algorithmic censorship, where genuine expressions are suppressed. Privacy issues arise from AI-based monitoring of online behaviour, raising concerns about surveillance and data misuse. Addressing these risks requires a balance between technological advancements and legal and ethical oversight. To be effective, AI systems must be transparent and enable effective human oversight, ensuring that their deployment serves both technical and societal needs. To enhance the effectiveness of AI in countering disinformation, research must focus on:

1. Developing Generalizable Models: Expanding the scope of training datasets to encompass diverse languages, cultures, and contexts will improve model adaptability and robustness.
2. Enhancing Real-Time Detection: Streamlining algorithms for faster analysis can enable proactive interventions before disinformation gains traction.
3. Fostering Collaboration: Multi-stakeholder partnerships between governments, tech companies, and academia are essential to create comprehensive countermeasures.
4. Establishing Regulatory Frameworks: Clear guidelines and standards for AI applications can ensure legal and ethical deployment and foster public trust (Gilbert & Gilbert, 2024). In this regard, Art. 4 of the AI Act mandates that educational institutions deploying AI systems should provide appropriate AI literacy training to the teaching staff. When appropriate, this should encompass training concerning the role of AI in spreading and countering disinformation. Currently, there seems to be a gap in the ethical guidelines for teachers concerning the responsible use of AI for disinformation education.

While the Commission's Ethical Guidelines on the use of AI and data in teaching and learning for Educators mention the need to promote digital skills such as management of information overload and recognising disinformation, they do not explore AI's role in this endeavour (European Commission 2022a, p. 9). Similarly, the Commission's Guidelines for teachers and educators on tackling disinformation and promoting digital literacy through education and training refer to technical aspects of disinformation, but do not provide an in-depth analysis of risks and benefits of using AI for teaching about disinformation (European Commission 2022b, p. 30).

AI represents a powerful ally in the fight against misinformation and disinformation, offering tools

for detection, analysis, and mitigation. By leveraging advancements in ML, NLP, and SNA, societies can counter the threats posed by digital falsehoods. However, a collaborative and ethically grounded approach is essential to ensure that these technologies serve the broader goals of truth, trust, and democratic integrity.

5.3 Strategies and future research: improve critical thinking

The problem of disinformation and misinformation is socio-technical in nature and therefore cannot be tackled solely through technical fixes.

The predominant focus on new challenges raised by AI in terms of fake news generation risks obscuring the human factor which remains fundamental in the successful spread of false information.

Critical thinking and media literacy are essential competencies in modern education. These skills enable students to critically evaluate information sources, discern truth from falsehood, and navigate the complexities of the digital world (Suffia, 2022). Integrating AI into educational frameworks can support these goals by fostering analytical skills and enhancing learners' ability to engage with content critically.

AI-powered tools, such as interactive fact-checking systems and automated credibility assessments, provide students with real-time feedback on the reliability of online information. An example is provided by the aforementioned NewsGuard which offers tools such as news reliability ratings based on impartial journalistic criteria or online tracing of false narratives. These tools not only assist in identifying falsehoods but also promote reflective learning by encouraging students to analyse the reasoning behind their judgments. AI can simulate scenarios where students must make decisions based on conflicting information, further strengthening their critical thinking abilities.

Incorporating AI literacy into education ensures that students are not passive recipients of technology but informed participants who can question and evaluate AI-generated outputs. Understanding the limitations and biases inherent in AI systems fosters a culture of scepticism and inquiry, vital for combating disinformation (Suffia, 2022).

Patterns of the behaviour of young people on social media, such as instantaneous sharing of news with the inner circle, growing engagement with influencers and the lack of appropriate media literacy are paramount in this regard (Valencia-Arias 2023). For example, a 2020 study involving high school students in Australia discovered that young people struggle to independently verify

the accuracy of social media posts or to note the possibility of manipulation of images and videos, often failing to recognise social and political biases (Johnston 2020). Studies of young people's approach to fake news in Spain and Poland have indicated that while students are aware of the phenomenon of disinformation on social media, they are not familiar with effective information verification methods (Perez-Escoda et al. 2021; Sitek 2024). Similarly, a study conducted in Finland found that children were able to discuss various intentions behind fake news generation but were less prepared to objectively evaluate the quality and consistency of evidence (Vartiainen et al. 2023).

Teaching media literacy should prioritise creating the competencies necessary for overcoming cognitive biases in the age of ML and offer practical tools to critically engage with news. This means that the curricula should be developed with less focus on following the disinformation checklists and more on the development of critical thinking which allows for independent evaluation of information (Johnston 2020). Moreover, the design of a media literacy curriculum should take into account country-specific attitudes of students' engagement with news, such as trust in traditional media or lack thereof (Selnes 2023).

Within the media literacy education programs, strategies can be deployed to harness the power of AI to address the growing spread of fake news. Tackling disinformation on a wide scale must necessarily involve human-machine cooperation. In particular, recent studies have confirmed that machines and humans tend to focus on different linguistic aspects of disinformation – while ML algorithms detect disinformation mainly through the association of names and expressions with news previously marked as false, humans tend to focus on more general features and emotional impact (Pawlicka et al. 2024). Thus, guided cooperation with explainable ML systems can boost human performance in fake news detection (Lai and Tan 2019). Therefore, AI systems have the potential to be used under teachers' supervision to educate students about the features of disinformation, enhancing their ability to spot suspicious content effectively. At the same time, media literacy education should explore the limitations of ML technologies, avoiding automation and confirmation bias.

Existing applications of AI in education and content analysis could be successfully leveraged to counter disinformation and support media literacy among students:

1. Adaptive Learning Platforms (ALPs): AI-driven adaptive learning platforms can personalize educational experiences, tailoring content to address students' individual needs and misconceptions. Popular ALPs such as Carnegie Learning, DreamBox Learning, Smart Sparrow or Knewton offer individualised learning paths in a variety of subjects and levels, leveraging the strengths and weaknesses of students to best address their learning needs (Dutta et al. 2024). By integrating modules on media literacy and critical evaluation of news, these platforms could reinforce key skills essential for identifying misinformation. For example, AI systems could analyse

students' interactions with digital content and provide tailored exercises that challenge their understanding of credibility and source evaluation. Such a personalised approach would ensure that students develop robust analytical skills over time.

2. Gamified Learning Experiences: Gamification, powered by AI, could transform the learning process into an engaging and interactive experience (Kurni et al. 2023). An example of such technology is provided by Seppo, a Finish learning platform which offers tools for turning existing course materials into interactive games (Goshevski et al. 2017). Games that simulate real-world scenarios involving misinformation, such as the spread of fake news or the identification of manipulated media, could effectively teach students about the dynamics of disinformation campaigns. These AI-driven simulations offer immediate feedback, enabling learners to see the consequences of their decisions in a controlled environment.

3. Real-Time Content Analysis Tools: AI tools capable of real-time content analysis, such as the Fake News Detection on Cloud system could be integrated into classroom settings to enhance learning outcomes (Cavus et al. 2024). These tools analyse text, images, and videos for signs of manipulation or falsehood, providing students with hands-on experience in detecting misinformation. By actively engaging with these applications, learners could develop practical skills that translate to their interactions with information outside the classroom.

4. Supporting Educators: AI also serves as a valuable resource for educators, helping them to optimise teaching methods through the use of AI-powered teacher dashboards (Molenaar et al. 2025, p. 269). These tools could be deployed to streamline the integration of media literacy into lesson plans, for instance by curating relevant teaching materials, assessing the credibility of sources, and generating illustrative examples of misinformation. These systems could enable teachers to focus on fostering critical thinking rather than spending excessive time on preparatory work.

5.4 Legal and ethical considerations and future directions

While AI offers promising solutions, its implementation in education must address several challenges. Ensuring the accessibility and affordability of AI tools is critical to prevent exacerbating existing inequalities in education. Additionally, transparency in AI algorithms is essential to build trust among educators and students. Educators must also be trained to use AI tools effectively and to interpret their outputs critically.

Legal and ethical considerations include safeguarding students' privacy and data security. The

deployment of AI in classrooms should prioritize the responsible use of technology, ensuring compliance with ethical standards and protecting students from potential harm.

To maximize the impact of AI in countering misinformation and disinformation in education, future efforts should focus on:

1. Developing Inclusive AI Tools: Ensuring that AI systems are designed to cater to diverse cultural and linguistic contexts, making them relevant and accessible to global learners.
2. Enhancing Teacher Training: Providing educators with comprehensive training on AI tools and media literacy to enable effective integration into teaching practices.
3. Encouraging Cross-Disciplinary Research: Promoting collaborations between educators, AI developers, and social scientists to create innovative solutions tailored to educational needs.
4. Implementing Policy Frameworks: Establishing guidelines for the legal, ethical and effective use of AI for countering disinformation in education settings, emphasizing data security and inclusivity.

AI holds immense potential to revolutionize education by equipping learners with the tools and skills needed to counteract misinformation and disinformation. Through the promotion of critical thinking, media literacy, and AI literacy, educational systems can empower students to navigate the complexities of the digital age with confidence and discernment. The integration of AI into education is not just a technological advancement but a critical step towards building resilient and informed societies (Suffia, 2022).

6 REPORT SUMMARY

This Report is part of the studies and publications of PAIDEIA project work package 2 and it is about how to use of AI in education in spreading disinformation/misinformation. A further objective of the study concerned evaluating the potential positive impact of AI in counteracting fake news.

The Report give a comparative evaluation of the impact of the use of AI in education in fighting misinformation and disinformation. This impact was analysed based on a state-of-the-art study of the phenomena and practical case-studies analysis. A further objective of the study was the elaboration of potential responding strategies.

It began with an analysis on artificial intelligence and with the elaboration of some definitions about the creation and dissemination of false, distorted or manipulated information, leading to the study of the correlation between the two themes.

Subsequently, the report focused on the opinion of experts in the field and the experience of teachers who face the challenges of using artificial intelligence and misinformation/disinformation daily. In order to explore these last two aspects in more detail, experts were interviewed and then the teachers answered questions in the context of two focus groups.

With the last section (number three), the use of artificial intelligence as a useful tool to counteract misinformation and disinformation was framed.

In the preparation of this report, several challenges emerged, and possible strategies were framed regarding the topics addressed.

With an awareness of how artificial intelligence is reshaping present and future society, the importance of keeping critical thinking alive and continuing to promote human intervention in activities involving the use of artificial intelligence was emphasised in several passages of the report.

Indeed, there are many concerns raised, challenges framed and, also, the advantages and strengths of AI.

Misinformation and misinformation are also complex phenomena that are increasingly influenced by the digital, now also by artificial intelligence.

In conclusion, this report has developed several reflections, observations, recommendations and strategies regarding the use of AI in education and countering misinformation and disinformation. These can be used in the development of the PAIDEIA project curriculum. Curricula that will have to adapt as far as possible to a complex, digitised society that has to deal with the challenges and strengths of AI.

7 BIBLIOGRAPHY

Wang, M., Sadeghi, M., Lin, C. (2025) "Chinese Chatbot Phenom is a Disinformation Machine", NewsGuard, https://www.newsguardtech.com/special-reports/deepseek-ai-chatbot-china-russia-iran-disinformation/

"DeepSeek Debuts with 83 Percent 'Fail Rate'" in NewsGuard's Chatbot Red Team Audit.

Bell, D. (1973). The coming of post-industrial society; a venture in social forecasting, New York, Basic Books.

Benanti, P. (2022). Human in the loop. Decisioni umane e intelligenze artificiali, Mondadori Università.

Bharathi Mohan, G., Prasanna Kumar, R., Vishal Krishh, P. et al. (2024). An analysis of large language models: their impact and potential applications, Knowl Inf Syst 66, 5047–5070, https://doi.org/10.1007/s10115-024-02120-8

Buitrago López, A., Pastor-Galindo, J., Ruipérez-Valiente, J. A.(2024). Frameworks, Modeling and Simulations of Misinformation and Disinformation: A Systematic Literature Review, ACM Comput. Surv. 1, 1 (June 2024), https://doi.org/10.48550/arXiv.2406.09343

Castells, M. (1996, 1997, 1998). The Rise of the Network Society, The Information Age: Economy, Society and Culture Vol. I.; The Power of Identity, The Information Age: Economy, Society and Culture Vol. II.; End of Millennium, The Information Age: Economy, Society and Culture Vol. III.; Cambridge, MA; Oxford, UK: Blackwell.

Cavus, N., Goksu, M. and Oktekin, B. (2024) 'Real-time fake news detection in online social networks: FANDC Cloud-based system', Scientific Reports, 14(1), p. 25954. Available at: https://doi.org/10.1038/s41598-024-76102-9

Commissione delle Comunità europee (1996). Libro verde vivere e lavorare nella società dell'informazione: priorità alla dimensione umana, Bruxelles, 24.07.1996 COM(96) 389 def.

Corazza, P. (2022). Filter bubbles e echo chambers: origini pre-digitali e elementi di novità. Riflessioni dalla prospettiva della media education. FORMAZIONE & INSEGNAMENTO, 20(1),

856-867 [10.7346/-fei-XX-01-22_76].

Crawford, K. (2021). Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence, Yale University Press.

Cugler de Moraes Silva, E., Carlos Vaz, J. (2024). What characteristics define disinformation and fake news? review of taxonomies and definitions, https://doi.org/10.48550/arXiv.2405.18339

Deepak, P., Bhadra, S., Jurek-Loughrey, A., Kumar, G., and Satish Kumar, M. (2024). Geo-Political Bias in Fake News Detection AI: The Case of Affect, AI and Ethics, 1–6. https://doi.org/10.1007/s43681-024-00494-7

Deibert, R. J. (1997). Parchment, Printing and Hypermedia. Communication and World Order Transformation, Columbia University Press.

Dutta, S. et al. (2024) 'Enhancing Educational Adaptability: A Review and Analysis of AI-Driven Adaptive Learning Platforms', in 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM). 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM), pp. 1–5. Available at: https://doi.org/10.1109/ICIPTM59628.2024.10563448

European Commission. Directorate General for Education, Youth, Sport and Culture. (2022a) Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators. LU: Publications Office. Available at: https://data.europa.eu/doi/10.2766/153756 (Accessed: 11 March 2025).

European Commission. Directorate General for Education, Youth, Sport and Culture. (2022b) Guidelines for teachers and educators on tackling disinformation and promoting digital literacy through education and training. LU: Publications Office. Available at: https://data.europa.eu/doi/10.2766/28248 (Accessed: 13 March 2025).

Floridi, L. (2020). Il verde e il blu. Idee ingenue per migliorare la politica, Raffaello Cortina.

Gilbert, C., and Mercy A. G. (2024). The Role of Artificial Intelligence (AI) in Combatting Deepfakes and Digital Misinformation, International Research Journal of Advanced Engineering and Science, Volume 9, Issue 4, pp. 170-181.

Goshevski, D., Veljanoska, J. and Hatziapostolou, T. (2017) 'A Review of Gamification Platforms for Higher Education', in Proceedings of the 8th Balkan Conference in Informatics. New York, NY, USA: Association for Computing Machinery (BCI '17), pp. 1–6. Available at: https://doi.org/10.1145/3136273.3136299

Greenwald, A.G., Hamilton Krieger, L. (2006) Implicit Bias: Scientific Foundations, «California Law Review», 94, 2006, n. 4, pp. 945-967.
https://newsguardrealitycheck.com/p/deepseek-debuts-with-83-percent-fail

Ichino, A. (2024). Che cos'è una teoria del complotto? (E perché questa è una domanda importante), "Rivista di filosofia, Rivista quadrimestrale" 2/2024, pp. 281-302, doi: 10.1413/114468.

Janiesch, C., Zschech, P. & Heinrich, K. (2021). Machine learning and deep learning. Electron Markets 31, 685–695, https://doi.org/10.1007/s12525-021-00475-2

Johnston, N. (2020). Living in the World of Fake News: High School Students. Evaluation of Information from Social Media Sites. Journal of the Australian Library and Information Association 69, no. 4: 430–50. https://doi.org/10.1080/24750158.2020.1821146

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, per- spectives, and prospects. Science, 349(6245), 255–260. https://doi. org/10.1126/science.aaa8415

Kahneman, D. (2011). Thinking, fast and slow, Farrar, Straus and Giroux.

Kumar, A., and James W. T. (2024). Feature Importance in the Age of Explainable AI: Case Study of Detecting Fake News & Misinformation via a Multi-Modal Framework, European Journal of Operational Research 317, no. 2: 401–13. https://doi.org/10.1016/j.ejor.2023.10.003

Kurni, M., Mohammed, M.S. and Srinivasa, K.G. (2023) 'AI-Enabled Gamification in Education', in M. Kurni, M.S. Mohammed, and S. K G (eds) A Beginner's Guide to Introduce Artificial Intelligence in Teaching and Learning. Cham: Springer International Publishing, pp. 105–114. Available at: https://doi.org/10.1007/978-3-031-32653-0_6 .

Lai, V., and Chenhao T. (2019). On Human Predictions with Explanations and Predictions of

Machine Learning Models: A Case Study on Deception Detection, Proceedings of the Conference on Fairness, Accountability, and Transparency, 29–38. FAT* '19. New York, NY, USA: Association for Computing Machinery, https://doi.org/10.1145/3287560.3287590 .

Machlup, F. (1962). The production and distribution of knowledge in the United States, Princeton, N.J., Princeton University Press.

McKenzie, S. (2024). NewsGuard "Monthly AI Misinformation Monitor of Leading AI Chatbots", https://www.newsguardtech.com/ai-monitor/december-2024-ai-misinformation-monitor/

Mienye, I.D.; Swart, T.G. (2024). A Comprehensive Review of Deep Learning: Architectures, Recent Advances, and Applications, Information 2024, 15, 755, https://doi.org/10.3390/info15120755

Molenaar, I. et al. (2025) 'Artificial Intelligence and Education: Different Perceptions and Ethical Directions', in N.A. Smuha (ed.) The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence. Cambridge: Cambridge University Press (Cambridge Law Handbooks), pp. 261–282. Available at: https://doi.org/10.1017/9781009367783.017

Montoro Montarroso, A., Javier C.-C., and Gómez Romero, J. (2023). Fighting Disinformation with Artificial Intelligence: Fundamentals, Advances and Challenges, https://doi.org/10.3145/epi.2023.may.22newsguardrealitycheck.com/p/deepseek-ai-chatbot-china-russia-iran-disinformation

Pariser, E. (2011). The filter bubble : what the Internet is hiding from you, Viking/Penguin Press.

Pawlicka, A., Pawlicki, M., Kozik, R., Andrychowicz-Trojanowska, A., and Choraś, M. (2024). AI vs Linguistic-Based Human Judgement: Bridging the Gap in Pursuit of Truth for Fake News Detection, Information Sciences 679: 121097. https://doi.org/10.1016/j.ins.2024.121097

Pérez-Escoda, A. (2022). Infodemic and Fake News Turning Shift for Media: Distrust among University Students, Information 13, no. 11: 523. https://doi.org/10.3390/info13110523

Sarker, I.H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions, SN COMPUT. SCI. 2, 160, https://doi.org/10.1007/s42979-021-00592-x

Saxena, D., Yasobant, S. (2022). Information Overload, in: Schintler, L.A., McNeely, C.L. (eds) Encyclopedia of Big Data, Springer, Cham. https://doi.org/10.1007/978-3-319-32010-6_374

Selnes, F. N. (2024). Fake News on Social Media: Understanding Teens' (Dis)Engagement with News, Media, Culture & Society 46, no. 2: 376–92. https://doi.org/10.1177/01634437231198447

Sitek, M. (2024). Przestrzeń informacyjna młodych Polaków: przegląd badań, Instytut Badań Edukacyjnych, https://open.icm.edu.pl/handle/123456789/25060

Skumanich, A., and Han Kyul, K. (2024). Modes of Analyzing Disinformation Narratives With AI/ML/Text Mining to Assist in Mitigating the Weaponization of Social Media, arXiv, https://doi.org/10.48550/arXiv.2405.15987

Spitale, G., Biller-Andorno, N., Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans, Science.org.

Suffia, G. (2022). Pulire l'infosfera. Intelligenza artificiale e contrasto alla disinformazione, Giuffré Francis Lefebvre.

Thirumala, A., Ferracane, E. (2023). Clickbait Classification and Spoiling Using Natural Language Processing, https://doi.org/10.48550/arXiv.2306.14907

Turing, A. M. (1950). Computing Machinery and Intelligence, Mind 49: 433-460.

Tversky, A., Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases, Science, 185, pp. 1124-1131.

Valencia-Arias, A., Arango-Botero, D. M., Cardona-Acevedo, S., Soledad Paredes Delgado, S., and Gallegos, A. (2023). Understanding the Spread of Fake News: An Approach from the Perspective of Young People, Informatics 10, no. 2: 38. https://doi.org/10.3390/informatics10020038

Vartiainen, H., Kahila, J., Tedre, M., Sointu, E., and Valtonen, T. (2023). More than Fabricated News Reports: Children's Perspectives and Experiences of Fake News, Journal of Media Literacy Education 15, no. 2: 17–30. https://doi.org/10.23860/JMLE-2023-15-2-2

World Health Organization, (2024). https://www.who.int/news-room/questions-and-

answers/item/disinformation-and-public-health

Wurman, R. S. (1989), Information Anxiety, Doubleday.

Xu, Z., Sanjay Jain, and Mohan Kankanhalli (2024). Hallucination Is Inevitable: An Innate Limitation of Large Language Models, arXiv, https://doi.org/10.48550/arXiv.2401.11817
Yankoski, M., Walter, S., and Tim, W. (2021). Meme Warfare: AI Countermeasures to Disinformation Should Focus on Popular, Not Perfect, Fakes, Bulletin of the Atomic Scientists 77, no. 3: 119–23. https://doi.org/10.1080/00963402.2021.1912093

Ziccardi, G. (2022). Diritti digitali. Informatica giuridica per le nuove professioni, Raffaello Cortina editore.