FERMI
FAKE NEWS
RISK MITIGATOR

Digital Trust: A Practical Path to Combating Disinformation and Fostering Resilience

"Digital Trust: A Practical Path to Combating Disinformation and Fostering Resilience" has been designed in the context of FERMI (Fake nEws Risk MItigator) [Project 101073980], a Horizon Europe project that studies and attempts to counter the root causes, spread and implications of disinformation and fake news. This training material is inspired and derived primarily from the insights shared during the FERMI webinar "Digital Trust in Action: Technological Approaches and Citizen Empowerment to Combat Disinformation" organised by Convergence on 04/12/2024.

The objective of this resource is to explore the challenges posed by disinformation, highlight the essential components of digital literacy, and provide tools and strategies to mitigate their impact. Designed as a practical resource, it includes reflective exercises and actionable insights to empower individuals in combating disinformation and fostering digital trust. This document should ideally be read before or after viewing the webinar recording (found on the FERMI website) thus offering a comprehensive package that provides in-depth knowledge, fosters understanding, and encourages critical engagement with the topics of digital literacy and disinformation.

For further context and additional materials, readers are encouraged to explore the FERMI website. Specifically, it is suggested to read "Navigating Disinformation: A Comprehensive Guide" and watch the first FERMI webinar, "A Dive into the Societal Landscape of Disinformation - Balancing between Law Enforcement and Fundamental Rights to Increase Digital Trust" held on 23/02/2024.

PARTNERS

Atos, netcompany intrasoft, itml innovation applied, INOV, BIGS, ianus consulting, the Lisbon council, CONVERGENCE, VUB Vrije Universiteit Brussel, KU LEUVEN, POLIISI, Police Fédérale, Polisen

Funded by the European Union

fighting-fake-news.eu

KEY THEMES: DISINFORMATION, DIGITAL LITERACY AND ARTIFICIAL INTELLIGENCE (AI)

The key topics that are going to be discussed in this document are disinformation and digital literacy, as well as the role of AI in both spreading and countering false information, whether intentional (disinformation) or unintentional (misinformation). Therefore, it is essential to provide a basic background to these concepts before we dive into their exploration.

DISINFORMATION

Disinformation poses a significant threat to societies worldwide, impacting political processes, public health, and social cohesion. Disinformation interferes with the quality of democracy; it diminishes democratic trust and can provoke polarisation online [1]. It can lead to real-world crimes and violence, making it crucial to develop effective strategies to counter them.

It is important to note that there are many different approaches to disinformation. Unfortunately, there is a gap in the definition of the term and there is still no common consensus on what constitutes it. A common definition of disinformation, used in policy-making settings, retrieved from a key EU document is the following:

> “Disinformation is understood as verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm” [2].

Disinformation can be used for political manipulation, financial gain, or to spread distrust. Some common tactics used to spread disinformation include creating and sharing fake news articles, spreading rumours and conspiracy theories, and using bots and fake accounts to amplify misleading content.

The FERMI project aims to address this challenge by developing a Comprehensive framework and a set of analytical tools to combat disinformation.

[1] Colomina, Carme, et al., The impact of disinformation on democratic processes and human rights in the world. Brussels: European Parliament (2021): 1-19.

[2] European Commission, Action Plan against Disinformation (Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 2018), p. 1. Available at: https://eur-lex.europa.eu/legal-content/GA/TXT/?uri=CELEX:52018JC0036

DIGITAL LITERACY

As digital technologies continue to grow and evolve rapidly, it has become increasingly important for individuals to develop the necessary skills and competencies to perform tasks and solve problems in digital environments. These capabilities are collectively referred to as digital literacy skills, as highlighted in various studies [3].

The concept of digital literacy was first introduced in 1997 and was defined as the ability to understand and use information across various digital platforms, going beyond basic computer skills. Digital literacy encompasses critical thinking and effective communication using digital technologies. The advancement of digital technology has altered our everyday routines and how we engage with our surroundings.

Digital literacy can assist raise awareness of the digital world by teaching people how to use technology and digital tools to meet their basic needs. Strong digital literacy makes it easier for people to meet their respective needs and adapt to the constantly shifting demands of the digital world [4].

The capacity to use digital technologies effectively and responsibly includes understanding how to use digital tools like computers, smartphones, and the internet, as well as being able to find, evaluate, and use information from online sources. Digital literacy is essential for navigating our increasingly digital world and for combating the spread of disinformation. It's important to mention that digital literacy goes beyond technical skills and encompasses ethical considerations, such as using online platforms responsibly and avoiding the spread of misleading content.

ARTIFICIAL INTELIGENCE (AI)

AI is a term that has become particularly well-known to the wider public over the last few years. It refers to a branch of computer science that involves creating intelligent agents that “simulate human learning, comprehension, problem-solving, decision-making, creativity and autonomy” [5]. Tools using AI technology, like ChatGPT, can perform tasks such as writing text, translating languages, and recognising images.

[3] Reddy, P., Sharma, B., & Chaudhary, K. (2020). *Digital Literacy: A Review of Literature*. International Journal of Technoethics (IJT), 11(2), 65-94. https://doi.org/10.4018/IJT.20200701.oa1

[4] Bashar, Ummul & Naaz, Ishrat. (2024). *Digital Literacy: The Importance, Initiatives and Challenges*. 10.56726/IRJMETS56658.

[5] IBM, *What is AI?* (Updated 16 August 2024, Contributors: Cole Stryker, Eda Kavlakoglu). Available at: https://www.ibm.com/topics/artificial-intelligence

AI is rapidly evolving, and its capabilities are increasing at an unprecedented rate. Undoubtedly, AI has the potential to profoundly transform many sectors, namely healthcare, education, and software engineering. For instance, AI-driven tools in healthcare are enabling faster diagnoses and personalised treatment plans. In education, AI supports adaptive learning platforms that cater to individual student needs, making education more accessible and engaging. Meanwhile, in software engineering, tools powered by generative AI are significantly enhancing productivity by automating routine coding tasks and assisting with debugging. While for some sectors (e.g., healthcare and education) great opportunities are presented, others like software engineering and customer service are already experiencing disruption. In the sense that AI tools, such as those used by major tech companies, are increasingly being integrated into workflows, enhancing productivity but also raising ethical and workforce-related concerns.

While these advancements highlight AI's transformative power, they also reveal its dual-use nature. AI tools that drive innovation and efficiency can also be weaponised for harmful and unethical purposes, including the creation and spread of disinformation. For instance, generative AI models capable of producing realistic text or images are increasingly used to create deepfakes or misleading content, which can amplify false narratives and erode public trust. This dual capability highlights the urgent need to address ethical and regulatory challenges as AI continues to shape our digital landscape.

DISINFORMATION: A CONTEMPORARY THREAT

Digital literacy is crucial in today's digital age, as it allows individuals to navigate the online world safely and responsibly. At present, the threat of disinformation is increasingly fuelled by the sophistication of AI technologies. One prominent example, as introduced in the previous section, is the creation of deepfakes, which leverage AI to produce realistic but fabricated content, posing significant risks for trust and authenticity in digital media. These tools, while potentially used for entertainment or educational purposes, are increasingly exploited by malicious actors for harmful activities, including spreading disinformation.

Disinformation campaigns are often orchestrated by extremist groups, who exploit digital platforms to spread propaganda and incite violence. These campaigns often target vulnerable communities, manipulating their fears and prejudices to achieve political or ideological goals. An example mentioned during the webinar, referring to the Romanian election influenced by a targeted TikTok campaign and the rise of far-right extremism, underscores the tangible impact of disinformation on democratic processes and societal stability.

By cultivating critical thinking skills, individuals can become informed and responsible digital citizens, capable of distinguishing fact from fiction and contributing to a trustworthy digital environment.

SECTION 1: UNDERSTANDING THE IMPORTANCE OF DIGITAL LITERACY

Digital literacy is described as a set of skills and knowledge needed to navigate the digital world effectively, responsibly, and critically. It includes the access to and proficiency of basic technological tools, which, although fundamental, are not guaranteed across the world, but it encompasses more than that. Rather than just knowing how to use technology, it's about understanding how technology works, how it impacts our lives, and how to use it safely and ethically.

In today's world, digital literacy is essential in a multitude of situations. Not only do many everyday processes in banking, health or government bureaucracy require digital skills, but there is also an emerging need for informed decisions regarding privacy protection. Furthermore, a vast majority of pursuits from education purposes to employment require at least some level of digital literacy.

In the current digital landscape, it is essential to understand the broader implications of technology, think critically about the information encountered online, and behave responsibly in the digital space. The most important element of digital literacy nevertheless seems to be the ethical and responsible use of digital tools, this involves understanding the potential for misuse, such as spreading misinformation, whether intentionally or unintentionally. Digital literacy is a constantly evolving field, and, as technology changes, upskilling is necessary.

THE USE OF BASIC DIGITAL TOOLS

As mentioned before, digital literacy is primarily connected to the proficient use of common digital tools. This proficiency is essential in the current digitised environment as it empowers individuals to effectively access information, connect with others, and participate in various aspects of society. We should underline that a high level of digital literacy does not have to be synonymous with programming skills. Some indicative examples of utilising digital tools are the following:

Use for communication and collaboration: This could be using emails or online collaboration tools (like Google Docs or Microsoft Teams) and being able to understand the nuances of online communication.

The ability to find and evaluate information: Being able to effectively use search engines, assess the credibility of websites and online sources, and distinguish between fact, opinion, and misinformation. For example, knowing how to spot signs of a fake news website (e.g., poor design, grammatical errors, sensationalised headlines) or being able to verify information by cross-referencing it with reputable sources is crucial.

ETHICAL AND RESPONSIBLE USE

On top of achieving adequate knowledge of the technological tools comes the ethical and responsible use of them. This implies carefully considering the potential impact of online actions and engaging to create a positive and respectful online environment. To prevent the spread of misinformation, verifying information before sharing it is crucial, as well as keeping a critical and cautious approach on which are some trustful sources. Overall, ethical digital citizenship translates as recognising that online actions have possible real-world consequences and being responsible for the created and shared content. More specifically, examples of ethical and responsible use can include:

- **Creation and sharing of content**: Ethical content creation encompasses not only using appropriate language and images, but also being mindful regarding privacy issues, understanding and complying with copyright laws, and fact-checking information before sharing it.
- **Data Privacy and Security**: The protection of personal information is a key component of responsible use. When using digital and online tools, individuals should understand how their data are collected and used, create strong passwords, safeguard their data, and be aware of common online scams and phishing attempts.

SECTION 2: DEVELOPING CRITICAL THINKING

CRITICAL THINKING: A POWERFUL TOOL

Another important aspect connected to digital literacy in the current digitalised environment is cultivating critical thinking skills. Critical thinking enables individuals to properly evaluate information retrieved online, distinguish between fiction and facts, and consequently avoid being easily manipulated. Going a bit further, being able to summarise and synthesise information, meaning that individuals can extract valuable insights from a source and present them clearly and understandably, is another crucial aspect of digital literacy, in a world with information overload.

Since critical thinking is an acquired skill, insufficient opportunities or efforts to cultivate it can lead to an observed lack of critical thinking. More and more people, especially the new generations, are relying heavily on AI tools like ChatGPT for writing tasks, jeopardising the development of their critical thinking skills. Since writing is a process that promotes deep engagement and requests inductive and pedagogical reasoning it is deeply associated with critical thinking, thus, the lack of engagement in this activity reduces the opportunities for improvement.

In addition, the overwhelming amount of available information online paired with the fast-paced nature of interactions, may lead to a tendency towards “fast thinking” as well as quick judgements. This impulse not only can easily lead to inaccurate conclusions but also interferes with the development of critical thinking.

PRACTICAL STEPS TO ENHANCE CRITICAL THINKING

The enhancement of critical thinking skills is an ongoing process and there are several ways to achieve it. The first step is to actively practice it, by trying to question the information encountered online, evaluating the source, trying to identify potential biases and more importantly the intent behind the shared message. Individuals can try to incorporate in their practice the concept of “thinking slow”, meaning to pause and reflect upon the information before sharing or reacting to it.
Possible questions to pose to oneself include:

- What is the source of this information, and is it credible?
- Could this information be biased or intentionally misleading?
- How might sharing this information affect others?

Another step is to engage in continuous learning through online resources and training programmes it is possible to better understand AI, digital tools, and disinformation tactics. Moreover, individuals can improve their information evaluation skills since, usually, misleading content carries some characteristics that can be recognised through practice. Some indicative examples include, but are not limited to, a lack of credible sources, inconsistencies in the information presented, grammatical errors, provocative headlines, and emotional appeals.

Furthermore, one must be mindful of their online behaviour to not spread disinformation unintentionally, by verifying information before sharing it. Finally, journaling can be a great habit that enables reflection on the information received, facilitates the filtering of thoughts, and promotes a more critical engagement with information.

CAN AI BE A FORCE FOR GOOD? FERMI's PLATFORM

While AI poses challenges and facilitates the spread of disinformation, it also offers tools to combat it, a robust example is provided by the platform created within the framework of this project. The FERMI project aims to create a platform that analyses disinformation campaigns, considering socio-economic factors that contribute to their spread. The FERMI platform comprises several AI-powered modules, that anticipate, analyse, and mitigate criminal activities instigated by the spread of false information.

These modules include:
- **Disinformation** Analyser which identifies and analyses disinformation campaigns on social media platforms like X and Mastodon.
- **Crimes Impact Predictor** which forecasts potential rises or falls in crime rates connected to the influence of disinformation.
- **Behaviour Profiler and Socioeconomic Analyser** which merges financial data from specific regions with the crime predictions generated by the Crimes Impact Predictor module to calculate the estimated impact of disinformation within this given region.
- **Community Resilience Modeler**, which assesses the likelihood of politically motivated crime and proposes countermeasures for law enforcement agencies.
- **Sentiment Analysis Module** which explores the emotional polarity of social media posts about disinformation.
- **Swarm Learning Module**, a module that utilises federated learning, allowing AI models to be trained on data from multiple law enforcement agencies while maintaining data confidentiality.

Concluding this section, it should be underscored that even with such sophisticated tools as the FERMI platform, which uses AI to analyse and predict the impact of disinformation, human critical thinking still remains essential. Although AI can assist in identifying and mitigating the spread of misinformation, AI systems are ultimately created by humans and trained on data that may contain inherent biases. Addressing these biases requires ongoing research, as well as transparency in how AI models are developed and trained. Users must remain vigilant and critically assess AI-generated content to ensure that biases are recognised and addressed effectively. Therefore, individuals need to develop their critical thinking skills to carefully evaluate information, consider the motivations behind online content, and engage in responsible online behaviour.

COUNTERING DISINFORMATION: A COLLABORATIVE APPROACH

Collaboration is key to harness the potential of AI while mitigating its risks. As discussed during the webinar, to counter disinformation and build digital trust, a collaborative effort involving individuals, tech companies, governments, and civil society is essential.

The value of digital literacy and critical thinking among citizens cannot be underestimated, but it is not the only factor that is necessary to tackle disinformation. The crucial role of governments should be highlighted, in promoting digital literacy initiatives, regulating the use of AI, and fostering a resilient digital ecosystem.

Moreover, civil society organisations play an important role in raising awareness about disinformation, through the development of educational campaigns and relevant resources, the promotion and support of media literacy, as well as the monitoring and reporting of disinformation and holding stakeholders accountable.

In addition, the responsibility of technology companies should not be overlooked in the prioritisation of AI safety and alignment; a need for the allocation of more resources to develop safeguards against the misuse of AI technologies is emerging. Through this collaboration among individuals, tech companies, governments, and civil society, technology can be leveraged while at the same time its risks are being mitigated for building a more trustworthy digital environment.

SECTION 3: USEFUL RESOURCES

TRAIN YOURSELF: ONLINE COURSES

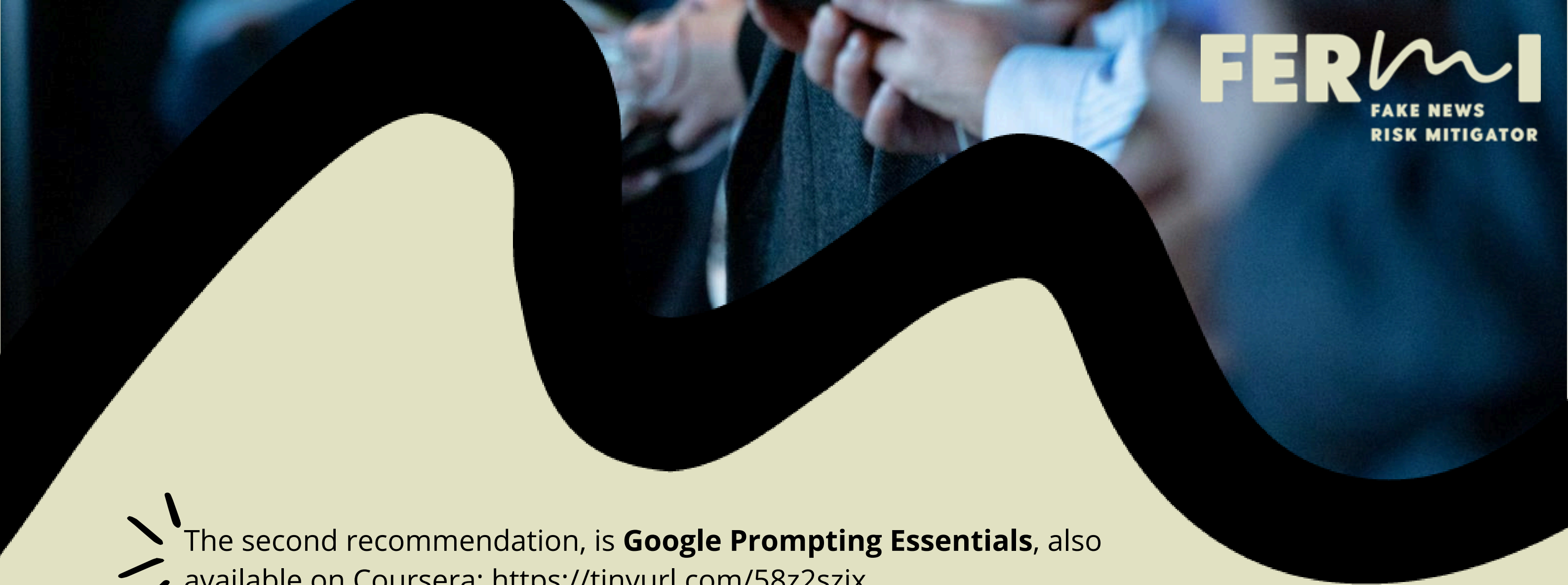
To remain relevant in a rapidly evolving digital landscape, individuals can proactively enhance their understanding of AI and its implications. Attending relevant courses can be a valuable starting point. Following are two indicative suggestions, but there is a great variety online depending on one’s needs [6].

The first recommendation is a course called **Google AI Essentials,** which is available online through Coursera: https://tinyurl.com/ye266e3n

The modules cover the following themes:
- Using AI tools to create content
- Training in clear and specific prompts
- Responsible use of AI
- Strategies to stay up to date in the emerging landscape of AI

There is a free option if the attendee doesn’t need a certificate, and a small fee is required for the acquisition of a certificate.

[6] The following courses as well as online tests (next section) were recommended during the webinar by David Timis, AI & Future of Work Expert.



The second recommendation, is **Google Prompting Essentials**, also available on Coursera: https://tinyurl.com/58z2szjx
This course covers the following themes:
- 5 steps system to write effective prompts
- Prompting techniques for everyday work tasks
- Prompting techniques for faster data analysis and creation of presentations
- Prompting techniques for the creation of AI agents to role-play conversations

PRACTICE SPOTTING DISINFORMATION: INTERACTIVE TESTS

As analysed herein, often users encounter online content and are asked to assess its accuracy and validity. Thus, it can be very effective to "train" themselves to recognise what disinformation and misinformation are and how to detect it. For that reason, two online tests follow, which can be a fun way to train oneself:

1. The first one is called **Find the FAKE** and it is addressed to the whole family. Through simple questions and pictures, the players have to guess whether a piece of information is “fact or fake” and they get trained on how to do fact-checking: https://tinyurl.com/54kp2n27

2. The second one is called **Real or Not** and asks the player to recognise whether an image is created by AI or not and then provides a score in the end: https://tinyurl.com/53fwpwjk

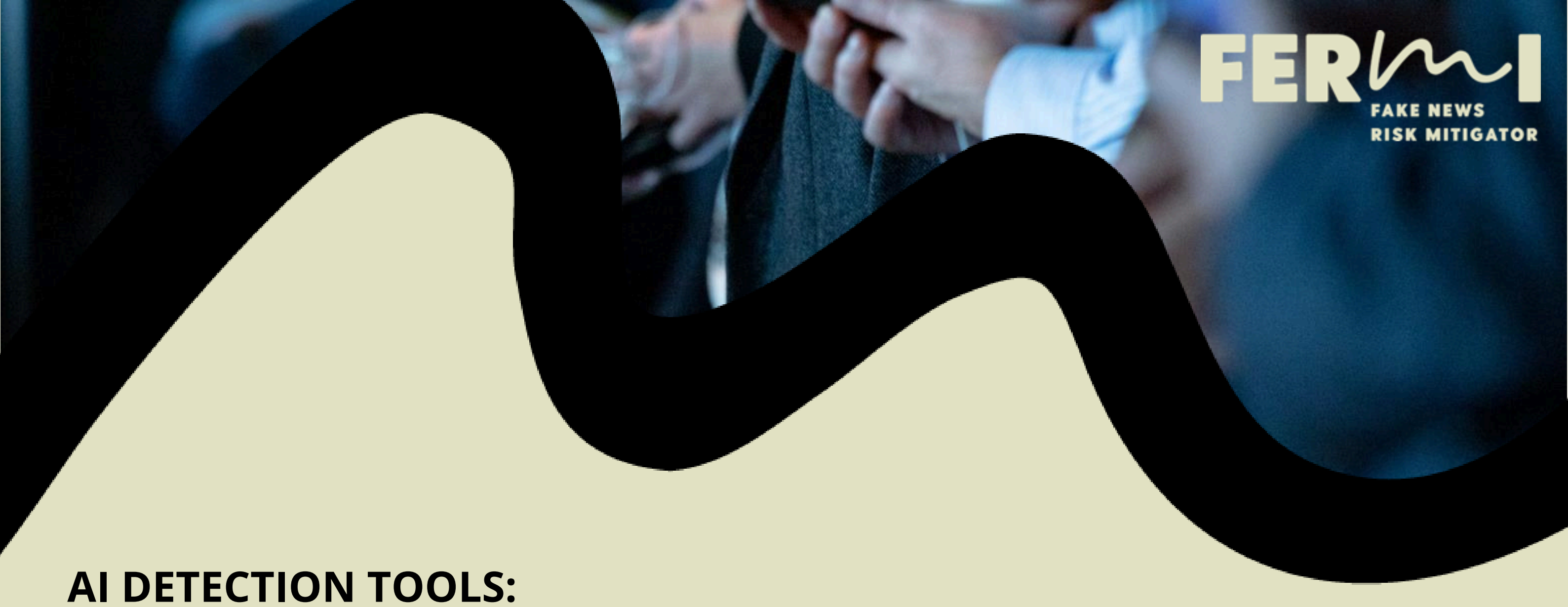
FACT-CHECKING: VERIFICATION TOOLS

Another useful resource in the fight against disinformation could be fact-checking tools. Fact-checking tools are digital resources designed to verify the accuracy and credibility of information found online to help users identify false or misleading content by cross-referencing claims with reliable sources or databases. Two recommendations are the following [7]:

1. **Google Fact Check Explorer:** A simple yet effective tool where users search with keywords regarding the validity of news or information. They get results based on articles from fact-checking organisations with ratings on the truthfulness of the information.
https://tinyurl.com/3v9hebhy

2. **AFP Fact Check:** AFP Fact Check is a department within Agence France-Presse (AFP), and it provides fact-checked articles from different websites.
https://tinyurl.com/36we6hnh

[7] The following fact-checking and AI detection tools were retrieved from **“A Toolkit for Identifying Disinformation and Strengthening Media Literacy”**, a resource created within the Erasmus+ project Anti-Rumour. Available at https://anti-rumour.eu



AI DETECTION TOOLS:

Given that we have already highlighted the widespread use of AI in content creation, this section proposes some useful AI content detection tools. These tools can play a vital role in identifying content generated or refined by AI, whether written text or images. These tools help users discern authentic material from AI-created fabrications, protecting individuals and organisations from being misled. Three tools are introduced:

1. **GPTZero:** This is a useful tool for detecting AI-generated text. It is easy to use online and it shows the probability that a certain text is created by AI, especially in English its accuracy is high however the rating should be treated carefully.
   https://tinyurl.com/5n7canch

2. **DEEPFAKE-O-METER:** This is an open platform that detects whether an image, a video, or an audio file was created using AI. Although it is free to use it requires setting up an account.
   https://tinyurl.com/46r4ee76

3. **Content at Scale AI Image Detector**: A simple tool that predicts whether an image was designed or photographed by a human or was AI-generated. Its use is free, and it is quite reliable, especially for images with high resolution. Moreover, in addition to directly uploading images, it supports providing their URLs.
   https://tinyurl.com/4pzvnmt4

KEY TAKEAWAYS

The technological landscape will continue to evolve at a rapid rate posing new challenges in building digital trust. The empowerment of citizens towards the creation of a more trustworthy digital environment emerges as a necessity. Thus, we must move beyond simply learning about digital literacy and responsible technology use and begin to actively apply these concepts and skills in our daily lives. This procedure, apart from using available training resources, could involve engaging in discussions about these important issues with peers, colleagues, and decision-makers to not only understand how to use digital tools effectively but also to be aware of the ethical implications of our digital actions. This is particularly important given the increasing accessibility and power of AI tools, which can be used for both positive and negative purposes. Another crucial aspect of individual empowerment is taking an active role in combating disinformation; this can include flagging inappropriate content on social media platforms and engaging in fact-checking to verify information before sharing it. Of course, the continuous effort to enhance our critical thinking skills will also play a valuable role in identifying misinformation and building a resilient digital environment.