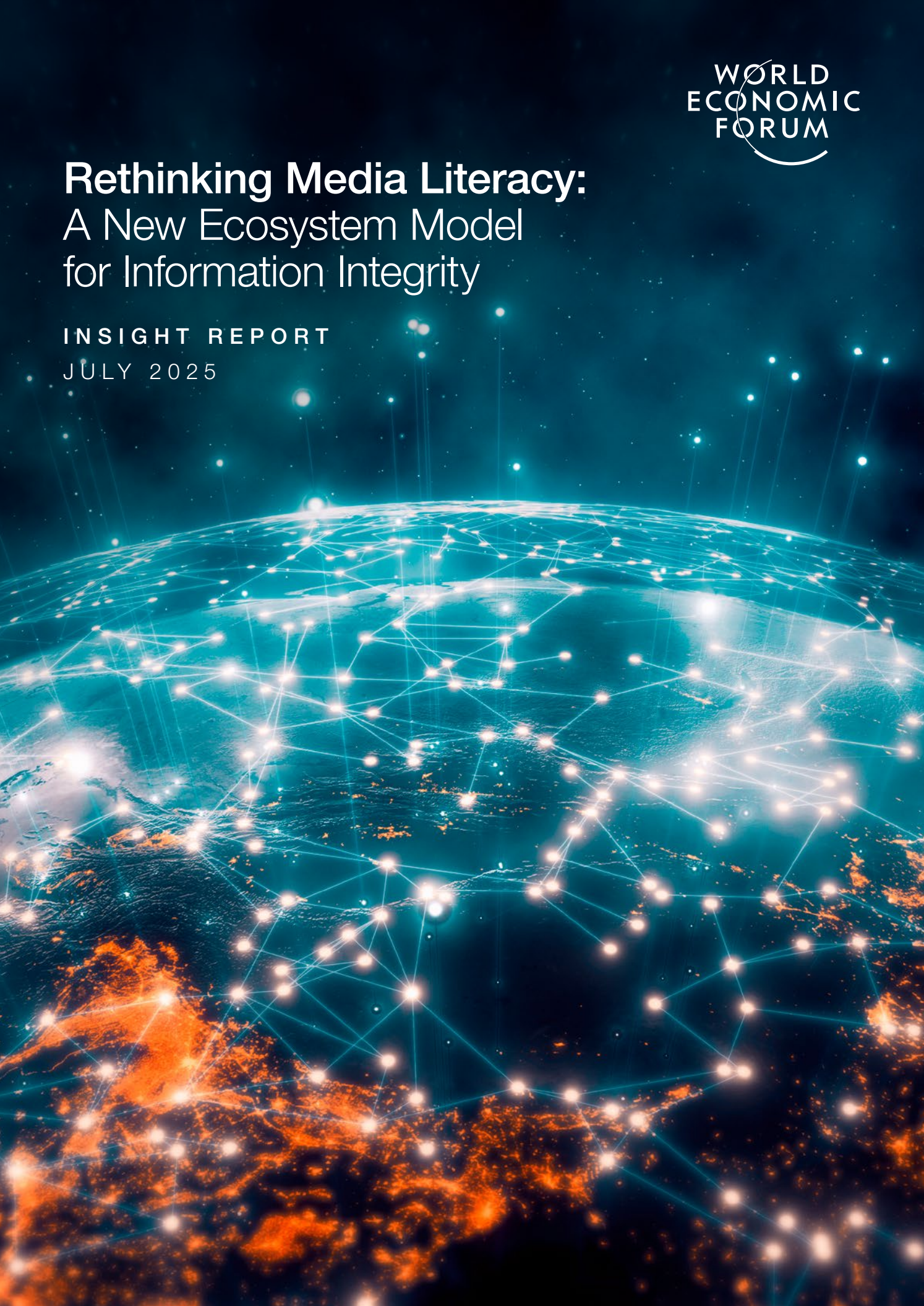
WORLD ECONOMIC FORUM

Rethinking Media Literacy: A New Ecosystem Model for Information Integrity

INSIGHT REPORT

JULY 2025



Images: Getty Images

Contents

Foreword

Sasha Havlicek
Chief Executive Officer,
Institute for Strategic Dialogue

Agustina Callegari
Lead, Technology Governance, Safety and International Cooperation, World Economic Forum

Daniel Dobrygowski
Head, Governance and Trust,
World Economic Forum

The ability to navigate and critically engage with information is a fundamental skill in today's digital age. Media and information literacy (MIL) lies at the core of efforts to strengthen information integrity and foster a safer, more trustworthy online environment. However, the onus cannot be solely on individual information consumers to assess credibility, when the mechanisms and tools to enable deception at ever-greater speed, scale, sophistication and complexity are ubiquitous. In such an environment, responsibility must be shared by a whole range of stakeholders. This includes the structures that exist to incentivize certain behaviours over others.

This report presents a new model for a comprehensive, whole-of-society approach to MIL. It aims to provide a framework for mapping the range of interventions needed and optimizing efforts to mitigate the proliferation and impact of disinformation while protecting free expression. It builds on the work of the Global Coalition for Digital Safety, particularly its focus on effective digital safety interventions, including education-based strategies.

The Coalition's work highlights the crucial role that multistakeholder approaches play in shaping healthier digital ecosystems. The aim of the Coalition is to help people develop the skills needed to better access, analyse, evaluate, create and responsibly consume media while at the same time encouraging interventions that disincentivize the production and viral distribution of disinformation. As seen in the *Global Risks Report 2025* – the World Economic Forum's annual survey of 900 experts worldwide on global threats – disinformation and misinformation were identified as the top risk over the next two years, reflecting mounting concerns about their impact on democracy, social cohesion and global stability.

This report not only highlights the urgent need for enhanced MIL but also offers practical approaches to embed it across different sectors, age groups and regions. It reinforces the idea that digital safety is a process extending beyond the individual to communities, public institutions, multilateral bodies and the private sector. The development of online spaces is a shared societal endeavour that must start with education and empowerment.

Moreover, the model presented in this report can serve as a valuable tool for understanding a wider range of challenges that affect the information environment beyond media literacy alone – including online harms, trust erosion and emerging threats driven by new technologies. It reflects the Global Coalition for Digital Safety's ongoing commitment to developing scalable, collaborative solutions that not only address immediate risks but also build the foundations for a healthier, more resilient digital ecosystem in the future.

Executive summary

Today's digital era has been shaped by the rapid evolution of social media platforms and, increasingly, the proliferation of generative AI.

As public confidence in information – whether from legacy media, public institutions or social media – has plummeted, the ability to critically engage with information and understand the technologies and mechanisms underpinning its distribution is essential for preserving democratic civic discourse, public safety and meaningful freedom of expression.

Media and information literacy is the interrelatedness of competencies regarding information; it is central to the safeguarding of information integrity as it equips individuals with the competencies to access, analyse, evaluate and produce information across various formats and platforms. It should support not only personal empowerment but also broader goals, such as democratic resilience, public trust, civic participation and social cohesion. As the information landscape grows more complex, MIL becomes essential for enabling people to distinguish between credible information on the one hand and falsehoods and manipulation on the other – especially as AI-generated media blurs the lines between fact and fiction.

However, while regulatory or platform-based approaches often struggle to keep pace with the emergence of new harms, evasive tactics and cultural norms, an overly narrow understanding of MIL risks obscuring the bigger picture and placing undue responsibility on individuals. The spaces that consumers inhabit must also allow credible information to be easily sourced and shared, and they must incentivize safety at the platform level. Increasingly, digital public commons are struggling to meet this need. Both visibility and engagement are often skewed, whether towards orchestrated deception, low-trust sources or other forms of divisive and manipulative content. As such, the challenge of embedding stronger MIL is combined with a more systemic need: to build healthy incentives into the information ecosystem at all levels.

This report aims to present a more holistic model, identifying all possible entry points for intervention and putting greater focus on the supply-and-demand dynamics of information spaces.

This new model, designed to aid the effectiveness and expansion of information integrity via a comprehensive approach to MIL, combines two conceptual frameworks:

- **The disinformation life cycle:** Five stages in the digital information life cycle that offer opportunities for intervention – pre-creation, creation, distribution, consumption and post-consumption

- **The socio-ecological model:** Five levels where interventions have an impact – at the individual, interpersonal, community, institutional and policy levels

By aligning these two approaches, the framework allows practitioners, policy-makers, the private sector and educators to identify gaps in the current MIL landscape as well as strategic points of potential engagement, with a view to enabling scaled impact over time.

Drawing from global case studies, this report demonstrates that MIL can be successfully embedded across diverse contexts, from youth organizations and media institutions to local governments and digital platforms.

Ultimately, strengthening information integrity requires coordinated action across education systems, civil society, governments and the private sector. The report offers a strategic foundation for that commitment, advancing a vision in which digital safety becomes a shared responsibility and a universal competency.

Introduction

Ensuring information integrity while safeguarding free expression is key to preserving public trust, democratic resilience, societal cohesion and individual empowerment.

The unprecedented speed and scale at which information travels, enabled by global connectivity, social media platforms and emerging technologies, have transformed how societies communicate, access knowledge and make decisions. Yet this transformation has also created vulnerabilities. The rapid circulation of misleading, inaccurate or manipulated information, whether intentional or not, can erode trust in public institutions, polarize communities and amplify social tensions. Perhaps most importantly, disinformation disrupts the ability of individuals to freely make informed choices about what is in their own best interest.[1]

The complexity of the information ecosystem, compounded by algorithmic amplification and the rise of synthetic media, has made it more difficult for people to discern reliable information from false or deceptive content. These challenges not only affect public health responses, electoral integrity and crisis management but also have profound consequences for the exercise of fundamental rights and the health of democratic societies.

The rise of generative artificial intelligence (GenAI) has further exacerbated the challenge. Tools capable of creating realistic images, audio, videos and text have lowered the barriers to producing and disseminating highly convincing and personalized false content. Even as early as 2020, a study found that in 50% of cases, humans could not differentiate between news created by a person and news generated by AI.[2] AI has substantially improved its ability to mimic human writing in the intervening years, along with audio, video and images. By age 11, children's confidence in evaluating online content often exceeds their actual competence,[3] while false information that has proliferated online is now being cited by AI large language models (LLMs), as individuals attempt to fact-check information which they encounter.[4]

As disinformation tactics evolve, so too must MIL initiatives, integrating insights from psychology, technology and education to remain effective in an ever-changing digital environment. The proliferation of user-friendly, relatively inexpensive and easily accessible applications has enabled the creation of synthetic media and the widespread interaction with non-human agents, such as AI-powered chatbots and virtual assistants. This shift underscores the need for teaching and training to equip people with the skills to critically evaluate synthetic media, discern credible information from AI-generated content (AIGC) and interact responsibly and ethically with AI systems.

In this context, MIL – which is defined as a set of competencies that empower individuals to access, understand, critically evaluate, create and responsibly share information and media content across different platforms and formats – is critical for building resilient societies and protecting individual freedoms. Beyond enabling individuals to defend themselves against manipulation or disinformation, the ability to seek, receive and impart information freely is a fundamental right, enshrined in international human rights frameworks such as Article 19 of the Universal Declaration of Human Rights. MIL acts as both a safeguard and an enabler of this right, ensuring that people are not silenced by manipulation, overwhelmed by disinformation or disenfranchised by their inability to critically engage with the information around them.

MIL serves as a foundational tool for furthering the education of informed digital citizens. It trains individuals to question sources, recognize biases and identify manipulative tactics. Moreover, it cultivates resilience against disinformation by promoting a culture of enquiry and reflection rather than passive consumption, and ideally it should provide citizens with an understanding of the digital information ecosystem in which they now live.

This report presents a holistic framework that situates MIL as one node across both the disinformation life cycle and the socio-ecological model (SEM) – a framework to capture the multiple, interacting layers of influence on digital safety, from personal behaviour to interpersonal and community dynamics, institutional obligations and policy levers. By applying this model, the report offers a structured approach to identify gaps in current interventions and supports organizations in more effectively targeting their strategies to strengthen information integrity and uphold fundamental rights. By analysing interventions at different stages – prevention, detection, response and resilience – and examining the influence of individual, community, institutional and societal factors, the framework offers a more comprehensive perspective for action. Indicative case studies offer practical examples of both MIL interventions and the application of this wider framework.

The objective of the report is two-fold: first, to assess the state of MIL efforts in the context of current information challenges; second, to unpack the disinformation life cycle and SEM, helping improve the design and targeting of more holistic interventions. It offers a new perspective on how to map and strengthen efforts that seek to bolster information integrity, including – but not limited to – MIL initiatives.

1 State of media and information literacy

More than ever before, people around the world require critical thinking skills to access and assess reliable information.

The importance of enhancing skills, knowledge and competencies in MIL has been widely recognized in international instruments, highlighting it as a key priority in the global fight against mis- and disinformation as well as hate speech and other online harms. The Pact for the Future adopted at the UN Summit of the Future in September 2024 positions MIL as an essential tool for promoting information integrity and fostering inclusive, resilient and informed societies.[5]

At the United Nations Educational, Scientific and Cultural Organization (UNESCO), MIL is a flagship programme and a vital contributor to sustainable development, information integrity and effective global governance of digital platforms. UNESCO has been the leading international organization in defining and expanding the scope of MIL and its essential competencies across the globe.

1.1 How to define media and information literacy

There are various definitions and interpretations of what constitutes media and information literacy, ranging from media literacy to digital education. UNESCO defines MIL as an umbrella concept stressing the interrelatedness of competencies related to information, particularly in media. These include: human rights literacy (focusing on the right to freedom of expression as an individual's right to communicate, seek and receive information and opinions); news literacy (which includes literacy about ethics and standards in journalism); advertising literacy; computer literacy; understanding of the so-called "attention economy"; intercultural literacy; privacy literacy; and so forth.[6] Today, AI literacy is also essential, given its growing role in shaping and disseminating information.

Additionally, different forms of literacy, such as visual, information, media, technology and foundational reading and writing, each offer distinct affordances and contribute in unique ways to how individuals engage with, interpret and respond to information. Recognizing and clarifying these differences is critical to understanding the specific functions and impacts of media literacy interventions.

MIL is concerned with a range of issues, including: people's understanding of content, how it is created and distributed, and by whom; how people make use of information (or not) and how they interact with libraries, the media and technology services (or don't); what knowledge, expertise, perspectives and approaches people need to evaluate information effectively; and how people can manage their engagement with information to achieve desired goals and objectives in their political, economic, social and cultural lives.[7]

In the digital age, MIL includes the fundamental skill set necessary for active participation using digital technologies, critical thinking, democratic engagement and intercultural understanding.

The core competencies of MIL may be categorized as:

- Understanding how information, media and digital communications can play a role in sustainable development and democracy
- Understanding content and its various uses
- Evaluation of information and media content, as well as media and information providers, for effectiveness, efficiency and the real-world implementation of ethical standards
- Comprehending the socio-cultural context of information, media and digital content
- Promoting MIL among learners/citizens and managing the necessary changes

1.2 How MIL can provide a response to the challenges of the digital age

The digital environment has democratized access to information, offering new avenues for the enjoyment of fundamental human rights, including freedom of expression. Today, however, digital platforms and online information threaten the very rights that internet connectivity once promised to fulfil. A few key trends that point towards the heightened need for enhanced MIL include:

Increased reliance on digital platforms to access public interest information

A recent UNESCO/IPSOS survey revealed that in 16 countries that were scheduled to have elections in 2024, social media was fund to be the primary source of information. Some 87% of citizens in these countries believed that online disinformation was already having a major impact on the political life of their country, and they feared its influence on election results.[8]

The rapid rise of digital platforms has created spaces where vast amounts of information are shared, which has significant social, political and economic impacts. However, these platforms can enable and accelerate the spread of misinformation, disinformation, hate speech and other harmful content, making it crucial to ensure they operate transparently and in alignment with human rights principles.

Reduced trust in traditional journalism and the growing influence of content creators as information channels

Studies show that people (especially youth) are increasingly turning to short-form video for news consumption. This type of format is particularly favoured by influencers and young news creators, who are increasingly becoming primary "news" creators and shaping public discourse on critical topics including elections, conflicts and environmental crises.[9] Video (and livestream, which is used by these same creators) creates an even greater moderation challenge for platforms that already struggle to apply their policies to harmful content.

In another concerning trend, a UNESCO-supported study showed that content creators (e.g. influencers) on digital platforms do not rely on traditional journalism to produce content, with mainstream news media ranked only as the third most common source used by these actors (36.9%). Alarmingly, 42% of content creators rely on likes and views as the primary indicator of credibility, indicating a shift away from traditional journalistic standards, where fact-checking and credibility are based on evidence and transparent citations.[10]

Such trends make a compelling case for increased MIL programmes that help users and individuals identify reliable news sources, understand the risks posed by artificial AI and mis- and disinformation on digital platforms and engage with content in an inclusive and ethical way. Further, MIL programmes must be designed to reinforce human rights, including the right to freedom of expression and access to information, empowering users to employ digital technologies and social media platforms in an open, safe and secure way.

Rapid developments in AI, including GenAI

Recent developments in AI are reshaping human society, influencing trust, media consumption and the broader information landscape.

AI has rapidly evolved from basic supervised and unsupervised learning models into highly complex deep learning (DL) algorithms, capable of handling unstructured data and performing advanced tasks such as image and text analysis, voice synthesis and predictive modelling. While these advancements offer significant benefits – such as improving healthcare diagnostics, streamlining content creation and enhancing personalized learning experiences – they also introduce critical risks, particularly concerning misinformation, bias and the erosion of trust in digital content. For example, deepfake technology – enabled by GenAI – has been used to fabricate realistic images and videos of public figures.[11] The increasing accessibility of such tools means that even non-experts can generate misleading content, further complicating efforts to embed information integrity.

The rise of GenAI models – such as such as those developed by OpenAI, Anthropic, Google and Meta among others – has added new layers of complexity to the challenge of MIL. These tools can produce convincing but misleading content, often blurring the line between what is human- or AI-generated. Research indicates that individuals already struggle with assessing the reliability of traditional search results, often assuming that higher-ranked pages are more credible. With AI-generated summaries becoming the default for many users,[12] there is a growing risk that misinformation, biases in training data or subtle manipulations could dictate public perception without users critically evaluating multiple sources.

Governments worldwide have responded differently to the rise of GenAI. Some countries, such as Italy, initially banned ChatGPT over privacy concerns before implementing regulatory measures,[13] while the European Union established the AI Act to provide more comprehensive regulatory oversight of AI systems, including requirements

for transparency, accountability and risk management.[14] The absence of global standards underscores the need for a parallel focus on societal resilience. Beyond governance, integrating AI literacy and data literacy into MIL curricula is essential. Currently, most MIL programmes do not include discussions on the political economy of AI-driven business models, despite their profound influence on information ecosystems. Addressing these gaps requires targeted interventions, such as embedding AI literacy in school curricula, training journalists to detect AI-generated content (AIGC) and equipping policy-makers with the necessary tools to assess AI-driven disinformation.

The challenge is particularly urgent for younger generations, who are increasingly relying on LLMs for information retrieval and research. Unlike traditional search engines, which encourage lateral reading by presenting multiple sources, LLMs provide a single answer, potentially discouraging critical analysis. This shift has major implications for digital media and information, as students and young professionals may be more inclined to trust AI-generated responses without verifying information through independent sources. Addressing this requires platforms to implement greater transparency in how AIGC is produced, while MIL programmes must adapt to equip individuals with the skills needed to navigate an AI-driven information ecosystem. Examples of emerging solutions include AI-detection tools such as Google DeepMind's SynthID, which watermarks and identifies AIGC. However, these efforts are still in their early stages and require significant scaling to achieve widespread adoption. Additionally, watermarking techniques have proven to be inconsistent and easily bypassed,[15] highlighting the need for more robust and multi-layered solutions.

In the absence of robust protection frameworks, any regulatory response to AI must be accompanied by efforts to strengthen public resilience to AI-driven disinformation. This includes proactive MIL interventions, partnerships with fact-checking organizations and collaborations with social media platforms to introduce friction in the sharing of deceptive AIGC. Without such measures, the rapid expansion of AI threatens to accelerate the spread of false information, further complicating an already volatile information landscape.

Increasing online harms and risks to digital safety

There are also growing concerns about wider online harms, such as hate speech and threats to digital safety, specifically for youth online. Some 78% of youth respondents to a survey conducted by the Office of the UN Secretary-General's Envoy on Youth reported having experienced digital threats, while 18% experienced them constantly.[16]

Online hate speech has become a pervasive issue, fuelling discrimination, inciting violence and deepening societal divides. The rise of mis- and disinformation, especially during global emergencies, has undermined public trust and stability, demonstrating the global impact of harmful content online.

Online harassment, threats and the non-consensual sharing of private information disproportionately target women, creating significant barriers to their participation in digital spaces and public life. Women journalists face attacks that aim to silence their voices, producing a "chilling" effect on freedom of expression.[17] Online harassment, abuse and disinformation campaigns are pervasive, often targeting women journalists with gendered threats of physical and sexual violence, and leading to self-censorship, psychological stress and even women leaving the profession.

As the US Surgeon General's Advisory found in 2023: "more research is needed to fully understand the impact of social media on children and adolescents; however, the current body of evidence indicates that while social media may have benefits, there are sufficient indicators that social media can also have a profound risk of harm to the mental health and well-being".[18] The glorification of mass shooters[19] and the accessibility of terrorist material on social media fuel radicalization, inspire copycat attacks and amplify violent extremist ideologies, posing significant security and societal risks.[20]

Recommendation feeds further create the potential to confine users to "echo chambers", hindering access to diverse sources of information. Nevertheless, exposure to different types of news sources is more likely on social media than in other types of media, and ranking algorithms do not have a significant effect on the ideological balance of news consumption on high-traffic websites such as Facebook or Google.[21] However, some algorithmic feeds prioritize posts with high engagement, which can highlight those posts that are more radical and emotionally charged, simply because they receive more engagement.[22] This phenomenon is especially dangerous in times of conflict or during elections.[23] GenAI, in particular, poses several risks to information integrity, specifically in terms of content creation – AI-generated deepfakes, "hallucinations"/inaccurate information,[24] rewriting of historical facts – and content distribution by perpetuating existing societal biases and amplifying discrimination. These risks undermine access to public interest information, which is the cornerstone of democratic societies.

1.3 How to build societal resilience to digital harms through MIL

MIL is increasingly being integrated into formal education by way of its inclusion in school curricula, with countries around the world developing or integrating MIL in primary, secondary or higher education. MIL is incorporated into school education through curricula designed to enhance students' critical thinking, communication and digital skills. Such a curriculum focuses on encouraging students to analyse, evaluate and produce information in various formats. This framework is embedded in subjects such as social studies, language and information and communication technology (ICT), while also fostering online safety, ethical media consumption and information verification skills.

It is also important to recognize that written literacy – the ability to read and write – cannot be universally assumed, even among groups typically expected to possess it. This reality underscores the need for media literacy interventions that go beyond written comprehension, embracing multimodal approaches that reflect the varied literacy levels and learning needs of diverse populations.

Beyond formal education, MIL curricula can be developed for:[25]

- Informal education that occurs at an individual level (e.g. self-study, massive open online courses [MOOCs], open educational resources [OER], accessing quality resources)
- Communities (e.g. study and reading groups, interactive public library sessions, community engagement workshops)
- Professional settings (e.g. workplace training and continued professional development [CPD]; tailored training for policy-makers, journalists, software developers, health workers, government officials, civil society organizations, electoral campaign managers)

The following implementation strategies can be used in the integration of MIL in formal and non-formal education:

Classroom learning

Integrating MIL into school curricula at different educational levels ensures that students develop critical thinking skills and the ability to navigate digital and information ecosystems effectively. MIL can be introduced as a standalone subject or embedded within existing disciplines such as social studies, languages, civics and science.

To enhance engagement and effectiveness, MIL instruction should go beyond theoretical discussions, incorporating interactive and experiential learning methods. This can include:

- **Discussion-based exercises:** Encouraging students to analyse real-world examples of misinformation, disinformation and media bias through debates, case studies and structured class discussions through open-ended questioning.
- **Practical assignments:** Tasking students with fact-checking exercises, media analysis reports and digital investigations. For example, students can compare how different news outlets cover the same event or track the spread of viral misinformation online.
- **Simulations and role-playing:** Designing activities where students take on roles as journalists, policy-makers or social media influencers to explore how information is created, disseminated and received. Red-teaming exercises can reinforce critical analysis skills.
- **Technology integration:** Using digital tools, online fact-checking resources and AI-driven media analysis platforms to help students verify sources, detect bias and assess credibility. Interactive apps and gamified learning experiences can make MIL more engaging.
- **Assessment and reflection**: Measuring learning outcomes through written reflections, multimedia presentations and assessments that require students to apply MIL concepts to real-life scenarios. Peer reviews and teacher feedback can help reinforce learning.

Standalone courses

MIL can be introduced as a mandatory course for students and teachers, either for academic credit or as a certification programme. Offered in offline, online or hybrid formats, these courses should be self-paced yet structured, using a blended pedagogical approach that combines reading materials with interactive multimedia resources such as videos, case studies and scenario-based exercises. Gamification elements and milestone certifications can further enhance engagement and retention.

MIL can also be fostered through the development of in-depth courses that are designed to equip different target audiences with specialized skills and knowledge. These courses should be tailored to address specific themes and challenges, ensuring relevance to various professional and societal contexts. For example, dedicated modules can focus on MIL as a tool for combating climate disinformation, equipping journalists, educators

and environmental advocates with strategies to counter false narratives. Similarly, a course designed for policy-makers can emphasize the role of MIL in shaping regulatory frameworks, promoting transparency and mitigating the spread of harmful misinformation in governance.

Workplace training/MIL training for professionals

Lifelong learning ensures that professionals across various sectors – ranging from software designers and content moderators to policy-makers, educators and journalists – can continuously develop the skills needed to navigate the evolving digital landscape. These training programmes should be designed to accommodate different career stages, allowing individuals to enhance their expertise in ways that are directly applicable to their professional responsibilities. By embedding MIL into workplace training and professional development initiatives, organizations can foster a workforce that is better equipped to identify, counter and mitigate the effects of mis- and disinformation.

Community engagement and outreach

MIL courses could also be offered through community engagement programmes such as rural adult education, urban-based literacy education, community media services, information campaigns ("pre-bunking")[26] and special programmes via national radio and television broadcasters. As an example, in 2020, UNESCO organized a virtual life skills training for more than 100 alternative learning system (ALS) implementers in the Philippines to enhance their MIL skills in the context of the COVID-19 pandemic. The training equipped participants with the skills to effectively navigate vast amounts of information and combat misinformation, covering topics such as online safety and critical thinking.[27] In addition, through "MIL cities", UNESCO also advocates for cities, municipalities, local government entities and various stakeholders, such as the transportation, entertainment and health industries, to empower residents with MIL competencies.[28]

Integration into the daily operation of youth organizations

As digital natives, young people are at the forefront of today's information ecosystems – both as consumers and creators of content. Their deep familiarity with digital platforms positions them as key actors in shaping the future of MIL. Consequently, youth-led organizations play a vital role in embedding MIL into non-formal education, peer-to-peer learning and community engagement. This integration goes beyond occasional workshops or campaigns; it involves embedding MIL principles into the very structures, practices and cultures of these organizations.

A notable example comes from 2020, when 27 youth organizations across south-eastern Europe and Turkey signed a Memorandum of Cooperation and Understanding to incorporate MIL into their daily operations.[29] This commitment included aligning internal policies with MIL values, training young members on critical thinking and fact-checking and using organizational platforms to combat disinformation and promote digital citizenship. Their example demonstrates how youth organizations can serve as incubators for grassroots MIL initiatives, fostering a generation that is not only media literate but also actively engaged in defending information integrity.

1.4 Opportunities, challenges and lessons learned

Implementing MIL programmes poses several systemic and practical challenges that call for multistakeholder solutions.

Commitment from the international community

There is a resounding commitment from the international community on the need and continued urgency to advance MIL, most recently reaffirmed in the UN Pact for the Future. However, differing interpretations of MIL's scope – encompassing digital education, media literacy and digital skills – along with the broad inclusion of various awareness-raising initiatives, have resulted in poor coordination, inconsistent monitoring and a lack of a cohesive strategy.

Fragmented MIL programmes

Fragmented efforts at monitoring and evaluating MIL programmes, especially in the long term, have reduced capacities to adapt MIL to new risks posed by technological developments and advancements in AI. It is crucial to evaluate how incorporating MIL in formal and non-formal education results in improved capacities to detect mis- and disinformation online, to enhance digital safety and to encourage participation in public discourse. Further, it is essential to set standards, ensuring quality control and adequate oversight of the operations of civil society actors and partner organizations implementing MIL initiatives.

National and regional MIL policies

It is vital that states develop national/regional MIL-related policies, elaborate strategies to sustain their efforts and allocate sufficient resources to implement MIL. According to a UNESCO Platform of Ibero-American Audiovisual Regulators (PRAI) study, 63% of audiovisual regulatory bodies in Ibero-America reported having a formal framework for action that outlines the importance, benefits and responsibilities related to MIL promotion and dissemination, while only 37% reported having resources or a budget earmarked specifically for actions on MIL.[30] In addition, the development and implementation of MIL policies must be carried out in coordination with key stakeholders, including regulatory authorities, media, civil society, fact-checking and youth-led organizations and educational institutions.

Adaptation and evolving initiatives

MIL initiatives must adapt and evolve as the information environment changes. A major barrier to successful MIL implementation is the lack of adequately skilled and trained educators, especially as digital platforms and media landscapes evolve at an unprecedented pace. Further, development of MIL strategies and curricula in the digital age needs to draw from expertise in interdisciplinary fields including education, journalism, psychology and behavioural science, law and policy, computer science and AI and cultural studies.

Beyond youth engagement

There is a growing need for MIL for adults, especially in the digital age. Given the high drop-out rates in lifelong learning contexts, short modules on MIL, with a focus on practical skills, need to be incorporated into professional training for institutional staff and duty-bearers. In the future, MIL should be a prerequisite for any job, ensuring that individuals in such roles are equipped to navigate, analyse and critically assess information effectively. This approach will promote informed, ethical and transparent decision-making across sectors.

Private-sector incentives

The private sector often lacks sufficient incentives or regulatory pressure to collaborate in MIL efforts, although many have made investments in MIL on their platforms. It is imperative to work with the private sector and technical community (especially companies working on AI research and development [R&D]) on designing and implementing MIL programmes. This will serve to promote transparency and accountability, which is necessary for empowering users on digital platforms.

2 Building a comprehensive model to strengthen information integrity

Understanding how disinformation spreads and the environments in which individuals engage with information is critical for creating an inclusive framework to reinforce information integrity.

The integrated model presented in this section combines the disinformation life cycle – capturing the stages from pre-creation and content production to distribution, consumption and post-consumption – with the socio-ecological model, which considers the layered influences on behaviour at the individual, interpersonal, community, institutional and policy levels. By aligning these two perspectives, the model enables a more strategic and systemic approach to digital safety and designing MIL interventions. It helps identify where current efforts may be concentrated or lacking, guides organizations in tailoring their strategies more effectively and underscores the importance of context-aware, multi-level responses to disinformation.

2.1 A holistic response to disinformation

While specific knowledge requirements for MIL may shift to deal with emerging threats, the core "scaffold" of competencies and attitudes tends to remain relevant across trends and over time. The greater challenge is ensuring that MIL is not narrowly conceived for one demographic (such as students or young people), and that strategies for provision are more creative about accessing groups beyond formal education (e.g. through employers, unions or faith-based institutions).

One goal of MIL is to help people source credible information and take greater responsibility for what they consume, where they consume it and how they choose to respond. It focuses on what can be taught, but must also consider what can be facilitated and encouraged within information ecosystems. At the individual level, this can be characterized as:

- **Competences (taught):** Understanding the nature and salience of disinformation in everyday life, with tools and skills to effectively analyse content regardless of its source.
- **Behaviours (facilitated):** Actively seeking out credible or evidence-based information and holding oneself and others to account in a constructive manner.
- **Attributes (encouraged):** Prioritizing nuance over simple binaries and seeing the benefit of more regulated, transparent and safe information spaces.

To tackle disinformation holistically, a wider range of socio-cultural factors becomes critical. These include, but are not limited to:

- Robust "taxonomies of harm" that evidence the impact disinformation can have on individuals, communities and societies at large
- Motivated champions for information integrity in the public sphere, including cultural and political figures across the ideological spectrum
- Universal access to fact-based and verifiable information sources, catering to a range of audiences both online and offline
- Clear legal definitions at the national and multilateral level – for example, where disinformation intersects with hate speech, incitement to violence or other criminal activity (e.g. electoral interference)
- Transparent standards and regulatory oversight for legacy media (press, radio, broadcast)
- Campaigns that explain and celebrate the process for producing credible journalism and scientific data
- Mechanisms to report harmful disinformation that are clear, fair and properly enforced (e.g. by tech companies or law enforcement)

Crucially, education alone will never be enough. Even if progress were achieved in all the areas listed, this would likely be eclipsed by a more systemic and pressing need: to build healthy incentives into the information ecosystem at all levels.

3 A socio-ecological model to tackle disinformation

Combating disinformation requires a whole-of-society approach, not just a focus on individuals and content.

To date, efforts to tackle disinformation have generally focused on consumers and content. The former includes a spectrum of MIL interventions, as outlined in Section 1 of this report, aiming to change individual habits and mitigate harms that may occur when disinformation is encountered "in the wild". The latter attempts to counter falsehoods via an arsenal of reactive fact-checks and debunking mechanisms, alongside ongoing research into the narrative and tactical playbooks used by malign actors. Monitoring may help surface threats before they reach a critical mass of exposure or engagement, while analysis can reveal the common techniques of disinformers and their apparent motivations (commercial, political, ideological, personal and so on).

3.1 Towards a whole-of-society strategy

These two strategies, monitoring and analysis, should operate on a feedback loop – exposing the anatomy of deception can help strengthen education and bolster overall resilience to such methods over time. However, they also risk underplaying the systems that enable disinformation or the malign incentives therein. Both routes have arguably placed too much onus on individuals and civil society while obscuring the larger structures at play. This includes the role of social media platforms, publishers, advertisers, digital service providers, governments and others in shaping our ecosystems at a fundamental level. If accountability is the goal, the burden cannot be shouldered only by those seeking information but must also include the intermediaries, hosts and brokers of information online and offline.

The need for a more holistic view is clearly defined in the UN's Global Principles for Information Integrity,[31] published in July 2024 and expanding on the *Our Common Agenda*[32] report launched for the body's 75th anniversary. These agendas look to reframe the public debate and articulate how policy can address the behaviours and systems that drive disinformation at scale. They also stress that "strengthening the good" cannot be the sole priority if recent trends are to be reversed; it is also critical to "weaken the bad" and create an environment where information integrity is both viable and self-sustaining.

Research has continually shown that disinformation outpaces facts or evidence, since this content helps drive the attention economy and systems optimized for engagement. As such, interventions that make wilful deceit harder to produce, riskier and less profitable are key. The playing field must be levelled so that credible information has a fighting chance. And, above all, disinformation must be undermined at every stage, from its design and distribution to its consumption and impact.

The information resilience mapping model in Figure 1 has been developed to support work under the Global Coalition for Digital Safety. It aims to strengthen a whole-of-society response by:

- **Unpacking the life cycle of disinformation (x-axis).** The model outlines all stages where disinformation could be discouraged, intercepted, weakened, challenged and countered at scale. This includes potential points of entry before content enters the public domain or starts gaining traction as opposed to a purely reactive approach. Debate and interventions tend to focus on the tail end of this pipeline, by which point mitigation is often more costly and must contend with a wider array of factors.

- **Identifying all stakeholders necessary for response (y-axis).** The model explores the role played by different groups and how their specific contributions interrelate. These are not intended

as a hierarchy, but rather to show how vital levers of influence exist and should be deployed at the individual, peer-to-peer, institutional and regulatory levels. It encourages dialogue with, and support for, a greater diversity of actors in the fight against disinformation and a recognition of how trust and power are distributed from the grassroots through to government.

– **Tackling supply- and demand-side dynamics.** The model considers the problem of disinformation through both supply and demand: those who deliberately create or amplify such content and those who consume it. It recognizes that consumption itself can be wilful or unwitting, and how interventions might effectively tackle each scenario. Disinformation can serve to convince people of something false, but it may also serve as a pretext for their existing beliefs or biases – a pattern clearly observed when disinformers offer more convenient stories to explain complex (and often distressing) realities. The model also addresses the "marketplaces" for ideas and how they can motivate, reward or drive certain behaviours and activity while prejudicing others.

– **Mapping current interventions and where gaps persist.** The model allows anyone engaged in countering disinformation to contextualize their efforts and frame objectives within a greater whole. This can help in shaping a unified theory of change and forging more effective links between projects and policies. In the process, it should also serve to expose areas where greater resource is urgently needed, whether financial or human.

FIGURE 1 **The information resilience mapping model**

Source: World Economic Forum.

4 The disinformation life cycle

Disinformation does not exist in a vacuum, and counter-efforts must be initiated before it achieves mass engagement or causes observable harm.

Tackling this issue at scale requires a broader view of the supply chain: what drives deceptive behaviour, which tools are used to produce and proliferate such content and how does it enter the mainstream in public life? The model considers five key stages in the disinformation life cycle from a "supply" and "demand" and marketplace perspective: pre-creation, content creation, distribution, consumption and post-consumption.

4.1 Pre-creation (initiation or impulse)

This phase largely relates to cultural norms, awareness and perceptions of accountability at the individual and collective levels. These factors create the soil in which disinformation will either thrive or struggle to take root, independent of any specific trend. Interventions must address psychosocial dynamics and barriers to accessing credible information across demographics and contexts. While critical thinking and rational analysis are foundational to media literacy, they are not sufficient on their own. Emotions, intuition and social identity play a powerful role in shaping how individuals engage with information. Also important is the role information plays in people's lives and how this is shaped by their experiences, including the theme of inequity. The core objective is to build resilience against disinformation in all its guises, embedding the means to navigate our information space in more deliberative and discerning ways.

Supply

Increase accountability for those who engage in the spread of disinformation. Highlight the downsides of such activity; for example, via campaigns showing how prior offenders incurred legal, financial, professional or personal costs. Profile where disinformation networks have been successfully exposed and held to account. In parallel, humanize the impact of disinformation through victim testimony.

Demand

Increase literacy about how information ecosystems function, including the business model for online platforms and the potential benefits or pitfalls of a "personalized web". Generate interest in fact-based journalism by showing how stories are produced, including the ethical standards and oversight applied to legacy media. In tandem, legacy media should strengthen partnerships with the influencers who have become primary conveyers of news to online audiences. Support people to explore their pre-existing biases and how these may be exploited by disinformers. Expose the mechanics and common traits of disinformation and bring to life its concrete harms.

Marketplace

Ensure credible information is available on an ongoing basis, with channels tailored to reach people of all backgrounds and profiles. Invest in local journalism, including stronger platforms for marginalized and underrepresented voices. Prevent predatory and/or micro-targeting of users in the online space, for example via the sale of personal data to advertisers. Support the development of "digital public squares", from social media to forums, in which high-trust information breaks through and systems are better geared for constructive discovery, debate and learning.

4.2 Content creation (production)

This phase considers the tools used to generate disinformation and when such activity is likely to surge. For the former, interventions should assess how easy it is to produce high-traction content and where barriers to entry could be raised. This question will become increasingly urgent as GenAI products are released on the mass market, with their potential to turbo-charge the scale, affordability, agility, targeting and persuasiveness of disinformation campaigns. For the latter, the focus should be on building collective readiness and ability to anticipate trends, as well as analysing where and why disinformation is proving effective.

Supply

Make it more expensive and labour-intensive to produce disinformation at scale. This may include stronger guardrails on GenAI tools such as text and image generators, as well as mixed media "deepfake" technology. Strengthen legal frameworks around copyright infringement (for example, the logo of a known media outlet) as well as non-consensual use or impersonation of someone's image, voice or identity. Platforms should take stronger action against for-profit human content farms and implement stricter recidivism strategies to prevent disinformation networks from rebuilding after removal.

Demand

Expose the common features of low-quality or low-trust information, including clickbait, content farms, propaganda, advertising and synthetic or manipulated media. Raise awareness of these red flags and champion signals of information integrity (e.g. clearly cited data and images). Embed access to tools that can support critical thinking, lateral reading and verification of sources in real time.

Marketplace

Mandate rigorous testing of GenAI tools and services before they enter the market, even in open-source models where decentralization can make enforcement challenging. This may include "red team" exercises that simulate the tactics of disinformers, helping to identify vulnerabilities and design more effective guardrails. Given that open-source AI can be modified and deployed by various actors, testing should occur not just before release but also through ongoing scrutiny and adaptation to emerging threats. Ensure transparent risk assessments that balance the intended value of a product (e.g. entertainment, efficiency, innovation, learning) with its potential misuse and the scale of related harm (e.g. automated or dangerous disinformation in response to a given prompt).

4.3 Distribution (dissemination and promotion)

This phase focuses on the role of platforms in the spread and mainstreaming of disinformation. It examines the structural factors that determine how content is surfaced, curated and amplified to consumers as well as the broader effects on public discourse. Transparency in these processes is critical – users, creators and researchers need greater visibility into how moderation decisions are made, how recommendation algorithms function and how content is promoted or suppressed. Without such transparency, public trust in platform governance erodes, fuelling perceptions of bias and allowing disinformers to exploit opaque systems to their advantage.

Online approaches should avoid relying too narrowly on content removal or post-level moderation – except in cases of acute and immediate harm. In some instances, what constitutes illegality can vary significantly across jurisdictions, and in some cases disinformation laws are deliberately crafted or applied in ways that suppress opposition speech and restrict legitimate political expression. However, content removal and moderation are necessary and proven tools that reduce the amount and visibility of harmful content on platforms. Instead, a broader review of the platform mechanisms that drive virality – such as algorithmic amplification, engagement-based ranking and advertising models – is needed. Platforms must be more forthcoming about how these systems operate, the criteria used to boost or demote content and the safeguards in place to prevent manipulation.

Any safeguards developed must distinguish between freedom of speech and freedom of reach, taking a human rights-based approach to both. This means preserving fundamental rights while ensuring that disinformers cannot abuse opaque platform systems to generate profit, spread falsehoods or undermine democratic discourse.

Supply

Ensure transparent processes for registering website domains, creating accounts on social media and administering in-platform groups or channels, with greater oversight of group admins, content creators and advertisers who may amplify misinformation unknowingly. Platforms should provide clearer disclosures on who is behind influential pages, groups and paid promotions, enabling users to assess credibility and accountability.

Enforce proportionate, clear and consistent action against “super-spreaders” of disinformation, including coordinated networks that operate within and between platforms. This should extend to advertisers as well as group admins.

Introduce tools that control how rapidly content can be shared, such as forwarding or tagging limits, while ensuring that these measures are applied transparently and equitably. Platforms must clearly communicate how these restrictions are implemented, who they apply to and how they contribute to reducing the spread of harmful content. Additionally, incorporate nudges – such as prompts encouraging users to verify information before sharing, or notifying them when they are about to engage with content flagged as misleading – that have proven particularly effective in slowing the spread of disinformation and fostering more thoughtful engagement.

Demand

Elevate trustworthy sources of information, including through partnerships between emergency responders and digital platforms. Develop campaigns that “inoculate” the public against persistent disinformation by exposing the tactics and motives behind misleading content. Strengthen public outreach by leveraging force-multipliers[33] and trusted actors – such as health and social workers, religious leaders, employers and trade unions – who can engage communities directly.

Additionally, partner with influencers and content creators who shape online discourse and drive engagement, ensuring that accurate information reaches audiences where they naturally consume news. These partnerships can help counter disinformation in a more organic and relatable way, fostering trust and improving the visibility of credible sources across different digital spaces.

Marketplace

Introduce “circuit breakers” for moments when disinformation surges, particularly around elections, conflicts, terrorist incidents, natural disasters and public health emergencies. Strengthen protocols that allow time for human review and verification before false claims can reach a critical mass. Platforms must also improve real-time transparency on the actions they take during crises – clearly communicating what content is being removed, downranked or labelled, and explaining the rationale behind these decisions to maintain public trust.

Label known disinformation using tools such as hashing, enabling quicker identification and mitigation as content spreads across and between platforms. Additionally, build stronger forums for intelligence-sharing between service providers, particularly when recurrent issues or tactics are detected – such as website spoofing, astroturfing networks,[34] coordinated inauthentic behaviour and identity fraud – ensuring a more unified response to emerging threats.

4.4 Consumption (engagement)

This stage considers people’s relationships to information in everyday life, including how healthier habits can be enabled or encouraged. It explores ways to encourage critical engagement with content and reduce susceptibility to disinformation. Efforts include applying fact-checks and warning labels to debunk falsehoods and enhancing peer-to-peer moderation for greater accountability.

Supply

Increase investment in local and citizen journalism that adheres to strong ethical standards, ensuring that high-quality information remains accessible and not prohibitively expensive or difficult to find. Rather than attempting to mimic rapidly shifting digital trends, focus on meeting key audiences where they already convene online, leveraging mixed media formats and diverse languages.

Support underserved communities in telling their own stories by providing training and resources that enhance their ability to produce credible content. Additionally, prioritize partnerships with influencers who have already built significant followings in the news and information space. Providing these creators with targeted media and digital literacy training can strengthen their role as ambassadors for information integrity and equip them to responsibly engage their audiences or counter disinformation in a trusted, authentic way.

Demand

Apply appropriate warning labels and fact-checks to disinformation online, as well as accounts repeatedly found to share such content. Increase accountability through tools for peer-to-peer moderation and education. For crisis events or topics most vulnerable to attack, create hubs of fact-based information and link these to relevant search terms. Educate consumers about phenomena such as echo chambers, which may narrow their understanding of the world or falsely imply that a viewpoint has universal consensus.

Marketplace

Strengthen in-platform safety features and user controls – for example, to limit screen time or order newsfeeds based on chronology. Allow consumers to withdraw consent for targeted advertising or the sale of their personal data to third parties. Ensure recommender algorithms do not produce a "rabbit hole" effect, whereby engaging with disinformation funnels users towards more content of this nature.

As sources of information – including local news outlets and independent journalism – continue to shrink or disappear, the information ecosystem grows more vulnerable as, in a market, competition fuels better products. This erosion not only weakens the marketplace of ideas but also undermines the context and credibility needed for media literacy efforts to succeed. Recognizing this reality, media literacy interventions must adapt by equipping users to critically evaluate the information landscape as it is, while also supporting efforts to revitalize or reimagine public-interest information infrastructure in both digital and offline spaces.

4.5 Post-consumption (influence and impact)

The final stage addresses how disinformation can affect individuals, groups and society at large. Disinformers are adaptive and will continue to exploit loopholes in policy or prevention efforts. It is important therefore to plan for scenarios where such content continues to spread or take up oxygen in public discourse. However, the existence or reach of disinformation should not be mistaken for a set of foregone outcomes. Interventions must focus on creating a stronger feedback loop, so that learnings from previous episodes can inform mitigation and response. This includes measures to undermine disinformation already in circulation and safeguard those affected.

Supply

Publish profiles on the actors behind disinformation campaigns, including their known tactics and suspected motivations. Invest in public databases of fact-checks and debunking mechanisms, alongside enhanced tools for reporting and "trusted flagger" schemes. Develop stronger legal frameworks that tackle disinformation in a fair and proportionate manner, grounded in human rights and a nuanced assessment of harm. Improve support and redress mechanisms for those victimized by disinformation, including digital safety and security training. Arm high-trust communicators with the knowledge and resources to compete in a saturated information space.

Demand

Conduct in-depth research to assess how people encounter disinformation in their everyday lives and its corresponding effects (e.g. on perceptions, attitudes and behaviours). Produce robust studies that quantify how such content can cause observable harm (e.g. mobilization to violence, threats to public health, disruption to emergency response or democratic processes). Adapt education materials and provision in line with these trends, in particular the emergence of new technologies.

Marketplace

Iterate both upstream policies (to disincentivize bad actors) and downstream protocols (to triage and respond to crises in a timely manner). This could include strengthened penalties for repeat offender accounts, changes to terms of service or platform functions, improved user controls or partnerships with high-trust media and other expert institutions.

4.6 Policy approaches to tackling disinformation

Since the early 2020s, governments around the world have increasingly turned to regulatory frameworks to tackle online disinformation while preserving fundamental rights. Although not without flaws, these efforts demonstrate how targeted interventions can disrupt the disinformation life cycle.

Australia and the United Kingdom have embedded "safety by design" principles into legislation, encouraging platforms to build user safety into their core infrastructure via different mechanisms. The UK's Online Safety Act 2023 (OSA) mandates proactive risk planning in platform development,[35] while Australia's Online Safety Bill 2021 promotes these standards through voluntary guidance.[36] This shift embeds harm prevention into the architecture of digital services.

The European Union's Digital Services Act (DSA) 2022 establishes a harmonized legal framework for digital services across the EU, with the overarching aim of creating a safer, more transparent and rights-respecting online environment for all users.[37] The DSA places binding obligations on very large online platforms (VLOPs) and very large online search engines (VLOSEs) – together referred to as VLOPSEs – to identify and mitigate a range of online harms – such as illegal content, disinformation and threats to public health or democratic processes – while upholding key fundamental rights, including freedom of expression and access to information.

The DSA addresses disinformation not by regulating content directly but by requiring platforms to assess and mitigate systemic risks linked to the design and functioning of their services, aiming to intercept disinformation at the pre-creation, creation and distribution points of the life cycle. VLOPSEs must evaluate how their recommender systems, monetization models and content moderation practices may facilitate the spread of both illegal content (as defined by national law) and legal but harmful content, such as health misinformation, coordinated harassment or falsehoods that undermine electoral processes or civic discourse. Regarding post-creation, VLOPSEs are required to demonstrate how they mitigate risks identified on their services, which could include deploying friction mechanisms, warning labels and source disclosures to reduce the impact of disinformation after exposure and to support more informed user decision-making.

Regulations in both the EU and the UK also strengthen platform accountability at the point of content consumption. The UK's OSA addresses disinformation through a layered approach that combines platform regulation with public empowerment – most notably via its strengthened media literacy duty. While the OSA does not directly regulate disinformation as a category, it places legal obligations on platforms to assess and mitigate risks from illegal content and content harmful to adults and empowers the regulator Ofcom to supervise compliance.

The DSA is still in early implementation stage but has begun reshaping platform behaviour by formalizing systemic risk governance and increasing scrutiny of recommender systems, content ranking and advertising transparency. VLOPSEs published their first transparency reports in early 2023, followed by systemic risk assessments submitted to the European Commission in August 2023. Independent audits were submitted in August 2024.[38] The DSA Transparency Database enables public access to content removal decisions and regulatory notices, strengthening civil society oversight.[39] While early implementation has driven improvements such as clearer content labelling and user control over personalization, challenges remain. The DSA enhances conditions for MIL by demystifying platform systems and enabling evidence-informed engagement by educators, researchers and users.

The UK's OSA complements this approach through a statutory media literacy duty. Ofcom's Media Literacy Strategy 2024–2027[40] is being implemented through research, pilot programmes and civil society partnerships aiming to improve public understanding of online harms and promote safer digital participation. This is bolstered by the UK Department for Science, Innovation and Technology's (DSIT) strategic priority principles on safety by design,[41] which guide platforms to proactively embed user protection and disinformation mitigation into service architecture. Together, these initiatives reflect a shift towards integrated regulatory ecosystems that support both systemic accountability and user empowerment.

5 A socio-ecological model

The strength of these models lies in their ability to highlight relationships between influencing factors and how they might work with or against each other.

This type of framework originated during the 1970s in developmental psychology in the work of Urie Bronfenbrenner, who created a model to organize the range of factors that affect an individual child's development. For example, children might vary in their genetics, talents and experiences, which, in turn, are influenced by their immediate environment (such as their families, peers and schooling), which is influenced by mass media, social services and local politics and which are ultimately influenced by the wider culture.

This nested organization acknowledges how the relationships between factors can act to amplify or dampen overall impact. An individual child may have an affinity for mathematics, but if they live in a family or go to a school where that is neither valued nor supported, this affinity may never grow into skill. However, if that same child lives in a culture with television shows that demonstrate the importance of mathematics and treat it as an impressive and desirable skill, or has a relative who plays a similar role, they may indeed find ways to develop in spite of other constraints. The range of factors that affect child development are manifold, but by using a socio-ecological lens it is possible to better organize these relationships and identify where intervention might be fruitful. The same is true for digital safety and the fight to embed, maintain and enhance information integrity.

SEMs have since been used in many different contexts to understand and organize influencing factors in complex, interconnected problem areas. In public health, SEMs have helped map intervention landscapes, such as the roles that social and economic conditions, community advocacy, corporate policies, education and individual choice all play in the adoption of healthy behaviours. SEMs have also been used in risk communication to assess the relationships between culture, education and timing in how an individual might best receive the information necessary to make impactful decisions through different channels.

Applying this lens to the landscape of counter-disinformation efforts helps establish where current investment is concentrated and where gaps in activity or attention persist. The addition of a timeline axis (the disinformation life cycle) helps illustrate how dynamics evolve throughout the process of creation and consumption.

This model helps configure the existing landscape, both for MIL interventions and other whole-of-society approaches to combat disinformation. The levels of the SEM are individual, interpersonal, community, institutional and policy.

5.1 Individual

SEMs often begin with the individual as the central focus, emphasizing the range of factors that shape their behaviours, preferences and development. These factors can include inherent abilities, personal affinities, psychological traits and learned skills. In the context of MIL, such considerations play a critical role in shaping how a person interacts with information. This includes the ability to identify and understand and interrogate their own biases and those present in external media, which requires critical thinking skills and awareness of different perspectives.

Emotional literacy also forms a significant component, as individuals must navigate how emotions influence their reactions to information, especially when encountering disinformation designed to provoke strong responses. Additionally, platform usage is a key individual concern, as the decision to share or withhold content on digital platforms contributes to the broader dissemination of information. By focusing on these individual elements, it is possible to better design interventions that empower people with the skills necessary to critically engage with media and resist the pull of disinformation.

5.2 Interpersonal

Some of the most direct influences on a person's experience stem from their close relationships, particularly within families, among caregivers and through interactions with peers. These relationships play a foundational role in shaping perceptions of what is considered normal and desirable behaviour, including how individuals engage with information and media. Trust within these relationships amplifies their impact, as advice, opinions and examples from loved ones are often internalized.

For MIL, interventions at the interpersonal level could leverage these dynamics to counter disinformation effectively. Social pressure within peer groups can discourage the sharing of false or misleading information, creating a culture where accuracy and verification are valued. Parents and caregivers can play a critical role by setting expectations for the responsible use of social media and actively guiding children in recognizing and addressing disinformation they encounter. This could include activities such as co-reviewing content, discussing sources and modelling critical questioning. Additionally, community-based workshops or campaigns can provide families and friends with tools to engage in constructive dialogue about media habits, fostering a shared commitment to information integrity in everyday interactions.

5.3 Community

Communities play a pivotal role in shaping norms, behaviours and attitudes, providing individuals with a sense of belonging and shared identity. These communities can take many forms, ranging from religious groups and local organizations to cultural associations, workplace networks or online forums dedicated to shared interests. Within these spaces, members often look to one another for guidance, reinforcement of beliefs and cues about acceptable behaviours, including how they consume and share information.

The role of community in MIL is especially critical because disinformation often spreads through trusted interpersonal and communal networks, where it can quickly gain traction. Some communities are disproportionately targeted or impacted by disinformation due to factors such as social vulnerabilities, language barriers or historical marginalization, making tailored interventions even more essential. Community norms, whether formalized or informal, significantly influence how disinformation is perceived and propagated, as members may feel pressure to align with the group's prevailing narratives or actions.

To address these dynamics, MIL initiatives should actively engage communities, empowering them to recognize and resist disinformation while reinforcing positive norms around information sharing. This can include community-driven workshops, partnerships with local leaders or influencers and the creation of resources tailored to specific cultural or contextual needs. By leveraging the trust and cohesion within communities, such efforts can foster a collective resilience to disinformation and encourage healthier, more informed interactions with media.

5.4 Institutional

Institutions are foundational components of social structures, shaping daily experiences and influencing behaviour on a broad scale. From workplaces and educational systems to healthcare providers, governments and platforms, institutions play a critical role in establishing norms, providing resources and enforcing standards. The trust people place in institutions gives these entities significant authority to influence perceptions, decisions and habits, making them key players in efforts to combat disinformation.

In the context of MIL, institutional-level interventions offer unique opportunities to address disinformation systematically. For non-platform institutions, such as employers or schools, interventions could include integrating MIL training into standard onboarding or professional development programmes, much like cybersecurity or privacy training. These programmes could focus on teaching employees or students how to critically evaluate information, identify disinformation and practise responsible sharing of content. Governments and healthcare institutions, as trusted sources of authority, could contribute by disseminating verified information and promoting public awareness campaigns to counter disinformation. However, if trust in institutions is lacking, the information they disseminate may fail to resonate with the public – and in some cases may even undermine efforts to counter disinformation.

Technology platforms, as central nodes of the information ecosystem, hold specific responsibilities in this landscape. Platforms can develop and refine algorithmic tools to flag or limit the reach of disinformation, including features such as content warnings, reduced visibility for harmful posts and stricter content moderation policies. They could also foster partnerships with fact-checkers, prioritize credible sources in search results and increase transparency around their moderation processes. Non-platform institutions can complement these efforts by advocating for regulatory standards that encourage accountability and ensure platforms take a proactive role in combating disinformation.

By aligning the efforts of platform and non-platform institutions, a comprehensive and multifaceted approach can be developed to foster resilience against disinformation. Leveraging their collective influence and resources, institutions can help build an information ecosystem where truth prevails and harmful narratives are systematically dismantled.

5.5 Policy

The policy landscape serves as the overarching framework shaping the environment in which individuals, communities and corporations operate. By establishing rules, incentives and priorities, policies influence the choices available at every level of society and define the parameters within which actions are taken. Effective policy interventions address structural and systemic factors, creating conditions that support better decision-making and equitable outcomes.

Policies play a critical role in shaping the broader context for combating disinformation. At the legislative level, this could involve implementing regulations that create incentives for platforms to prioritize accuracy, transparency and accountability in their content moderation and recommender systems. Policies could mandate greater oversight of algorithmic processes or require platforms to publicly report on their efforts to combat disinformation.

Budgetary support is another essential aspect of policy intervention. Governments can ensure that MIL initiatives are adequately resourced, funding educational programmes, public awareness campaigns and the development of tools to improve digital literacy. These investments can also extend to underserved communities, helping to close gaps in access to reliable information and critical-thinking skills.

Moreover, policy can drive progress by funding and prioritizing research into MIL best practices. This includes identifying effective interventions, understanding disinformation's psychological and social dynamics and exploring innovative approaches to equip individuals and communities with the tools to navigate the digital information landscape.

At the same time, policies can harm issues of freedom of expression and access to information. Legislative efforts in some jurisdictions – such as proposals to restrict discussion of structural social factors in education – pose a significant challenge to the open, critical inquiry that media literacy requires. These developments signal a broader trend that could constrain the scope and impact of media literacy programmes worldwide. In navigating these pressures, it becomes even more urgent to defend academic freedom, support educators and design interventions that are adaptable, transparent and rooted in democratic values.

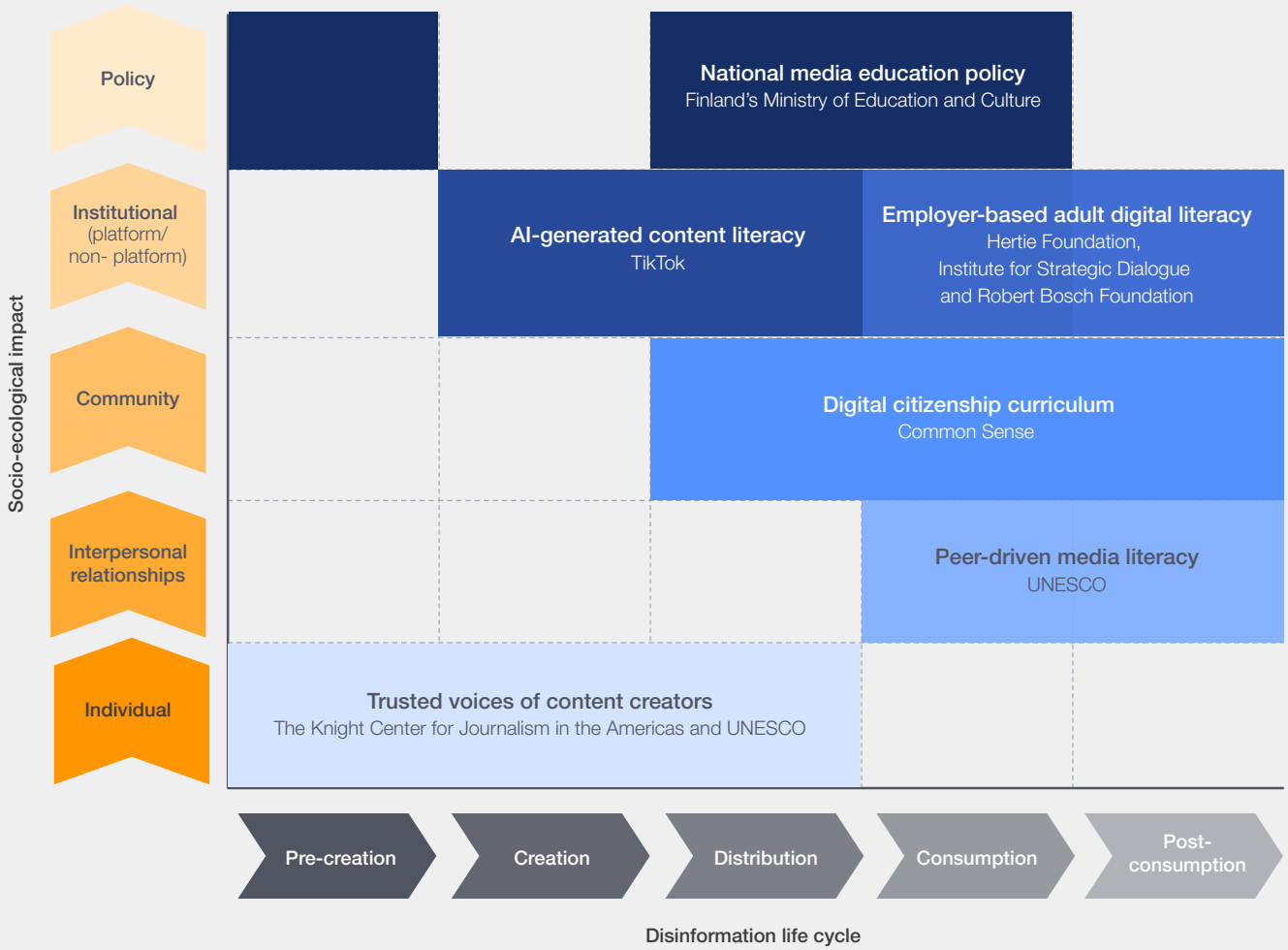
Ultimately, policy interventions must operate at a macro level, aligning stakeholders across sectors and ensuring that MIL efforts are both sustainable and scalable. By embedding MIL within the broader policy framework, governments and institutions can foster an information environment where truth and integrity are safeguarded and disinformation is systematically diminished.

6 Mapping MIL interventions

Case studies can demonstrate how a whole-of-society approach is the most effective response to disinformation.

The case studies used in this report and mapped in Figure 2 play a crucial role in illustrating the diverse ways in which interventions can address disinformation across multiple socio-ecological levels and stages of the disinformation life cycle. Through real-world examples, policy-makers, educators and digital safety practitioners can better understand how interventions operate, as well as identify where there is a dearth of activity targeting a certain area. Case studies provide tangible evidence of how these levels intersect, demonstrating that an effective disinformation response often requires a multi-tiered approach. Ultimately, they serve as a blueprint for best practices, guiding future initiatives in refining and scaling across different social, technological and regulatory contexts.

FIGURE 2 **The information resilience mapping model: case studies**



Source: World Economic Forum.

6.1 Trusted voices of content creators

The Knight Center for Journalism in the Americas, in partnership with UNESCO, has launched an initiative exploring the evolving role of influencers and digital content creators in the news media industry. As the lines between journalism and content creation continue to blur, this project seeks to bridge the gap between the two fields, fostering a shared commitment to credibility, fact-checking and ethical storytelling.

A UNESCO survey of digital content creators revealed that 62% do not carry out rigorous and systematic fact-checking of information prior to sharing it; however, 73% wish to be trained to do so.[42]

This multifaceted effort culminated in the launch of a free, multilingual online course: Digital Content Creators and Journalists: How to Be a Trusted Voice Online. The course provides journalists and digital creators with the tools to navigate the challenges of the digital era, equipping them with essential skills in media literacy, disinformation detection and audience engagement. The initiative is further supported by the publication of an open-access e-book, *Content Creators and Journalists: Redefining News and Credibility in the Digital Age*, offering insights from media professionals worldwide.[43]

Socio-ecological level

The intervention operates at the individual level of the socio-ecological model, focusing on enhancing the skills, knowledge and ethical awareness of journalists and content creators. As digital platforms become primary sources of information, individuals play a crucial role in shaping public discourse and influencing news consumption. By equipping participants with fact-checking techniques, ethical storytelling practices and media literacy skills, the initiative empowers them to navigate the digital landscape responsibly and combat the spread of misinformation.

Through the free, multilingual online course, participants develop the ability to critically assess information, verify sources and understand the impact of their content. The initiative also fosters self-reflection and professional growth, encouraging journalists and content creators to adopt best practices in credibility and audience engagement. Given this shift, it is essential that content creators and journalists possess the individual capacity to uphold ethical standards and counter disinformation. In turn, they can become force-multipliers and shift the cultural norms for their (online) audience, extending the impact of such a programme from the individual to community level.

Disinformation life cycle level

Due to the initiative's focus on digital content creators and journalists, it operates at the pre-creation, creation and distribution stages of the life cycle. By equipping journalists and digital content creators with the skills necessary to produce credible, ethical and fact-based content, the initiative helps shape the integrity of information before it is created and shared.

At the pre-creation stage, the initiative fosters critical awareness and ethical responsibility among its participants. Through the free, multilingual online course and collaborative discussions at the Knight Center roundtable, they can develop a deeper understanding of misinformation risks, media credibility and the societal impact of their content. This stage focuses on preparing individuals to recognize bias, question sources and apply fact-checking techniques before producing content.

During the creation stage, the initiative strengthens responsible storytelling and digital journalism. The programme emphasizes best practices in media ethics, fact-checking and audience engagement, ensuring that participants are equipped to produce high-quality, accurate content. Determining what information to share is increasingly important for content creators, as 42% of respondents said they used "the number of 'likes' and 'shares' a post had received" on social media as the main indicator for assessing the credibility of information.[44] Content Creators and Journalists: Redefining News and Credibility in the Digital Age compiles insights from industry professionals, highlighting how journalistic integrity and digital content creation can coexist in an evolving media landscape.

At the distribution stage, the initiative addresses how content reaches audiences and combats the spread of misinformation and harmful narratives. By encouraging responsible content dissemination and algorithmic awareness, participants learn how platform mechanics, engagement metrics and virality influence information flow.

Outcomes

The Digital Content Creators and Journalists: How to Be a Trusted Voice Online course has made a significant impact in strengthening the credibility and ethical standards of digital content creation. Participants reported increased confidence in applying fact-checking techniques, verifying sources and producing responsible media content, leading to a more informed approach to journalism and digital storytelling. The initiative's free online course, with more than 9,000 participants from 172 countries, has

expanded access to essential media literacy training, equipping a global network of content creators and journalists with the tools to combat misinformation.[45]

Beyond individual skill-building, the initiative has fostered collaboration between journalists and digital influencers, encouraging cross-industry learning on best practices for credibility and audience engagement. The roundtable discussions and sessions at the International Symposium on Online Journalism (ISOJ) created a space for critical reflection on the evolving role of influencers in news distribution, leading to greater awareness of ethical storytelling and media accountability. Content Creators and Journalists: Redefining News and Credibility in the Digital Age further solidifies these insights, providing an enduring resource for media professionals worldwide.

This initiative ensures that journalists and digital creators are equipped to navigate the challenges of the digital age, especially as content creators increasingly become a primary source of news and information. By fostering ethical storytelling, fact-checking and media literacy, the programme empowers participants to produce credible content in an era where audiences rely more on digital influencers for news consumption.

6.2 Peer-driven media literacy

The Think Critically, Click Wisely! initiative is part of UNESCO's broader effort to promote MIL. It emphasizes the importance of access to reliable and fact-based information for making informed decisions and participating in society. This initiative recognizes that while new technologies offer opportunities for education, debate and self-expression, they also present challenges such as mis-and disinformation, hate speech and privacy infringements.

The initiative is designed to leverage social relationships in countering disinformation at the consumption and post-consumption stages of the disinformation life cycle. Recognizing that false information often spreads through personal networks – family, friends, colleagues and online communities – this intervention emphasizes peer-based learning and trust-driven correction mechanisms. By fostering interpersonal discussions on misinformation, equipping individuals with effective corrective techniques and promoting responsible digital engagement, it aims to create a socially reinforced culture of media literacy.

Implemented across various social settings – including schools, workplaces and community organizations – the intervention is grounded in research on digital literacy's role in shaping social norms. Studies indicate that misinformation is more likely to be believed and shared when reinforced by personal connections. The initiative therefore enhances individuals' ability to evaluate information collaboratively and intervene against disinformation within their networks.

Socio-ecological level

This intervention operates at the interpersonal level, recognizing that individuals rarely consume or engage with digital content in isolation. Social relationships play a pivotal role in shaping how information is processed, shared and reinforced. By equipping individuals with the skills to critically engage with misinformation in everyday conversations, the initiative helps disrupt the spread of false narratives within families, peer groups and professional networks.

Participants trained in media literacy often serve as informal educators within their social circles, guiding friends and family members in identifying misleading content. Research has shown that misinformation correction is most effective when it comes from a trusted source rather than an external authority. This intervention encourages conversational corrective techniques, teaching individuals how to challenge false information in ways that minimize defensiveness and increase receptivity.

Additionally, the initiative fosters intergenerational learning, as younger digital natives frequently assist older relatives in navigating online misinformation. This was evident in evaluations where participants reported helping family members identify deceptive content related to health, elections and financial scams. Beyond individual actions, community-driven fact-checking initiatives – such as neighbourhood digital literacy groups and workplace media awareness programmes – further reinforce social accountability in information consumption.

Disinformation life cycle level

The initiative strategically intervenes at the consumption and post-consumption stages of the disinformation life cycle, equipping individuals with the necessary skills to critically engage with misleading content and take corrective action within their social networks. By addressing both how people initially encounter and interpret false information and how they respond after exposure, the initiative fosters a proactive and socially embedded approach to combating disinformation.

At the consumption stage, individuals are often exposed to misleading or deceptive content through social media, messaging apps and interpersonal discussions. Research indicates that people tend to trust and accept information shared by friends, family and colleagues, often without questioning its accuracy. To counteract this, the initiative focuses on developing critical consumption habits by teaching participants to recognize disinformation tactics, verify sources and cross-check information and analyse how algorithms, engagement metrics and virality shape the spread of false information.

At the post-consumption stage, misinformation continues to exert influence, even after individuals realize it is false. This is due to the lingering effects of initial exposure, the reinforcement of falsehoods within social circles and the psychological barriers that prevent individuals from correcting misinformation. To address this, the initiative encourages participants to adopt corrective behaviours that disrupt the continued spread of disinformation. This includes initiating constructive conversations, sharing reliable sources within peer groups and establishing digital safety norms. By shifting the expectation from passive misinformation exposure to active intervention, the initiative empowers individuals to take responsibility for the accuracy of information circulating within their networks.

Outcomes

Think Critically, Click Wisely! has demonstrated significant success in disrupting the social spread of misinformation. Participants reported increased confidence in engaging peers and family members in discussions about digital literacy, leading to a measurable reduction in misinformation sharing within their networks. Post-intervention assessments revealed that individuals were more likely to verify information before forwarding content and more willing to engage in constructive conversations when encountering false claims.

Beyond personal behaviour, the initiative fostered collective resilience, with communities adopting social norms that prioritize critical media engagement. Schools and workplaces that implemented the programme observed sustained engagement with media literacy principles, with participants continuing to practise and promote responsible digital habits. These outcomes have informed broader policy discussions on integrating social-based media literacy strategies into national digital safety frameworks, underscoring the power of interpersonal networks in combating disinformation.

By embedding media literacy within everyday social interactions, this MIL interpersonal initiative ensures that critical engagement with information becomes a shared societal responsibility – one conversation at a time.

6.3 Digital citizenship curriculum

The Common Sense media literacy intervention is designed to equip students and educators with the kind of critical thinking needed to engage effectively with the complexities of the modern digital world. Recognizing that misinformation and disinformation thrive in an environment of low media literacy, this initiative integrates digital literacy, digital citizenship and well-being into school curricula. By fostering both educator confidence and student engagement, the intervention aims to create a holistic learning environment where young people can critically analyse information, recognize biases and understand the ethical implications of online content.

Implemented in schools in and around London and Essex by Common Sense, the intervention is grounded in research conducted by the London School of Economics (LSE). A study evaluating the programme's effectiveness found that students demonstrated notable improvements in media literacy, misinformation detection and responsible digital behaviour after as little as six weeks of exposure.[46]

Socio-ecological level

The intervention targets the community level (students in school) but also extends beyond the classroom, fostering a broader cultural shift in how misinformation is recognized and addressed. The programme acknowledges that students do not consume or engage with digital content in isolation – parents, caregivers, educators and local communities play an integral role in shaping their digital habits and critical-thinking skills.

Students who develop media literacy skills in school frequently share their knowledge with parents and caregivers, many of whom lack formal digital literacy education. This intergenerational learning was evident in the LSE's independent evaluation, where students reported helping family members navigate misleading or dangerous content on social media, particularly on topics such as online scams, health misinformation and political disinformation.

Teachers are not only trained to deliver media literacy lessons but also encouraged to host discussions with parents and local organizations, reinforcing digital safety as a shared responsibility. Recognizing that misinformation often spreads through family and social networks, Common Sense provides accessible resources to help parents navigate online risks alongside their children, covering topics including media balance, misinformation detection and AIGC. Schools also collaborate with libraries, youth organizations and community centres to expand access to digital literacy resources beyond the classroom. Additionally, the intervention allows educators to tailor discussions to regional concerns, ensuring communities are equipped to address locally relevant misinformation, from election falsehoods to health myths.

By fostering digital literacy at the community level, the Common Sense intervention ensures that media literacy is not confined to formal education settings. Instead, it becomes a shared societal responsibility, where young people, educators, parents and local institutions work collectively to build resilience against misinformation.

Disinformation life cycle level

The Common Sense media literacy intervention affects multiple stages of the disinformation life cycle by equipping students, educators and communities with the skills to critically engage with digital content. At the distribution stage, it educates students on how algorithms, engagement metrics and virality influence the spread of false information, encouraging more mindful sharing habits. At the consumption stage, students develop the ability to detect misinformation through source verification, lateral reading and exposure to real-world digital dilemmas. Finally, in post-consumption, the programme promotes corrective behaviours, such as debunking misinformation, discussing digital dilemmas with peers and family and understanding the broader societal impact of false narratives.

Outcomes

The intervention has led to significant and measurable outcomes in strengthening digital resilience. Students demonstrated improved misinformation detection, with a heightened ability to critically assess misleading content and verify sources. Post-intervention assessments revealed higher digital literacy scores, particularly in understanding digital privacy, online identity and the implications of AIGC. The programme also fostered greater student engagement, with participants finding the lessons both relevant and applicable to their everyday online experiences. Beyond the classroom, the intervention created a ripple effect on families, as many students reported helping parents identify fake news and navigate misinformation on social media. Its success has contributed to scalability and policy influence, with findings used to advocate for integrating media literacy into national education policies and broader digital safety frameworks. By embedding structured, research-backed media literacy education in schools and extending its impact to communities, the initiative is cultivating a more informed, critical and responsible digital generation.

6.4 AI-generated content literacy

A survey by MediaWise found that while most adults today are concerned about misleading and AI-generated images online, they often lack the skills and confidence to identify them.[47]

When it comes to content generated with AI, TikTok has a comprehensive approach that includes firm safety policies, reporting and labelling tools and media literacy campaigns to encourage the responsible use of AI on the platform. The approach has been informed by partnering with peers and experts (including Safety Advisory Councils as well as external partners such as the Content Authenticity Initiative) to share learnings and solutions to the collective challenges TikTok is seeing in relation to AIGC.

Socio-ecological level

TikTok's intervention operates at the institutional level – for example, with its Community Guidelines,[48] which require individuals to label AIGC or heavily edited media that depicts realistic-appearing people or scenes and prohibit certain kinds of realistic AIGC, such as content falsely depicting a public figure making an endorsement they did not make. The organization also prohibits harmful misinformation, non-consensual sexual imagery, impersonation and other harmful content, regardless of whether it is AI-generated.

When it comes to reporting tools, as AI evolves, TikTok continuously updates and builds new detection models to identify content that violates its policies, while also enabling its community to report potentially violative content for review. TikTok also partners with more than 20 fact-checking

organizations as well as a range of experts who help it stay ahead of new kinds of potentially harmful content. From October-December 2024, over 98% of violative content on TikTok was removed before a user reported it to the platform.[49]

Disinformation life cycle level

TikTok's intervention addresses three key stages of the disinformation life cycle: content creation, distribution and consumption.

At the content creation stage, TikTok was the first platform of its kind to launch a tool that enables creators to easily label their AIGC and give viewers important context.

In terms of distribution, TikTok was the first video-sharing platform to start implementing the Coalition for Content Provenance and Authenticity's (C2PA) content credentials technology, which enables TikTok to automatically label AIGC created on some other platforms, while informing C2PA of AIGC made on TikTok.[50]

Finally, at the consumption stage, TikTok invests in media literacy campaigns and other strategies to increase awareness of its AIGC tools and controls so users can feel confident about navigating content on its platform. While experts welcome AIGC labelling measures, they emphasize that media literacy is also crucial for driving adoption and clarity. To help its community navigate AIGC and misinformation online, TikTok has launched media literacy resources that were developed with guidance from experts.

In 2024, TikTok launched an AIGC transparency in-app campaign (developed with expert guidance from WITNESS and Mutale Nkonde, Chief Executive Officer and founder of AI for the People and a member of TikTok's US Content Advisory Council) to help the TikTok community feel more confident about navigating content on the platform and beyond. The campaign had two objectives:

1. Drive awareness of product features (AIGC labels and the reporting tool) that can help people spot and act upon AIGC misinformation

2. Increase familiarity with TikTok's approach to AIGC

As part of this campaign, TikTok used AI technologies such as dubbing, translation and video generation to create two in-app videos that highlighted different use cases and common scenarios of AIGC. The goal of using GenAI was to help illustrate both the creative power and potential risks of this technology. Each video opened with an engaging hook and disclosed that the content is made with the help of AI early on, highlighting that AIGC is not inherently harmful, and can be a powerful and positive creative tool when used properly. The scripts were intended to be entertaining, educational and positive in tone. These videos were launched in five different languages in five countries.[51]

Outcomes

The videos reached 90 million users in total in five countries and led to a knowledge hub that had more information and tips on how to spot AIGC. The campaign helped drive awareness of and familiarity with TikTok's AIGC tools and controls as well as highlighting the importance of AI transparency more broadly. This helped users spot and act upon potentially misleading AIGC, reinforcing TikTok's commitment to educating communities on media literacy and providing a safe, welcoming and trustworthy creative environment.

This campaign is an example of how TikTok has been working with experts to develop media literacy strategies that can help its community identify and think critically about AIGC and misinformation.

6.5 Employer-based adult digital literacy

The Business Council for Democracy (BC4D) is a joint initiative from the Hertie Foundation, Institute for Strategic Dialogue and Robert Bosch Foundation designed to deliver tailored interventions in workplaces across Germany in partnership with private and public employers.[52] The first national programme of its kind, BC4D enables employers to embed the protection of democracy as part of their corporate culture, upskilling personnel with eight 60-minute sessions designed to develop better digital hygiene and the awareness and skills to identify and address issues such as online hate, conspiracy narratives and mis- and disinformation. It is specifically designed to create an open forum for debate, combining expert-led workshops, interactive modules and facilitated discussions led by specialists.

BC4D not only provides employees with the skills and tools necessary to protect themselves, their families and their workplace environment from digital harms, it also empowers them to become ambassadors for a healthy democratic (online) discourse. In doing so, it seeks to foster more inclusive and successful workplace cultures and extend those outcomes beyond the professional domain.

Socio-ecological level

The BC4D programme delivers impact at the institutional level of the model, although it touches on other levels due to the knowledge gained potentially impacting other aspects. For employees taking part, the modules raise critical awareness and embed skills and tools necessary to protect themselves from digital harm – including how to critically assess information, how to recognize the common tropes of violent extremist rhetoric and targeted hate speech and how to respond to online harm when encountered in their everyday browsing. Participants also become force-multipliers in their workplace environment and outside.

At the institutional level, BC4D recognizes the privileged access and sustained relationship that employers have with their staff. As shown by Edelman's 2025 Trust Barometer, businesses remain the most trusted intuitions and, despite all-round declining trust, 75% of respondents still trust their employers to do what is right.[53] Delivering activity for adults is a notorious challenge, especially in relation to MIL – beyond formal education, the points of entry for engagement become more limited, less unified and harder to develop into long-term partnerships. This programme leverages the unique environment of the workplace, in which people from often diverse backgrounds need to collaborate and mitigate conflict in service of common goals.

The case for employer engagement in tackling online harms is grounded in three key imperatives:

- **Legal**: As remote work and digital service delivery expand, employers must extend their duty of care to digital spaces. German employment law requires safe, inclusive workplaces across protected characteristics such as race, gender and religion. That includes protecting staff from online risks such as hate speech, harassment, doxxing, conspiracy theories and extremist narratives – including when these appear on internal office platforms. Fulfilling this mandate means investing in staff training, awareness and clear reporting pathways to address digital harms effectively.

- **Moral**: Most employers seek to foster a fair, respectful and inclusive work environment. Investing in initiatives such as BC4D supports employee well-being and reduces stress, discrimination and division. It is also a powerful ethical stance, as the workplace is one of the most direct access points for adults to learn about digital safety. Employers have a role in shaping norms of dignity, mutual respect and cohesion that extend beyond professional tasks.

- **Business-oriented**: Consumers increasingly judge companies based on values as much as products, prompting a rise in "triple bottom line" models. Today's leading brands incorporate themes of justice and inclusion into their public messaging, responding to demand from both the market and potential hires. Taking a stand against disinformation and online hate builds trust, strengthens reputation and makes the workplace more appealing – especially to younger, purpose-driven talent. A proactive stance on digital safety sets the BC4D network apart in both mission and identity.

Disinformation life cycle level

BC4D touches upon many stages of the life cycle, but in particular consumption and post-consumption. For consumption, modules expose participants to features and vulnerabilities in the digital sphere that may incentivize or exacerbate harm – such as algorithmic recommender systems, echo chamber effects, monetization of content, micro-targeting and other ad-based platform design. Armed with this knowledge and practical skills, those graduating from BC4D should be more attuned to the benefits and pitfalls of a "personalized web" and how they can diversify their media diet. Finally, BC4D recognizes that, even with healthier habits, participants are still likely to encounter certain online harms and must be equipped to respond. The programme

therefore provides a range of accessible tools to mitigate harm – including in-platform user controls, how to report harmful content, legal definitions of hate speech and ways to support victims – which is often complemented by a specific focus on company policies to safeguard employees and ensure an inclusive, cohesive workplace for all.

Outcomes

BC4D has engaged more than 85 companies and 2,200 participants, including multinational firms in Germany such as SAP, Bosch, Nokia, VW and two of the biggest German football clubs, Hamburg SV and St Pauli. Key outcomes include the training of trainers (ToT) and cascade models for delivery. After completing the courses, participating companies have institutionalized the learnings from the trainings – for instance, by adapting their internal codes of conduct, updating their definitions and improving their compliance processes. Evaluations also indicate a tangible moderation of discourse within these organizations, contributing to a more inclusive and respectful workplace environment.

The multiplier effect of the initiative has proven significant, with 83% of participants actively sharing the training content with personal or professional contacts. Many companies have since extended their efforts, incorporating elements of BC4D into mandatory, company-wide training and internal academies for continued professional development (CPD). In terms of addressing online harms, the external evaluation results were particularly impressive. Some 99% of respondents said they had acquired the ability to confront hate speech, 98% felt equipped to detect and counter disinformation and 95% felt more comfortable dealing with conspiracy narratives. Additionally, 99% of participants valued their employers' commitment to democratic values, highlighting the profound impact of the initiative on organizational culture. Overall, BC4D positively transformed participating companies, specifically with regard to fostering digital literacy and strengthening democratic discourse.

6.6 National media education policy

Finland's Media Literacy in Finland policy, published in 2019 by the Ministry of Education and Culture in collaboration with the National Audiovisual Institute (KAVI), represents a landmark update to the country's media education framework originally set out in 2013.[54] It represents a comprehensive, whole-of-society approach designed to cultivate robust media literacy skills across all age groups, with particular emphasis on children and adolescents. Recognizing the rapid digital transformation and the rising threat of disinformation, the policy aims to embed MIL as a fundamental part of civic education and lifelong learning. It seeks to empower citizens to critically evaluate media content, engage responsibly in digital environments and contribute to a healthy democratic society.

The policy promotes broad collaboration among multiple government ministries – including Education and Culture, Transport and Communications, Justice and the Prime Minister's Office – alongside key institutions such as schools, libraries, youth organizations, cultural bodies and non-governmental organizations (NGOs). A special focus is placed on tackling hate speech and political disinformation, particularly during sensitive times such as elections and public crises, thereby enhancing societal resilience against information manipulation.

Socio-ecological level

At the policy level, Finland's media education strategy is characterized by strong governmental leadership and a coordinated, cross-ministerial approach. Key ministries collaborate to ensure media literacy is embedded as a national priority. The Finnish National Agency for Education plays a central role in institutionalizing media literacy by integrating it into the national curricula and teacher training programmes, ensuring that educators across the country are equipped to teach critical media skills to children from early childhood onwards.

KAVI operates as a pivotal government agency tasked with developing educational resources, providing training for professionals and advising policy-makers on emerging challenges related to disinformation and digital media use. Finland's policy framework also aligns with European Union directives and regulations, such as the Audiovisual Media Services Directive, ensuring national measures comply with broader legal standards for safe and transparent media environments.

Disinformation life cycle level

The Finnish media literacy policy strategically addresses the pre-creation, distribution and consumption stages of the disinformation life cycle by embedding media literacy education and awareness-building across society.

At the pre-creation stage, the Finnish National Agency for Education (EDUFI) plays a central role by integrating media literacy into the national school curriculum from early education onwards. This approach aims to develop critical thinking skills, ethical understanding and digital responsibility among students before they begin creating or sharing content themselves, thereby reducing the likelihood of generating or inadvertently contributing to disinformation. By fostering a strong foundation in evaluating sources and understanding the impact of information, these education efforts work to prevent harmful content at its inception.

Regarding the distribution stage, organizations such as the Finnish Communications Regulatory Authority (FICORA) and various civil society groups promote responsible information-sharing behaviours among citizens. Additionally, Finland enforces EU-level regulations such as the Audiovisual Media Services Directive, which obligates platforms to address harmful content as well as promoting transparency.

At the consumption stage, initiatives led by EDUFI, media literacy NGOs and public broadcasters such as Yleisradio (Yle) focus on equipping citizens with the skills to critically evaluate the information they encounter. Projects like Facts Against Hate equip individuals, especially youth, with skills to identify and counteract hate speech, extremist content and deliberately misleading narratives.

Outcomes

Finland's media education policy has yielded a resilient, multidimensional ecosystem that integrates media literacy into the very fabric of education, governance and community engagement.

It has successfully mainstreamed media and information literacy competencies across national curricula, teacher training and public education initiatives, creating widespread awareness and practical skills to counter disinformation. The coordinated involvement of multiple ministries and specialized agencies has fostered strong institutional support and sustained funding, ensuring policy longevity and adaptability. Access to free, high-quality educational materials and training services has democratized media literacy, reached diverse demographic groups and reduced digital inequalities.

The policy's emphasis on combating political disinformation and hate speech during critical moments such as elections has reinforced Finland's democratic resilience.

7 Action points

Addressing the evolving challenges of disinformation requires a coordinated, multistakeholder response.

As this report demonstrates, MIL is a foundational tool for increasing information integrity, equipping individuals and communities with the skills necessary to critically engage with digital content. However, to maximize impact, interventions must be strategically deployed where they are most needed.

7.1 Organizational recommendations

The framework presented in this report provides a structured method for mapping existing interventions along the disinformation life cycle and across socio-ecological levels. This approach enables policy-makers, educators, civil society organizations and technology companies to identify areas where efforts are lacking or misaligned.

Organizations should therefore:

- **Assess existing interventions:** Use the combined life cycle and socio-ecological models to evaluate existing strategies. Are efforts overly concentrated on reactive debunking at the point of consumption? Are earlier and later stages – such as pre-creation (e.g. narrative inoculation) or post-consumption (e.g. critical reflection) – neglected?

  Example: A platform can conduct life cycle audits of its content moderation and educational interventions to ensure MIL is embedded in both pre-upload creator guidance and post-viewer feedback loops.

- **Identify gaps and expand efforts:** While formal education remains essential, disinformation affects people of all ages and professions. Tailored MIL efforts should be developed for adults in healthcare, journalism, law enforcement and public administration.

  Example: Governments can work with medical boards to provide MIL certifications for healthcare workers exposed to health-related misinformation, such as vaccine conspiracy theories.

- **Engage vulnerable populations:** Disinformation often exploits linguistic divides, access gaps and socio-political marginalization. Interventions must reach non-native speakers, rural populations and those with low digital literacy.

  Example: WhatsApp's helpline model in India allows users to forward suspected misinformation to verified fact-checkers in local languages, offering an example of both linguistic accessibility and grassroots engagement.[55]

- **Strengthen cross-sector collaboration:** Effective MIL requires coordination across civil society, regulatory bodies and tech platforms to reinforce rather than duplicate efforts.

  Example: The European Digital Media Observatory (EDMO) brings together fact-checkers, academic researchers and media regulators to build shared standards and data across EU member states.[56]

- **Leverage community-based learning:** MIL messages are more likely to resonate when they come from trusted community figures. Training local leaders, influencers and faith-based organizations can help normalize critical consumption habits.

  Example: In Kenya, a UNESCO programme partnered with community radio stations and youth to address the spread of hate speech and mis- or disinformation online ahead of national elections.[57]

– **Adapt to emerging threats:** Disinformation evolves with technology, especially with GenAI, deepfakes and algorithmic manipulation. MIL must remain agile and incorporate the latest insights on digital deception.

Example: Initiatives can equip journalists, educators and others with tools to identify AIGC and teach others to do the same.

7.2 Scaling up effective interventions

While progress has been made, MIL initiatives remain fragmented and underfunded. To build a more resilient society, stakeholders should:

– **Integrate MIL into AI governance:** Regulate synthetic media by requiring clear labelling of AIGC and introducing platform-level friction (e.g. interstitial warnings or slowed sharing).

Example: An individual platform's new policies on AI content labelling should serve as an initial step, but standards should be harmonized across platforms.

– **Embed MIL in everyday digital experiences:** Users should encounter MIL nudges – such as credibility warnings, source reminders or contextual explanations – during routine engagement online.

Example: X's community notes allow users to collaboratively add context to misleading tweets, crowd-sourcing fact-checking in a way that encourages active participation and reflection.

– **Educate users on the why and the how:** While users often learn how to fact-check or spot bias, they are rarely taught why they see certain content. AI-driven algorithms in combination with user preferences, platform features and supply-and-demand dynamics amplify content that captures attention.[58]

Example: Integrate teachings on the political economy of social media along with other MIL teachings.

– **Support lifelong-learning initiatives:** Expand MIL access beyond schools by incorporating modules into workforce development, civic education programmes and digital onboarding.

Example: BC4D includes MIL training for adult employees, not just students.

– **Incentivize private-sector participation:** While technology companies have now taken steps to support MIL, there is a need to build on and scale these efforts across all industries. For these companies not doing it, encourage them to co-create with educators, fact-checkers and civil society MIL experiences beyond content moderation.

Example: TikTok's election hubs that direct users to authoritative information (for example, national election commission sites) demonstrate how platforms can embed media literacy interventions at scale.[59]

– **Enhance evaluation mechanisms:** Impact measurement is essential for scaling what works. Invest in independent, longitudinal evaluations that track behavioural change across populations and platforms.

Example: The UK's Ofcom regulator conducts regular Digital Literacy Tracker surveys on attitudes of adults.[60]

– **Promote long-term investment strategies:** Shift from one-off grants to sustained, multi-year support for MIL ecosystems. This ensures consistency, local ownership and better adaptation to new threats.

Example: Multi-year and long-term investments in national MIL strategies and infrastructure by organizations.

– **Develop "pre-bunking" initiatives:** Rather than only correcting misinformation after exposure, invest in pre-bunking strategies that teach people to recognize common disinformation tactics.

Example: Google and Jigsaw's YouTube ads based on inoculation theory – which teach users about disinformation techniques such as scapegoating – have shown measurable improvements in users' resistance to manipulation.[61]

7.3 Policy recommendations

A robust policy and regulatory approach are critical to embedding MIL into national strategies and ensuring it addresses the full disinformation life cycle. Policy-makers should:

- **Integrate MIL into broader digital resilience strategies:**
  Position MIL as a core pillar of national digital resilience frameworks by embedding it into cybersecurity strategies, media regulation policies and digital governance structures. This integrated approach ensures that MIL is not treated in isolation, but as a cross-cutting tool for protecting democratic discourse, national security and social cohesion.

  Example: Develop MIL alongside cybersecurity preparedness and strategic communications, empowering citizens and institutions to identify and resist information threats as part of a whole-of-society approach to hybrid threats.

- **Ensure rights-respecting enforcement and oversight of digital regulation:**
  In democratic societies where social media or digital platform regulation is already in place, governments should ensure the rights-respecting enforcement of these rules. This includes conducting regular assessments of the impact on freedom of expression and civic participation, alongside mechanisms for independent oversight and public learning.

  Example: The EU Digital Services Act requires VLOPs to conduct annual risk assessments of systemic harms, such as disinformation, with oversight by independent regulators and civil society participation.

- **Incentivize safety-by-design through innovation-oriented policy:**
  Governments should create regulatory and financial incentives for technology developers to prioritize safety-by-design, particularly in emerging technologies such as GenAI.

  Example: Countries could provide tax breaks, R&D grants, public procurement preferences or inclusion in digital innovation hubs for companies that embed user safeguards and transparency mechanisms from the outset.

- **Mandate MIL in education systems:**
  Ensure that digital literacy, critical thinking and MIL are woven throughout national curricula.

  Example: Estonia has embedded MIL into its national curriculum from primary through to secondary school as part of its digital nation strategy.[62]

- **MIL in teacher certification:**
  Make MIL competencies a requirement for educators across all subjects – not only media or IT.

  Example: Localize MIL teaching standards into national teacher accreditation systems.

- **Support lifelong-learning initiatives:**
  Codify MIL into adult learning, vocational training and public communications campaigns.

  Example: The UK's Share Checklist campaign encouraged users to check information regarding COVID-19 before sharing.[63]

7.4 Unified global commitment

Disinformation is not just a technological issue, it is a societal challenge that threatens democracy, public health and social cohesion. A successful response requires sustained commitment, political will and shared responsibility. Governments must integrate MIL into national strategies for digital resilience.[64] Educational institutions must embed MIL into curricula. The private sector must implement responsible design principles. And civil society must continue to advocate for informed and empowered digital citizens. International organizations must facilitate dialogue, standard-setting and resource-sharing across borders to address disinformation as a global issue.

By leveraging the tools and models outlined in this report, stakeholders can move beyond ad hoc interventions and build a comprehensive, long-term strategy for combating disinformation.

Conclusion

Media literacy is pivotal for protecting society from harm and building a digital future where technology benefits everyone.

In an era defined by rapid technological advancements, global connectivity and shifting information dynamics, the integrity of information stands as a foundational element of resilient, democratic and rights-respecting societies, but achieving it can seem illusive. While interventions are often reactive and outpaced by the sheer scale and complexity of emerging technologies, the tide is beginning to shift. This report aims to serve as both a challenge and call to action, showing how efforts can be distributed along the information supply chain and create change from the grassroots through to governments.

The model presented seeks to reframe how we conceptualize the issue, plan activity, allocate resources and make use of all available levers of influence at our disposal. It can be used to map interventions in a more nuanced and comprehensive manner and, in turn, identify urgent gaps and opportunities for investment, advocacy or partnership across the globe. In doing so, it can enable a more holistic and impactful approach, not only to bolster information ecosystems but to build more inclusive and constructive public discourse. Crucially, it should shift the focus to upstream interventions and ways to anticipate or pre-empt bad actors and efforts to manipulate, rather than purely firefighting after the fact. Strengthening information integrity is not merely about countering threats, it is about upholding and empowering the right to freedom of expression, access to information and meaningful participation in civic life.

MIL remains pivotal; however, this report demonstrates how the focus should not only be on education. Rather, a more expansive understanding of MIL is needed, one that encompasses stages across the disinformation life cycle and stakeholders at each level of the socio-ecological model. In this sense, MIL can and must include smarter platform design and governance, enhanced channels of communication with the public, more diverse news media and campaigns that shift community and wider cultural norms. As information environments continue to evolve, investments in MIL must be scaled, diversified and embedded across education, professional development, community engagement and policy initiatives.

Above all, these interventions should be unified by a common goal: developing informed, engaged, resilient communities who claim their rights and enact their civic duty in the information space. By championing MIL, we are not only protecting societies from harm, we are affirming our collective commitment to building a digital future grounded in fundamental rights and global public goods.

MIL offers a powerful rights and research-based pathway to achieve this goal. By equipping individuals with critical thinking skills, ethical media practices and the ability to navigate a complex digital landscape, MIL empowers people to actively claim their rights and responsibilities in the information society. It supports the development of informed, engaged and resilient communities capable of sustaining democratic principles and inclusive public discourse.

As information environments continue to evolve, investments in MIL must be scaled, diversified and embedded across education, professional development, community engagement and policy initiatives. Upholding information integrity is a shared societal endeavour, requiring collaboration across governments, private sector actors, civil society, academia and individuals themselves.

Contributors

Lead authors

Adeline Hulin
Head of Unit for Media and Information Literacy and Digital Competencies, United Nations Educational, Scientific and Cultural Organization (UNESCO)

Jennie King
Director of Climate Research and Policy, Institute for Strategic Dialogue

Daegan Kingery
Specialist, Digital Safety and Trustworthy Technology, World Economic Forum

Ashley Walker
User Experience Researcher, Trust and Safety, Google

World Economic Forum

Agustina Callegari
Lead, Technology Governance, Safety, and International Cooperation

Daniel Dobrygowski
Head, Governance and Trust

Cathy Li
Head, AI, Data and Metaverse; Member of the Executive Committee

Acknowledgements

This paper is a combined effort based on numerous interviews, discussions, workshops and research. The opinions expressed herein do not necessarily reflect the views of the individuals or organizations listed below.

Sincere appreciation is extended to the following working group members, who spent numerous hours providing critical input and feedback on the drafts. Their diverse insights are fundamental to the success of this work.

Jeanne Bourgault
President and Chief Executive Officer, Internews

Sandra Braman
Professor of Media and Information and Senior Scholar, Quello Center of Media and Information Policy, Michigan State University

Isabelle Frances-Wright
Director of Technology and Society, Institute for Strategic Dialogue

Sasha Havlicek
Chief Executive Officer, Institute for Strategic Dialogue

Peggy Hicks
Director, Thematic Engagement, Special Procedures, and Right to Development Division, Office of the High Commissioner for Human Rights (OHCHR)

Adeline Hulin
Head of Unit for Media and Information Literacy and Digital Competencies, United Nations Educational, Scientific and Cultural Organization (UNESCO)

Avinash Kaul
Chief Executive Officer, Network 18 Media & Investments

Irene Liu
Regional Director, Asia and the Pacific, International Fund for Public Interest Media

Jeffrey McGregor
Chief Executive Officer, Truepic

Mauro Miedico
Director, United Nations Counter-Terrorism Centre, United Nations Office of Counter-Terrorism (UNOCT)

Susan Ness
Nonresident Senior Fellow, the Atlantic Council's Europe Center, the Atlantic Council

Bruno Osorio
Outreach and Partnerships Manager, LATAM – Trust and Safety, TikTok

María Paz Canales Loebel
Head of Legal, Policy and Research, Global Partners Digital

Joanna Rubinstein
President and Chief Executive Officer, World Childhood Foundation USA

Liz Thomas
Director Public Policy, Digital Safety, Microsoft

Nicholas Thompson
Chief Executive Officer, The Atlantic

Hayley van Loon
Deputy Chief Executive Officer,
Crime Stoppers International

Steven Vosloo
Digital Foresight and Policy Specialist,
United Nations Children's Fund (UNICEF)

David Wright
Chief Executive Officer, SWGfL

John Zoltner
Global Lead, Protection from Digital Harm,
Save the Children International

Production

Michela Liberale Dorbolò
Designer, World Economic Forum

Simon Smith
Editor, Astra Content

Endnotes

1. United Nations General Assembly. (2024, April 5). *Resolution adopted by the Human Rights Council on 3 April 2024: 55/10 – Role of States in countering the negative impact of disinformation on the enjoyment and realization of human rights.* https://docs.un.org/en/A/HRC/RES/55/10#:~:text=Calls%20upon%20all%20States%20to,6
2. Kreps, S., Miles, R. M., & Brundage, M. (2020). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science, 9*(1), pp. 104–117. https://www.cambridge.org/core/journals/journal-of-experimental-political-science/article/abs/all-the-news-thats-fit-to-fabricate-aigenerated-text-as-a-tool-of-media-misinformation/40F27F0661B839FA47375F538C19FA59
3. SWGfL. (2024). *Evaluation of the ProjectEVOLVE database: Understanding online safety delivery and assessment in schools*. https://swgfl.org.uk/research/evaluation-projectevolve-database/
4. Sadeghi, M., & Blachez, I. (2025, March 6). *A well-funded Moscow-based global "news" network has infected Western artificial intelligence tools worldwide with Russian propaganda*. NewsGuard. https://www.newsguardrealitycheck.com/p/a-well-funded-moscow-based-global?utm_campaign=email-half-post&r=1o6y5f&utm_source=substack&utm_medium=email
5. United Nations General Assembly. (2024, September 22). *Resolution adopted by the General Assembly on 22 September 2024: 79/1 – The Pact for the Future*. https://docs.un.org/en/A/RES/79/1
6. UNESCO. (2021). *Media and information literate citizens: Think critically, click wisely!*. https://unesdoc.unesco.org/ark:/48223/pf0000377068
7. Carlsson, U. (ed) (2019). *Understanding media and information literacy (MIL) in the digital age*. Department of Journalism, Media and Communication (JMG). University of Gothenburg. https://www.gu.se/en/jmg/understanding-media-and-information-literacy-mil-in-the-digital-age
8. UNESCO & Ipsos. (2023). *Survey on the impact of online disinformation and hate speech*. https://www.unesco.org/sites/default/files/medias/fichiers/2023/11/unesco_ipsos_survey.pdf
9. Newman, N., Fletcher, R., Robertson, C. T., Ross Arguedas, A., & Nielsen, R. K. (2024). *Reuters Institute digital news report 2024*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024
10. Ha, L. (2024). *Behind the screens: Insights from digital content creators; understanding their intentions, practices and challenges*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000392006
11. Le Fevre Cervini, E. M., & Carro, M. (2024, October 25). *An overview of the impact of GenAI and deepfakes on global electoral processes*. Italian Institute for International Political Studies. https://www.ispionline.it/en/publication/an-overview-of-the-impact-of-genai-and-deepfakes-on-global-electoral-processes-167584
12. Aguilar, N. (2024, June 4). *AI overviews are taking over Google search. How to turn it off*. CNET. https://www.cnet.com/tech/services-and-software/ai-overviews-are-taking-over-google-search-how-to-turn-it-off/
13. McCallum, S. (2023, April 1). *ChatGPT banned in Italy over privacy concerns*. BBC. https://www.bbc.com/news/technology-65139406
14. European Union. (2024). *The EU Artificial Intelligence Act*. https://artificialintelligenceact.eu/
15. Knibbs, K. (2023, October 3). *Researchers tested AI watermarks – and broke all of them*. WIRED. https://www.wired.com/story/artificial-intelligence-watermarking-issues/
16. Izsák-Ndiaye, R. (2021). *"If I disappear": Global report on protecting young people in civic space*. United Nations. https://unoy.org/downloads/if-i-disappear-global-report-on-protecting-young-people-in-civic-space/
17. UNESCO. (2023, February 10). *How to combat hate speech and gendered disinformation online? UNESCO provides some ideas*. https://www.unesco.org/en/articles/how-combat-hate-speech-and-gendered-disinformation-online-unesco-provides-some-ideas?utm_source=chatgpt.com
18. Office of the Surgeon General. (2023). *Social media and youth mental health: The U.S. Surgeon General's Advisory*. https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf
19. Ayad, M., & Frances-Wright, I. (2024, January 24). *Minors exposed to mass shooter glorification across mainstream social media platforms*. Institute for Strategic Dialogue. https://www.isdglobal.org/digital_dispatches/minors-exposed-to-mass-shooter-glorification-across-mainstream-social-media-platforms/
20. Frances-Wright, I., Ayad, M., & and Jacobs, E. (2024, August 7). *One click away: Foiled plot targeting Taylor Swift concert highlights minors' access to terrorist content online*. Institute for Strategic Dialogue. https://www.isdglobal.org/digital_dispatches/one-click-away-foiled-plot-targeting-taylor-swift-concert-highlights-minors-access-to-terrorist-content-online/
21. Barberá, P. (2020). Social media, echo chambers, and political polarization. In N. Persily & J. A. Tucker (Eds.), *Social media and democracy*, pp. 34–55. Cambridge University Press. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/E79E2BBF03C18C3A56A5CC393698F117/9781108835558AR.pdf/Social_Media_and_Democracy.pdf?event-type=FTLA
22. Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S., & Dragan, A. D. (2025). Engagement, user satisfaction, and the amplification of divisive content on social media. *PNAS Nexus, 4*(3). https://academic.oup.com/pnasnexus/article/4/3/pgaf062/8052060

23. UNESCO & Ipsos. (2023). *Survey on the impact of online disinformation and hate speech*. https://www.unesco.org/sites/default/files/medias/fichiers/2023/11/unesco_ipsos_survey.pdf
24. Newsham, J. (2025, May 27). *AI hallucinations in court documents are a growing problem, and data shows lawyers are responsible for many of the errors*. Business Insider. https://www.businessinsider.com/increasing-ai-hallucinations-fake-citations-court-records-data-2025-5
25. UNESCO. (2019). *Global standards for media and information literacy curricula development guidelines*. https://www.unesco.org/sites/default/files/medias/files/2022/02/Global%20Standards%20for%20Media%20and%20Information%20Literacy%20Curricula%20Development%20Guidelines_EN.pdf
26. Techniques on how to build psychological resistance and reduce susceptibility to manipulation.
27. UNESCO. (2020, November 30). *Lifeskills training on financial literacy and basic business management skills for ALS learners*. https://www.unesco.org/en/articles/lifeskills-training-financial-literacy-and-basic-business-management-skills-als-learners
28. UNESCO. (2024). *Operational guidelines: Constructing UNESCO Media and Information Literacy Cities*. https://unesdoc.unesco.org/ark:/48223/pf0000391748
29. UNESCO. (2021, February 2). *Network of youth organisations in south east Europe adopts MIL policies*. https://www.unesco.org/en/articles/network-youth-organisations-south-east-europe-adopts-mil-policies
30. UNESCO. (2024). *Mapping of media and information literacy initiatives in Ibero-America – 2023*. https://unesdoc.unesco.org/ark:/48223/pf0000390030_eng
31. United Nations. (2024). *United Nations Global Principles For Information Integrity*. https://www.un.org/en/information-integrity/global-principles
32. United Nations. (2021). *Our common agenda – Report of the Secretary-Genera*l. https://www.un.org/en/common-agenda
33. Actors who have a high outreach and can multiply their impact, e.g. content creators.
34. Coordinated deception that mimics grassroots behaviour to manipulate public opinion while hiding the true origins of the user (e.g. a fake account in which someone pretends to be a national from one country complaining about a domestic issue, but actually hails from a different country).
35. UK Government. (2023). *Online Safety Act 2023*. https://www.legislation.gov.uk/ukpga/2023/50
36. Australian Government. (2021). *Online Safety Bill 202*1. https://www.aph.gov.au/Parliamentary_Business/Bills_LEGislation/Bills_Search_Results/Result?bId=r6680
37. European Union. (2022). *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a single market for digital services and amending directive 2000/31/EC (Digital Services Act*). Official Journal of the European Union, L 277, 1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj
38. European Commission. (2024, November 26). *Very large online platforms and search engines to publish first risk assessment and audit reports under the Digital Services Act* [Press release]. https://digital-strategy.ec.europa.eu/en/news/very-large-online-platforms-and-search-engines-publish-first-risk-assessment-and-audit-reports
39. European Commission. (n.d.) *Welcome to the DSA Transparency Database!* Retrieved June 6, 2025, from https://transparency.dsa.ec.europa.eu/?lang=en
40. Ofcom. (2024, October 7). *Ofcom's three-year media literacy strategy*. https://www.ofcom.org.uk/media-use-and-attitudes/media-literacy/ofcoms-three-year-media-literacy-strategy
41. Department of Science, Innovation and Technology. (2025, May 8). *Draft statement of strategic priorities for online safety*. https://www.gov.uk/government/publications/draft-statement-of-strategic-priorities-for-online-safety/draft-statement-of-strategic-priorities-for-online-safety
42. UNESCO. (2024, November 26). *2/3 of digital content creators do not check their facts before sharing, but want to learn how to do so (UNESCO survey)* [Press release]. https://www.unesco.org/en/articles/2/3-digital-content-creators-do-not-check-their-facts-sharing-want-learn-how-do-so-unesco-survey
43. Harlow, S. (Ed.). (2024). *Content creators and journalists: Redefining news and credibility in the digital age*. Knight Center for Journalism in the Americas. https://knightcenter.utexas.edu/wp-content/uploads/2024/10/Content-Creator-and-Journalists.pdf?utm_source=journalismcourses&utm_medium=download&utm_campaign=ebook_eng
44. UNESCO. (2024, November 26). *2/3 of digital content creators do not check their facts before sharing, but want to learn how to do so (UNESCO survey)* [Press release]. https://www.unesco.org/en/articles/2/3-digital-content-creators-do-not-check-their-facts-sharing-want-learn-how-do-so-unesco-survey.
45. Mioli, M. (2024, November 19). *Free online course on influencers and journalists starts with 9,000 participants from 172 countries. Registration is still open!* Knight Center for Journalism in the Americas. https://journalismcourses.org/free-online-course-on-influencers-and-journalists-starts-with-8000-participants-from-149-countries-registration-is-still-open/
46. Abades-Barclay, F., & Banaji, S. (2024). *LSE – Common Sense digital citizenship curriculum evaluation*. Department of Media and Communication, London School of Economics. https://www.lse.ac.uk/News/News-Assets/PDFs/2024/LSE-%E2%80%94-Common-Sense-Digital-Citizenship-Curriculum-Evaluation-Report-high-res-web-version-V6.pdf
47. Poynter staff. (2023, October 25). *Survey identifies media literacy skills gap amidst rise in AI-generated content*. Poynter. https://www.poynter.org/fact-checking/media-literacy/2023/adults-worry-misleading-ai-images-lack-media-literacy/
48. TikTok. (2024). *Edited media and AI-generated content (AIGC)*. https://www.tiktok.com/community-guidelines/en/integrity-authenticity#3.

49. TikTok. (2025, March 28). *Community guidelines enforcement report*. https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2024-4

50. TikTok. (2024, May 9). *Partnering with our industry to advance AI transparency and literacy*. https://newsroom.tiktok.com/en-us/partnering-with-our-industry-to-advance-ai-transparency-and-literacy

51. TikTok. (2024). *Learn about AI-generated content with TikTok*. https://activity-i18n.tiktok.com/magic/eco/runtime/release/670693b8e6773c031f2ff1cc?isFromGCP=1&appType=muse&magic_page_no=1&lang=en-GB&region=GB

52. Business Council for Democracy. (n.d.). *Das Programm* [German language]. Retrieved June 6, 2025, from https://www.bc4d.org/

53. Edelman. (2025). *2025 Edelman Trust Barometer*. https://www.edelman.com/trust/2025/trust-barometer

54. Ministry of Education and Culture, Finland. (2019). *Media literacy in Finland: National media education policy*. https://julkaisut.valtioneuvosto.fi/handle/10024/162065

55. Meta. (2024, February 19). *MCA's WhatsApp helpline: Curbing the spread of AI-generated misinformation In India*. https://about.fb.com/news/2024/02/mcas-whatsapp-helpline-curbing-the-spread-of-ai-generated-misinformation/

56. European Commission. (2024). *European Digital Media Observatory (EDMO)*. https://digital-strategy.ec.europa.eu/en/policies/european-digital-media-observatory

57. UNESCO. (2024). *"Unganisha jamii": Community radios empower Kenya's youth to unify communities and build peace online*. https://www.unesco.org/en/articles/unganisha-jamii-community-radios-empower-kenyas-youth-unify-communities-and-build-peace-online

58. Bhargava, V. R., & Velasquez, M. (2020). Ethics of the attention economy: The problem of social media addiction. *Business Ethics Quarterly, 31*(3), pp. 321–359. https://doi.org/10.1017/beq.2020.32; Hosseinmardi, H., Ghasemian, A., Clauset, A., Mobius, M., Rothschild, D. M., & Watts, D. J. (2021). Examining the consumption of radical content on YouTube. *PNAS, 118*(32). https://doi.org/10.1073/pnas.2101967118; Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S., & Dragan, A. D. (2025). Engagement, user satisfaction, and the amplification of divisive content on social media. *PNAS Nexus, 4*(3). https://doi.org/10.1093/pnasnexus/pgaf062

59. Cade. (2024, February 14). *TikTok to launch in-app Election Centers in EU states to prevent misinformation ahead of election year*. https://cadeproject.org/updates/tiktok-to-launch-in-app-election-centers-in-eu-states-to-prevent-misinformation-ahead-of-election-year/

60. Ofcom. (2025, May 7). *Adults' media use and attitudes report 2025*. https://www.ofcom.org.uk/media-use-and-attitudes/media-habits-adults/adults-media-use-and-attitudes

61. Gerken, T. (2022, August 25). *Google to run ads educating users about fake news*. BBC. https://www.bbc.com/news/technology-62644550

62. Yee, A. (2022, January 31). *The country inoculating against disinformation*. BBC. https://www.bbc.com/future/article/20220128-the-country-inoculating-against-disinformation

63. Department for Culture, Media and Sport. (n.d.). *Check before you share toolkit*. Retrieved June 6, 2025, from https://dcmsblog.uk/check-before-you-share-toolkit/

64. SWGfl. (2024). *Evaluation of the ProjectEVOLVE database: Understanding online safety delivery and assessment in schools*. https://swgfl.org.uk/research/evaluation-projectevolve-database/

WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org