Building adult community resilience to disinformation during health emergencies through information literacy

Summary note of a workshop held on 27 – 28 November 2023

Background

This note provides a summary of workshop discussions exploring information literacy as a potential solution to the impact of scientific disinformation on adult communities. The workshop was jointly hosted by the Royal Society, the British Broadcasting Corporation (BBC), and Wikimedia UK on 27 and 28 November 2023 at Carlton House Terrace and Broadcasting House. The workshop focused on i) the role of information literacy in tackling disinformation during health emergencies, ii) the challenges of developing information literacy campaigns that can build community resilience to disinformation within adult groups, particularly minority ethnic groups in the UK, and iii) how information literacy must adapt to mitigate the risks of AI-generated disinformation. For the purposes of this note, we use the term 'disinformation', which refers to deliberately fabricated or manipulated information. However, the term 'misinformation', referring to false information shared unintentionally, was also central to discussions[1].

The workshop was convened following the publication of the Royal Society's 2022 report, *The online information environment: Understanding how the internet shapes people's engagement with scientific information*[2]. It builds on related work, including a workshop on digital content provenance and a 'public service internet' organised by Royal Society and the BBC in September 2022[3].

Recommendation 9 of the Royal Society report calls for the UK Government to invest in lifelong, nationwide, information literacy initiatives so that people can effectively evaluate online content for reliability, credibility and relevance of information. These skills are important for adults outside formal education and, specifically, those from minority ethnic groups, which refers to groups of people based on a combination of 'racial' characterisations, national identity, cultural expression, or religious affiliation other than the White majority, such as Black, Asian, Mixed, and other ethnic backgrounds[4]. These groups were disproportionally affected by the COVID-19 virus during the pandemic, including higher rates of death compared to the White British majority[5].

This note summarises a range of lightning talks and group discussions and presents suggestions for action and further research. References are included both to illustrate points raised in the workshop and to reflect the evolving challenges in the information ecosystem since it took place. This note is not intended as a verbatim record of the proceedings and does not necessarily represent the views or positions of the participants or organisations who took part. It was drafted by Royal Society staff and reviewed by staff at the BBC and Wikimedia UK.

1. The Royal Society. 2022 The online information environment. See https://royalsociety.org/topics-policy/projects/online-information-environment (accessed 15 May 2024).
2. *Ibid*.
3. The Royal Society. 2023 Generative AI, content provenance and a public service internet. See: https://royalsociety.org/news-resources/publications/2023/digital-content-provenance-bbc (accessed 15 May 2024)
4. Ethnicity facts and figures. List of ethnic groups. See https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups (accessed 18 December 2024)
5. The Royal Society. 2021 COVID-19 and its impact on minority groups in the UK. See https://royalsociety.org/-/media/policy/Publications/2022/covid-19-and-minority-ethnic-groups-workshop-report-royal-society-and-British-Academy.pdf?la=en-GB&hash=2BE0BD6F2AE0BF3EE761DB23ABBA4937 (accessed 10 January 2025)

The Royal Society

The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity. The Society's strategic priorities emphasise its commitment to the highest quality science, to curiosity-driven research, and to the development and use of science for the benefit of society. These priorities are:

- The Fellowship, Foreign Membership and beyond
- Influencing
- Research system and culture
- Science and society
- Corporate and governance.

The British Broadcasting Corporation (BBC)

The BBC is the world's leading public service broadcaster. Founded in 1922, its mission is to act in the public interest, serving all audiences through the provision of impartial, high-quality and distinctive output and services which inform, educate, and entertain. The BBC is established under a Royal Charter which sets out its public purposes. These purposes are:

- To provide impartial news and information to help people understand and engage with the world around them
- To support learning for people of all ages
- To show the most creative, highest quality and distinctive output and services
- To reflect, represent and serve the diverse communities of all of the United Kingdom's nations and regions and, in doing so, support the creative economy across the United Kingdom
- To reflect the United Kingdom, its culture and values to the world

Wikimedia UK

Wikimedia UK is the national charity for the global Wikimedia movement, bringing together practical and policy expertise about Wikipedia and the other Wikimedia projects. The charity engages with thousands of people every year through education and outreach — including media and information literacy programmes — while their work in partnership with the cultural sector is transforming the public's engagement with cultural heritage collections and delivering increased reach and impact for some of the world's leading galleries, libraries, archives and museums.

Wikimedia UK's vision is of a more informed, democratic and equitable society through open knowledge, and their work focuses on knowledge and communities that have been left out by structures of power and privilege; breaking down the barriers that prevent people and organisations from accessing and contributing to open knowledge, and supporting the development of people-centred and technical solutions to help eradicate inequality and bias on the Wikimedia projects.

Summary of key takeaways

1. Information literacy is the ability to use information and communication technologies to find, evaluate, create, and communicate information both online and offline, enabling individuals to make informed decisions. Poor information literacy becomes a particular concern during health emergencies which may be coupled with high rates of disinformation, otherwise known as an 'infodemic'. It is an essential, lifelong competency which should be encouraged across adult communities.

2. Trust is considered to be a cornerstone of community resilience to disinformation, as such it should form the basis of information literacy campaigns. Trust is a process, not an end state, and it should extend beyond the information itself to include messengers, governments, services, and products. This will require sustained collaboration across sectors, before, during, and after crises.

3. Effective information literacy initiatives are those which are tailored to address the diverse needs of various communities. This may require the translation of content into multiple languages as well as contextualising it to meet the cultural and practical needs of different populations.

4. Effective engagement of minority ethnic communities and other underserved groups requires a multi-platform approach, utilising both traditional and digital media to meet the public where they are. Public health bodies will need to support media platforms in disseminating accurate, timely information, while empowering community-based organisations to combat disinformation and deliver trustworthy public health messaging.

5. AI is enabling the proliferation of disinformation in new ways such as deepfakes, unreliable AI-generated news sites (UAINS), and AI chatbots. Therefore, efforts in information literacy must evolve to confront these emerging challenges posed by AI-driven disinformation. Information literacy efforts need to incorporate AI literacy into their strategies. This may include developing a foundational understanding of AI systems, recognising the potential for biased training data, and developing skills to critically evaluate the credibility and trustworthiness of AI-generated content.

Information literacy: What is it?

Information literacy is the ability to use information and communication technologies to find, evaluate, create, and communicate information both online and offline[6,7]. In practice, this includes skills such as fact checking, assessing URLs[8], using reverse image searches[9], and identifying deepfakes[10]. These capabilities can empower people to evaluate online content for reliability, credibility and relevance, helping to build resilience against disinformation or misleading content online. The terms *media literacy*[11] and *digital literacy*[12] were also referenced during this workshop. These concepts share overlapping goals with information literacy.

Information literacy skills can be encouraged through a variety of methods. For instance, the UK Government's Online Media Literacy Strategy[13] addresses the need for enhanced media literacy in the UK, with Ofcom, the UK's communications regulator, playing a central role. Through its media literacy programme, Ofcom conducts research into digital habits and promotes public awareness about online safety through targeted campaigns[14]. In addition to governmental efforts, various programmes in the UK contribute to advancing information literacy. For example, NewsWise, a partnership between the Guardian Foundation, the National Literacy Trust, and the PSHE Association, teaches primary school children to spot fake news, recognise media biases, and create their own news stories, fostering critical thinking from a young age[15].

Moreover, advocacy and counter-disinformation organisations such as Full Fact find, expose and counter disinformation online in an effort to reduce its harm and give the public, journalists and policymakers accurate information[16].

Despite the UK's relatively high media literacy rate and approximately 170 organisations delivering information literacy initiatives[17,18], the needs of adults are often overlooked. Adults are particularly vulnerable to disinformation as they are more likely to be targeted by and susceptible to misleading information than younger people[19]. Claire Levens, Head of Media Literacy Policy at Ofcom, presented Ofcom research on community-based media literacy interventions showing that certain groups – including older adults, minority ethnic groups, non-English speakers, and individuals with mental health challenges or disabilities – remain underserved by media literacy initiatives[20].

Barriers such as lower socioeconomic status, limited internet access, and insufficient digital literacy skills often compound the challenges of reaching minority ethnic communities, as does low awareness or indifference to the need for information literacy. These findings highlight the importance of adopting targeted strategies to address both access challenges and the unique vulnerabilities faced by these groups. Moreover, this suggests that information literacy should be considered as a lifelong learning process[21].

6. Guess et al. 2020 A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proc. Natl. Acad. Sci. U.S.A.* 117 (27) 15536-15545. (http://dx.doi.org/ 10.1073/pnas.1920498117)
7. InfoLit. Definitions & models. See https://infolit.org.uk/definitions-models (accessed 12 June 2024)
8. Polizzi G. 2020 Fake news, COVID-19 and digital literacy: Do what experts do. London School of Economics. 17 June 2020. See https://blogs.lse.ac.uk/medialse/2020/06/17/fake-news-covid-19-and-digital-literacy-do-what-the-experts-do (accessed 25 November 2024).
9. *Ibid*.
10. Microsoft. Spot the Deepfake. See https://www.spotdeepfakes.org/en-US (accessed 25 November 2024)
11. Gov.UK. Online Media Literacy Strategy. See https://www.gov.uk/government/publications/online-media-literacy-strategy (accessed 12 June 2024)
12. American Library Association. Digital literacy. See https://literacy.ala.org/digital-literacy (accessed 12 June 2024)
13. HM Government. 2021 Online Media Literacy Strategy. See https://www.gov.uk/government/publications/online-media-literacy-strategy (accessed 20 January 2025)
14. Ofcom. Media Literacy. See https://www.ofcom.org.uk/media-use-and-attitudes/media-literacy (accessed 12 June 2024)
15. The Guardian Foundation. Newswise. See https://theguardianfoundation.org/programmes/newswise (accessed 12 June 2024)
16. Full Fact. See https://fullfact.org (accessed 12 June 2024)
17. OSIS. BG. 2022 How It Started, How It is Going: Media literacy Index 2022. See https://osis.bg/?p=4243&lang=en (accessed 13 June 2024)
18. *Compared to 41 European countries, using indicators for media freedom, education and trust in people*.
19. Moore R, Hancock J. 2020 Older Adults, Social Technologies, and the Coronavirus Pandemic: Challenges, Strengths, and Strategies for Support. Social Media + Society, 6(3). (http://dx.doi.org/10.1177/2056305120948162).
20. Ofcom. Improving online media literacy in local communities. See https://www.ofcom.org.uk/media-use-and-attitudes/media-literacy/improving-online-media-literacy-in-local-communities (accessed 13 June 2024)
21. Gov.UK. Media literacy uptake among hard to reach citizens. 29 September 2023. See https://www.gov.uk/government/publications/media-literacy-uptake-amongst-hard-to-reach-citizens (accessed 13 June 2024)

In empowering individuals to evaluate online information, information literacy can foster a healthier online environment and build communities that are more resilient to disinformation. As this is an evolving issue, information literacy initiatives require sustained evaluation to monitor their long-term effectiveness. Julian McDougall, Professor in Media and Education at Bournemouth University, presented the Centre for Excellence in Media Practice media literacy evaluation framework which uses a theory of change to help measure the impact of media literacy initiatives for four elements:

- Full and safe access to digital technology and media,
- Critical awareness of media representations and what content and information can be trusted,
- The capability to use their media literacy actively, rather than as passive consumers, and
- The critical understanding of the consequences of their actions in the media ecosystem and how to use their capabilities for positive consequences.

This approach suggests that information literacy should be assessed not just by skills and competencies but by its impact on people's lives, including how they engage with media, adapt their media behaviours, and use media literacy to drive positive actions in the world[22]. This expanded approach to media and information literacy not only equips individuals with practical skills but also nurtures active engagement with the digital world and it lays the foundation for more informed and resilient communities.

Building resilient communities: Lessons from the 'infodemic'

In 2019, UN Secretary-General António Guterres described a global 'Trust Deficit Disorder'[23], a phenomenon that soon spiralled into a global "infodemic" of disinformation during the COVID-19 pandemic[24]. Widespread disinformation during the pandemic exacerbated its impact on public trust in political institutions and contributed to polarisation across many communities. In this context, 'community' refers to groups of people bound by shared identities, languages, cultures, or geographic locations.

These groups often operate outside formal institutional frameworks, media, and information ecosystems. Community resilience to disinformation involves the capacity of these groups to resist, adapt to, and recover from the impacts of false or misleading information which may require strong internal trust networks, improving digital literacy and ensuring access to credible, context-specific information.

Discussions at the workshop focused on how to rebuild and maintain trust with marginalised communities during health emergencies. In her presentation, Dr Tina Purnat, Visiting Senior Scholar at the University of Memphis and Team Lead for Infodemic Management at the World Health Organisation (WHO), highlighted how trust depends not just on the integrity of information but also its messengers, governments, services, and products, and as such requires sustained collaboration across sectors before, during, and after crises. To achieve this, the following framework for trust was suggested:

- **Establish trust before an emergency:** Build trust through consistent, transparent communication from trusted sources and engage the community through outreach and educational campaigns across all demographics. Additionally, providing feedback mechanisms between local communities and government can help to address community concerns and ensure health policies are implemented to reflect public needs.
- **Leverage and maintain trust during an emergency:** Engage trusted local institutions such as healthcare providers, libraries, and community networks to disseminate accurate information. This messaging must be timely, accurate, credible and accessible.
- **Reflect on trust after an emergency:** Post-crisis assessments are useful for identifying where trust was sustained and where it was lost, specifically which partnerships were able to disseminate trusted information and how disinformation exploited vulnerabilities. These assessments can guide improved communication strategies and highlight successful collaborations for future health emergencies, ensuring better preparedness and resilience.

22. Bournemouth University. 2023 Evaluating Media Literacy with a Theory of Change. See https://www.bournemouth.ac.uk/research/projects/evaluating-media-literacy-theory-change (accessed 10 January 2025)
23. Guterres A. 2019. *Remarks at opening of the 74th Session of the UN General Assembly*. See https://www.un.org/sg/en/content/sg/speeches/2019-09-17/remarks-opening-of-74th-session-of-unga (accessed 18 December 2024)
24. Guterres A. 2020. *This is a time for science and solidarity*. See https://www.un.org/en/un-coronavirus-communications-team/time-science-and-solidarity (accessed 18 December 2024)

Dr Purnat noted that building sustained community resilience to disinformation is an ongoing process, which requires adapting to the changing needs of communities and ensuring trust remains a priority long after an emergency subsides. This can be challenging within minority ethnic communities as historical experiences of institutional racism and medical mistreatment have contributed to mistrust of the healthcare system and its interventions[25]. During the COVID-19 pandemic, decreased trust in public health authorities was associated with lower vaccination rates, contributing to poorer health outcomes[26].

Some participants suggested that gaps in public health communication disempowered minority ethnic and non-English speaking communities during the pandemic, leaving them with insufficient information to make informed health decisions. Presentations at the workshop focused on the specific challenges of delivering effective public health messaging to these communities, including the need for linguistically and culturally appropriate messaging and information literacy programmes.

Wikipedia has long been a key resource for people seeking information online and this continued to be the case during the pandemic. Throughout the pandemic, Wikipedia provided extensive, real-time coverage of public health information in over 300 languages, reaching billions of users monthly[27]. This free resource relied on the contributions of Wikimedia's global volunteer community. Dr Netha Hussain, medical doctor and volunteer Wikimedia contributor, detailed the work involved in writing, updating and translating essential public health content, in an effort to bridge the digital gap for non-English speaking communities and tackle disinformation[28].

Community-based groups can also be important providers of public health information in culturally sensitive contexts. Dr Bnar Talabani, medical doctor, clinical lecturer at Cardiff University and co-founder of Muslim Doctors Cymru, shared how their campaign took a comprehensive approach to combating vaccine misinformation among local communities in Wales during the pandemic[29]. Their approach included hosting open dialogues at mosques, addressing concerns at vaccine centres, and collaborating with faith leaders, charities and local businesses to highlight accurate information. The outreach was conducted in multiple languages, in-person and through webinars, and addressed community-specific questions such as whether vaccines are halal or cause infertility[30].

Dr Talabani's work shows the potential of community-based groups to promote information literacy and trust in their communities. Their trusted status and ability to bridge language and cultural gaps could make them key partners for public health institutions; however, their efforts can be limited by lack of resources or coordination with public health authorities. Workshop discussions emphasised the importance of public health bodies such as the NHS and governmental departments working more closely with trusted community groups. By providing resources, support, and opportunities for collaboration, public health bodies can empower these groups to continue the crucial work of building trust and disseminating information in their communities.

25. British Medical Association (BMA). Rebuilding trust in medicine among ethnic minority communities. See https://www.bma.org.uk/news-and-opinion/rebuilding-trust-in-medicine-among-ethnic-minority-communities (accessed 17 December 2024)
26. World Economic Forum. People have lost trust in healthcare systems because of COVID. How can the damage be healed? 25 March 2022. See https://www.weforum.org/agenda/2022/03/trust-health-economy-pandemic-covid19 (accessed 25 July 2024)
27. Wikipedia. Wikipedia and the COVID-19 pandemic. See https://en.wikipedia.org/wiki/Wikipedia_and_the_COVID-19_pandemic (accessed 7 January 2025)
28. Wikimedia Foundation. Responding to COVID-19: How we can help in this time of uncertainty. See https://wikimediafoundation.org/covid19/#section-1 (accessed 15 May 2024)
29. Welsh Government. 2023 St David Awards: Muslim Doctors Cymru. 9 March 2023. See https://www.gov.wales/st-david-awards/muslim-doctors-cymru (accessed 15 May 2024)
30. Cardiff University, Islam UK Centre. 2021 Muslim Doctors Cymru and Covid Vaccine Questions. 9 February 2021. See https://blogs.cardiff.ac.uk/islamukcentre/muslim-doctors-cymru-and-covid-vaccine-questions (accessed 18 December 2024)

Meeting the public where they are: the role of the media

During health crises, access to timely, accurate, and credible information is essential. However, case studies discussed in the workshop revealed that inconsistent public health messaging often led to individuals seeking alternative sources of information, such as social media and private messaging chats, potentially leaving them more susceptible to misleading information.

Parents were among the groups negatively affected by inconsistent messaging during the pandemic, with some being targeted by anti-vaccine campaigners outside the school gates[31]. Vicki Shotbolt, CEO of Parent Zone, a social enterprise that teaches parents and children to navigate the digital world safely, presented findings at the workshop from their 2020 report, *Left behind in Lockdown*[32]. The report found that confusing public health messaging weakened the relationship between parents and schools, with many parents turning to online platforms such as Facebook and Mumsnet or informal WhatsApp chat groups for advice. While these sources often provided immediate access to information, they also exposed users to unverified and potentially misleading content[33].

The pandemic also saw people rely on news and entertainment media, such as television and radio programmes, for critical information. Kamlesh Purohit, Managing Editor at BBC Radio Leicester, noted that during heightened lockdown measures in Leicester in December 2021, people in the South Asian community, particularly non-English speakers and those with limited digital literacy, turned to local radio stations for trustworthy, community-specific information. In response, BBC Radio Leicester and BBC Asian Network tailored their services to meet these needs, providing videos and advice bulletins in multiple South Asian languages to ensure essential information was accessible to these communities[34].

Seeking information from social media, group chats or entertainment media is not a new phenomenon, and it is important to understand how information can travel during health emergencies. Iain Dodgeon, Director at Opening Knowledge across Research & Entertainment (OKRE), a non-profit agency which connects professionals from the entertainment industry and academia, noted that the public engages with information sources not only based on their credibility but also its relatability, and that relatable sources (eg Facebook or a family group chat) often hold more sway. This makes entertainment media a powerful tool for information dissemination. For example, soap operas such as the BBC's *EastEnders* have successfully addressed HIV/AIDS misinformation and stigma through relatable storylines, demonstrating how entertainment can play a strong role in conveying accurate health information to mass audiences[35].

Workshop discussions indicated that to engage the public effectively during health crises, services should meet people where they are, whether that is meeting parents at the school gate or in trusted group chats, or by supporting and leveraging news and entertainment media to communicate public health messages to large audiences. Tailoring communication channels to the needs and preferences of different communities could help to ensure that essential health information is accessible.

The workshop also explored the role of community media in addressing gaps in information access and trust. Community media refers to locally embedded media platforms, including radio stations, online outlets, newspapers and TV channels that serve specific communities. Unlike mainstream media that targets broad national or global audiences, community media focuses on delivering culturally relevant and context-specific content, making it a critical resource for fostering trust and combatting disinformation, particularly within marginalised groups.

31. Schraer, R. 2021 Covid: Misleading vaccine claims target children and parents. *BBC News*. 14 December 2021. See: https://www.bbc.co.uk/news/health-58783711 (accessed 13 January 2025).
32. Parent Zone. 2020 Left Behind in Lockdown. See https://parentzone.org.uk/sites/default/files/2021-12/PZ_Left_Behind_In_Lockdown_2020_0.pdf (accessed 15 May 2024)
33. The Royal Society. 2022 The online information environment. See https://royalsociety.org/topics-policy/projects/online-information-environment (accessed 15 May 2024).
34. BBC Radio Leicester. Gujarati update. See https://www.bbc.co.uk/programmes/p08jjfkw (accessed 15 May 2024)
35. Terrence Higgins Trust. 2023 Eastenders HIV storyline: What you need to know. 16 January 2023. See https://www.tht.org.uk/news/eastenders-hiv-storyline-what-you-need-know (accessed 12 June 2024)

Examples such as the now-disbanded Media Diversified, a UK-based nonprofit media and advocacy organisation[36], illustrate how community media platforms can provide more accessible and relatable information for marginalised groups[37], encouraging community resilience to disinformation. Dr Henna Zamurd Butt, Associate Lecturer at Goldsmiths, University of London, King's College London, and former Editor of Media Diversified, spoke about how minority ethnic communities are often marginalised in mainstream media[38]. These groups may also harbour distrust toward media and government-led communications, stemming from historical patterns of strategic information manipulation and institutional disinformation targeting marginalised populations[39]. However, she noted that community platforms face significant challenges, including underfunding and resource constraints, and a lack of technical skills to address disinformation.

Recognising the importance of these platforms, the workshop aimed to include representatives from community media outlets to share perspectives on tackling disinformation and promoting information literacy, particularly those producing non-English language media such as Channel S, which serves the British-Bangladeshi community, and 92 News UK, a UK-based Pakistani news channel. However, efforts to engage these organisations proved difficult, and none were able to attend this workshop. Ensuring the meaningful inclusion of community media in these discussions is essential for developing solutions to disinformation. A priority for future research in this space should be proactive and have sustained engagement with these groups.

AI and the changing face of disinformation: adapting for future health emergencies

The World Economic Forum's Global Risks Report 2024[40] named disinformation as a top risk, with the potential to destabilise societies and undermine democratic processes[41]. The report emphasises the urgent need for action as generative AI, a machine learning system that can generate new text, images, audio, or video in response to user input[42], may increase both the volume and credibility of falsehoods in the information ecosystem[43]. While the full extent of AI's role in shaping disinformation remains unclear, recent examples from various 2024 elections indicate that AI-generated disinformation is a growing risk that necessitates expanding the scope of information literacy to include AI literacy.

Workshop presentations and discussions highlighted key concerns for disinformation during future health emergencies, including unreliable AI-generated news sites (UAINS), AI-generated deepfakes and consumer use of large language models (LLMs).

Unreliable AI-generated news sites

Generative AI systems that can create new text, images, audio or video in response to user input using machine learning techniques[44] are increasingly impacting the trustworthiness of news information, particularly through the rise of unreliable AI-generated news sites (UAINS)[45]. Veena McCoole, Head of Communications and Marketing at NewsGuard, a company that monitors and rates news sites for their transparency and reliability, highlighted how UAINS impacts the information ecosystem with low-quality or false content.

36. Media Diversified. See https://mediadiversified.org (accessed 18 December 2024)
37. Beazer A, Walter S, Eldridge, SA, Palicki SK. 2023 On the Margins: Exploring Minority News Media Representations of Women during the COVID-19 Pandemic. *Digital Journalism*, 1–20. (http://dx.doi.org /10.1080/21670811.2023.2206039)
38. Quintero Johnson JM, Saleem M, Tang L, Ramasubramanian S, Riewestahl E. 2021 Media Use during COVID-19: An Investigation of Negative Effects on the Mental Health of Asian versus White Americans." *Frontiers in Communication* 6. (http://dx.doi.org 10.3389/fcomm.2021.638031)
39. Marwick A, Kuo R, Jones Cameron S, Weigel M. 2021 Critical disinformation studies. Center for Information, Technology, and Public Life (CITAP). See https://assets.pubpub.org/affm8wbg/31651593248343.pdf (accessed 18 December 2024)
40. World Economic Forum. 2024 Global Risks Report 2024. 10 January 2024. See https://www.weforum.org/publications/global-risks-report-2024 (accessed 20 December 2024)
41. World Economic Forum. 2024 The big election year: how to protect democracy in the era of AI. January 29 2024. See https://www.weforum.org/stories/2024/01/ai-democracy-election-year-2024-disinformation-misinformation (accessed 20 December 2024)
42. The Alan Turing Institute. Defining data science and AI. See: https://www.turing.ac.uk/news/data-science-and-ai-glossary (accessed 20 December 2024)
43. Ryan-Mosley T. 2023 How generative AI is boosting the spread of disinformation and propaganda. *MIT Technology Review*. 4 October 2023. See https://www.technologyreview.com/2023/10/04/1080801/generative-ai-boosting-disinformation-and-propaganda-freedom-house (accessed 25 July 2024)
44. The Alan Turing Institute. Defining data science and AI. See: https://www.turing.ac.uk/news/data-science-and-ai-glossary (accessed 20 December 2024)
45. WebStat. 2024 Proliferation of AI-Generated News Sites Raises Concerns Among Analysts. 11 December 2024. See https://webstat.net/ai-fake-news/proliferation-of-ai-generated-news-sites-raises-concerns-among-analysts (accessed 20 December 2024)

NewsGuard has identified over 500 such sites, many of which plagiarise credible content[46]. These trends threaten to further undermine trust in journalism, making it harder for the public to discern reliable information. These risks are amplified by the growing sophistication of generative AI, which has been used to spread state-sponsored propaganda, healthcare hoaxes, and false narratives about global conflicts[47].

Deepfakes

Deepfakes are hyper-realistic synthetic audio, video or imagery created by machine learning algorithms[48]. As powerful generative AI systems gain prevalence, deepfakes have also become more widespread and easier to create. A 2023 research experiment published by the Royal Society found that only 22% of participants were able to correctly identify a deepfake when asked to pick it out from a set of videos[49]. During health emergencies, such hyper-realistic images and videos can easily spread disinformation and be difficult for the average viewer to detect, even with content warnings[50]. Deepfakes have also been used to imitate political leaders or health officials to spread false or dangerous disinformation. For example, in the United States, a deepfake of the former Chief Medical Adviser, Professor Anthony Fauci ForMemRS, was used in a campaign video by a candidate in the Republican Presidential nomination race.[51] This tactic could easily extend to future health emergencies and risks further eroding trust in public health information[52].

To mitigate the impact of AI-generated content and encourage information literacy, the BBC has adopted proactive measures. Flora Carmichael, Editor of BBC Trending, discussed the work of the BBC Verify team[53], who combine human fact-checking, video verification, and the Coalition for Content Provenance (C2PA) technical standard to trace the origin of different types of media[54].

These tools not only help journalists to discern credible news sources, but also explain the verification process to the public and teach them about the importance of evaluating content provenance. By fostering informed decision-making and reinforcing trust in the information ecosystem, these efforts contribute to equipping individuals with the new information literacy skills needed to navigate a complex digital landscape.

Consumer use of large language models

While AI is reshaping the media landscape, it is also increasingly being used in place of online search engines or other consumer research tools. Large language models (LLMs) such as ChatGPT and Gemini have increased the distance between consumers and original information sources as outputs are synthesised from vast datasets rather than being directly attributed to specific, verifiable sources. This lack of transparency can make it easier for disinformation to spread unnoticed, as people might trust or share the content without questioning its accuracy, authenticity or intent. At an October 2023 workshop hosted by the Royal Society, health and climate postgraduate students demonstrated this risk by red-teaming a LLM to produce scientific disinformation, highlighting how personal use of AI tools for information gathering during health emergencies can perpetuate false narratives[55].

46. NewsGuard. Misinformation Monitor: December 2023. See https://www.newsguardtech.com/misinformation-monitor/december-2023 (accessed 12 June 2024)
47. NewsGuard. Overview. See https://www.newsguardtech.com/special-reports/ai-tracking-center (accessed 12 June 2024)
48. The Alan Turing Institute. Defining data science and AI. See: https://www.turing.ac.uk/news/data-science-and-ai-glossary (accessed 18 December 2024)
49. Lewis A, Vu P, Duch R M, Chowdhury A. 2023 Deepfake detection with and without content warnings. *Royal Society Open Science*/ 10,11 (http://dx.doi.org/10.1098/rsos.231214)
50. *Ibid.*
51. Vincent, J. DeSantis attack ad uses fake images of Trump embracing Fauci. *BBC News*. 8 June 2023. See: https://www.theverge.com/2023/6/8/23753626/deepfake-political-attack-ad-ron-desantis-donald-trump-anthony-fauci (accessed 13 January 2025).
52. World Economic Forum. 2024 Global Risks Report 2024. 10 January 2024. See https://www.weforum.org/publications/global-risks-report-2024 (accessed 20 December 2024)
53. BBC News. BBC Verify. See https://www.bbc.co.uk/news/reality_check (accessed 12 June 2024)
54. Coalition for Content Provenance and Authenticity (C2PA). See https://c2pa.org (accessed 12 June 2024)
55. The Royal Society. 2024 Red teaming large language models (LLMs) for resilience to scientific disinformation. See https://royalsociety.org/news-resources/publications/2024/red-teaming-llms-for-resilience-to-scientific-disinformation (accessed 13 June 2024)

These risks are compounded by AI's ability to exploit societal biases and amplify divisions. Dr Stefanie Felsberger, Research Associate at the Minderoo Centre for Democracy and Technology, explained that AI-driven disinformation campaigns often manipulate public opinion by targeting fault lines in race, gender and class. With AI models often trained on data from WEIRD (White, Educated, Industrialised, Rich and Democratic) populations, systemic harms inherent to biased data can be perpetuated, harming diverse populations seeking information during health emergencies[56]. Dr Henna Zamurd Butt highlighted the need to expand our understanding of disinformation[57] to account for the 'echoes of colonialism' embedded in AI systems[58], which may produce harmful health guidance for minority communities.

Dr Julia Slupska, Head of Policy, Research and Campaigns at Glitch, which campaigns against online abuse, pointed out that minority ethic groups, particularly Black women, are 34% more likely to be the target of online abuse and disinformation[59]. Without intentional intervention, AI tools risk worsening the inequalities felt during health emergencies[60]. Workshop participants suggested that organisations focused on information literacy should integrate aspects of AI literacy into their strategies. This may include a foundational understanding of AI systems, such as data inputs, algroithms and biases. Moreover, AI literacy could equip people to critically engage with AI outputs to evaluate their reliability and trustworthiness.

Conclusion

The challenges of disinformation, trust, and the rapidly evolving information landscape highlight the need for adaptable approaches to information literacy. As AI-based technologies advance and societal dynamics shift, strategies for information literacy must adapt to ensure individuals and communities are equipped with the tools to critically engage with information and build resilience to misleading content. Collaboration among governments, media, community groups, and educators will be essential to ensuring that information literacy remains a cornerstone of public resilience and trust, both in times of crisis and beyond.

56. Royal Society and Department for Science, Innovation and Technology workshop on horizon scanning AI safety risks across scientific disciplines, October 2023. See https://royalsociety.org/current-topics/ai-data (accessed 13 May 2024)
57. Marwick A, Kuo R, Jones Cameron S, Weigel M. 2021 Critical disinformation studies. Center for Information, Technology, and Public Life (CITAP). See https://assets.pubpub.org/affm8wbg/31651593248343.pdf (accessed 18 December 2024)
58. ManyfestoAI. AI Decolonial Manyfesto. See https://manyfesto.ai (accessed 18 December 2024)
59. Glitch. 2023 Glitch x EE Hope United: Online Safety Bill Campaign. *Medium*. 13 April 2023. See https://medium.com/@glitchuk_/glitch-x-ee-hope-united-online-safety-bill-campaign-fa047430f508 (accessed 12 June 2024)
60. Royal Society and Department for Science, Innovation and Technology workshop on horizon scanning AI safety risks across scientific disciplines, October 2023. See https://royalsociety.org/current-topics/ai-data (accessed 13 May 2024)

Attendees

The Royal Society would like to thank the following people for supporting this event.

Event participants
Maryam Ahmed, BBC
Nicola Aitken, Meta
Ros Atkins, BBC
Ruth Carlyle, NHS England
Flora Carmichael, BBC Trending
Katherine Chadd, Parent Zone
Iain Dodgeon, OKRE
Stefanie Felsberger, Minderoo Centre for Technology and Democracy
Nathan Forrester, Wikimedia Foundation
Blathnaid Healy, BBC
Jonathan Heawood, Public Interest News Foundation
Netha Hussain, Wikimedia Foundation
Charles Inskip, University College London
Emily Jewell, Arts Council England
Liz Jolly, The British Library
Sue Lacey-Bryant, NHS England
Claire Levens, Ofcom
Alba MacGillivray, Centre for Information Resilience
Veena McCoole, NewsGuard
Julian McDougall, Bournemouth University
Liz Moorse, Association for Citizenship Teaching
Chris Morris, Full Fact
Carissa Mwananshiku, LadBible Group
Nalini Naidoo, Parent Zone
Jenny Niven, Edinburgh International Book Festival
Sarah Parkes, Age UK
Luis Pereira, City, University of London
Hannah Perry, Demos
Tina D Purnat, University of Memphis
Kamlesh Purohit, BBC Radio Leicester
Franziska Putz, Wikimedia Foundation
Sophie Randall, Patient Information Forum
Jane Secker, City, University of London
Vicki Shotbolt, Parent Zone
Julia Slupska, Glitch
Bnar Talabani, Cardiff University
Glen Tarman, Full fact
Sheila Webber, University of Sheffield

Event participants (continued)
Catherine Walters, Parent Zone
Henna Zamurd Butt, Goldsmiths, University of London and King's College London

Royal Society staff
Nicole Mwananshiku, Policy Advisor, Data and Digital Technologies
Areeq Chowdhury, Head of Policy, Data and Digital Technologies
Denisse Albornoz, Senior Policy Advisor, Data and Digital Technologies (until June 2023)
James Hannaford, Programme Coordinator, Data and Digital Technologies
Charise Johnson, Policy Advisor, Data and Digital Technologies

BBC staff
Laura Ellis, Head of Technology Forecasting
Bill Thompson, Head of Public Value Research

Wikimedia UK staff
Lucy Crompton-Reid, Chief Executive
Daria Cybulska, Director of Programmes and Education