Produced by the AI Governance Pillar at AI Singapore, now part of the NUS Artificial Intelligence Institute

AI Governance Roundtable #4: Mis-, Dis-, and Mal-information with AI

Can and should industry and regulators distinguish between AI- and human-generated content (through watermarks, etc.)? How can industry work with the government to manage mis-, dis-, and mal-information produced by generative AI?

This is the fourth of a series of roundtables convened by AI Singapore for representatives from industry, government, and academia to discuss responsible AI. Such discussions are typically too narrow and too broad. Too narrow in that a few voices dominate the discussion – notably those in the United States and Europe, with China sometimes included. Too broad in that discussion is often limited to generalities and principles. This project aims to address both aspects of this problem, involving a wider set of stakeholders — in particular those from Southeast Asia — in more focused discussions of specific challenges in the application of Responsible AI to particular questions.

Rapporteur: Tristan Koh Ly Wey and Dawei Chen, AI Singapore
Location : Meta Singapore

1 Introduction

The problem of fake news is not new. Despite Donald Trump's occasional assertions that he has coined the term[1], "fake news" was featured as a headline in the *New York Times* over a century ago[2]. However, the emergence, advances and proliferation of generative artificial intelligence (genAI) capable of generating seemingly new content in the form of text, images, videos or other data has brought the age-old problem of fake news back to the forefront of public discussion. This problem of fake news could be seen as a subset of the problem of mis-, dis- or mal-information. **Misinformation** refers to false information spread

[1] Callum Borchers, 'Trump Falsely Claims (Again) that He Coined the Term "Fake News"', *Washington Post* (26 October 2017).
[2] Capt. Bartlett, '"Fake" News for Spain', *New York Times* (6 October 1901).

without malicious intent. **Disinformation** involves false information created and disseminated with malicious intent. **Malinformation** is true information used with harmful intent. This roundtable mainly focused on misinformation and disinformation.

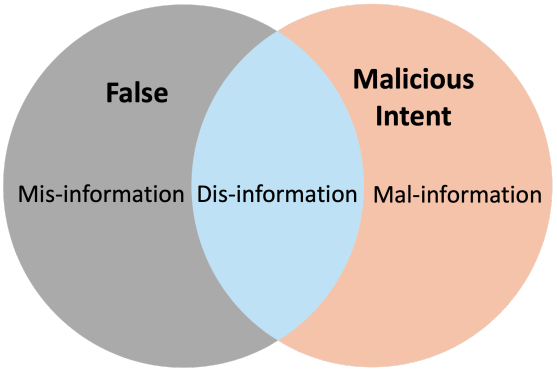

Figure 1. Categorization of Information Disorder[3]

GenAI impacts the state of information disorder in three ways. First, genAI could significantly increase the **quantity** of misinformation as the cost and effort to mimic human generated content is reduced. GenAI users can easily create content mimicking different styles, languages, and more[4]. Second, genAI could increase the **quality** (i.e., persuasiveness, fraudulence, difficulty of discerning from real content) of misinformation because genAI is capable of learning statistical patterns from massive real data, and even personalise misinformation to specific vulnerable individuals. Third, apart from genAI, other AI technologies (e.g., recommendation systems) could speed up the **dissemination and consumption** of misinformation such as the creation of echo chambers.

Considering these impacts, genAI might **challenge existing mitigation strategies**. Manual validation of content veracity and quality cannot cope with the increasing scale of content. Automated technical methods that detect misinformation might have **low reliability** and are easily circumvented and tampered by bad actors. Furthermore, the **virality of misinformation** combined with increased speed of dissemination poses challenges for law enforcement and harm mitigation efforts.

For the general public, misinformation may become more readily available and harder to distinguish. Relying on such misinformation results in **harm** and **reduces trust in institutions and platforms**. Consequently, **identifying misinformation** is no longer confined to institutions and experts that have traditionally been concerned about misinformation. Taken

[3] The figure is adapted from Claire Wardle and Hossein Derakhsha. "Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making." Council of Europe, DGI(2017)09, September 2017.

[4] Feuerriegel, et al. "Research Can Help to Tackle AI-generated Disinformation." *Nature Human Behaviour* 7, no. 11 (2023): 1818-1821.

to an extreme, the "**Liar's Dividend**" could emerge, where the abundance of misinformation **diminishes the value** of true information because identifying misinformation becomes a futile effort for the average individual. Resultantly, bad actors that produce fake news can more easily **escape responsibility**.

To tackle these issues, solutions can block, lower the speed of transmission, add context to or educate consumers about misinformation. Other solutions may involve distinguishing between trustworthy and untrustworthy, or harmful and harmless content at both the creation and distribution stage of information.

The AI Governance pillar of AI Singapore convened a roundtable with industry, government, and academia to discuss ***whether industry & regulators can and should distinguish between AI- and human-generated content, and how can industry work with the government to manage mis-, dis-, and mal-information produced by genAI****?*

2 Understanding the Problem

Understanding the problems with AI-generated misinformation first involves a discussion of the ***comparative importance*** *and* ***relevance of provenance, truth, trust, and harm*** *of AIGC in relation to HGC.* Investigating these issues provides clearer rationales for downstream solutions.

2.1 What is AIGC?

There is a spectrum of AIGC. AIGC is typically understood as content produced by genAI tools like ChatGPT, Claude, or Gemini, where text prompts generate images or other outputs. However, there are other types of content **modified by AIGC to varying degrees**. One example is Zoom filters where built-in functions are used to enhance profile photos or add a virtual background. Should the resulting video be considered AIGC? Another category is human-edited content, such as images refined using Photoshop. Finally, there is purely human-generated content, such as text written without any AI assistance or an unedited original photo. The **line between AIGC and HGC** might depend on what type of AI tool is used and how it is applied in the content creation process. For instance, an image originally generated by genAI tools, even if followed by manual editing, may be more likely to be considered AIGC. Conversely, content originally created by a human and then edited by genAI tools may be more likely to be considered as HGC.

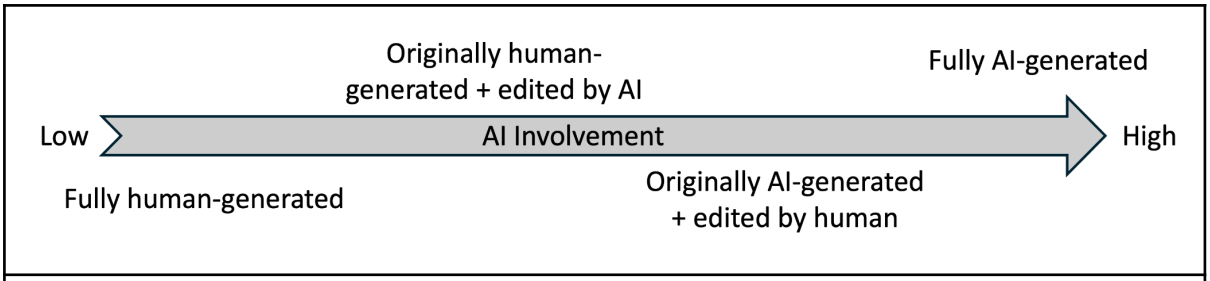

Figure 2. Content Categorization

2.2 Provenance vs. Truth

An approach to mis- and dis-information focuses on determining **whether content is truthful rather than whether it is AI-generated** (i.e. provenance). Conceptually, there is no inherent connection between provenance and truthfulness. AIGC could be either true or false. However in practice, the truthfulness of AIGC is perceived to be **more difficult to discern** compared to human-generated content because of the increased volume and perceived quality which has **no necessary relation** to truth. As one researcher puts it, LLMs are "incidental truth engines"[5]. Thus, there could be possible correlations between truth and provenance along the spectrum of AIGC types.

2.3 Truth vs. Trust vs. Harm

Other than tracking provenance, there are interrelated **issues of trust and harm that are influenced by truth and provenance**. Trust is a broad concept. **Digital trust** has been defined as "the expectation by individuals that digital technologies and services - and the organisations providing them - will protect all stakeholders' interests and uphold societal expectations and values"[6]. Within the context of misinformation, untrue information can sometimes lead to **harm for society and individuals**, and therefore **reducing trust**.

Trust could also refer to **trusted sources of information** such as public institutions or reputable news outlets where **truth and transparency** are highly valued. For the **government**, truth is important to maintain public trust for various reasons. Truth and transparency enables citizens to make **informed decisions** about their lives and ensures **democratic accountability**. For **news outlets and social media platforms**, these serve as main sources of **information and discussion** for the public. They have the **responsibility** to ensure that their platforms are truthful sources of information. To improve transparency and trust, these platforms typically disclose information like **content sources**, and implement crowd-based and platform **verification** which provide relevant details for individuals to assess and consume content.

---

[5] Sandra Wachter, Brent Mittelstadt and Chris Russell, "Do Large Language Models Have a Legal Duty to Tell the Truth?" *Royal Society Open Science* 11, no. 8 (2024).

[6] Kate Whiting, "Social Media in the Crossfire: This is How You Establish 'Digital Trust'", *World Economic Forum* (20 February 2024).

Within these contexts, truthfulness is positively correlated with trustworthiness and harmlessness because **truthful sources can be relied upon, reducing the likelihood of harm caused by misinformation**. However, with AIGC, even information from traditionally trusted sources may be false because of the issues mentioned above. This perceived difficulty in discerning truth indicates that consumers might need to be **more cautious** when relying on traditional trusted sources of information because AIGC potentially makes consumers more vulnerable to harm caused by mis- and dis-information.

2.4 Impact of AI-generated Misinformation

GenAI **enables** harms traditionally caused by human-generated misinformation by rendering traditional enforcement mechanisms less effective. Some harms are identified below.

Level of negative impact	Description
Societal	**Reduced trust** in public institutions, national security concerns and crime enabled by genAI. Some examples: ● Influencing voter opinion, political disinformation campaigns, radicalisation and incitement of violence ● Generation of child sexual abuse material, deepfakes of notable personalities used for fraud[7]
Company / Organisational	Creates **reputational, compliance & legal uncertainty and risks**. Some examples of risks: ● Content creators may be hesitant to use genAI tools because of concerns that they may be held responsible for any inaccuracies and insensitivities in AIGC ● Potential vicarious liability of employers when their employees produce AIGC that harms the company ● Company leadership unsure of the minimum threshold for legal & compliance when adopting genAI tools
Consumer	**Reduced trust** in institutions and companies, overt and subtle **manipulation** of consumer behaviour, values and beliefs. Some examples: ● Scams and fraud ● Influencing consumer behaviour to further a covert objective by a bad actor

[7] One example was then-Prime Minister Lee Hsien Loong's deepfaked video encouraging viewers to invest in a crypto scam.

These negative impacts do not diminish the fact that genAI can be **a force for good**, such as helping content creators better articulate their vision and provide productivity gains for companies. As ever, the problem is how to mitigate the negative effects while maximising the potential of the technology.

3 Potential Solutions

Solutions to combat misinformation can come in at both the creation and distribution stage of misinformation. **At creation**, there are solutions that track the provenance of information. Even if the information was AI generated, the information may not be false, untrustworthy or harmful. Only with further context **at distribution** can the information be more likely deemed as false, untrustworthy or harmful.

Solutions can include existing methods that have been traditionally designed for misinformation before genAI**.** While these methods similarly apply to AI-generated misinformation, they may require supplementing with new solutions to combat the speed and quality. Taking a **risk-based approach** to select and design appropriate solutions could allow for a more holistic strategy.

Type of solution	Description
Add context	**Tracking provenance** at creation, then disclosing to consumers at distribution
Educate	Educating consumers about the **responsible use** of genAI at creation, and **identifying misinformation** at distribution Educating law enforcement on **genAI-enabled crimes** (e.g. scams & fraud, disinformation campaigns) for more effective crime detection and enforcement
Slow down transmission	Mechanisms and guidelines that detect and mitigate misinformation **before it becomes viral** Regulations that require information platforms like social media to **disclose** details of recommendation algorithms and **compulsory cooperation** with governments when material shared on platform is potentially illegal
Block or correct	Community standards defined along **specified categories of harm** that allow the takedown of content

	Tech-agnostic regulations that allow **government takedown or correction** of content when content causes specific harms (e.g. detrimental to public interest, reputational damage, foreign interference) Technical solutions that block false content from being generated during content creation such as **reducing hallucinations** or **introducing digital noise** which prevents re-use of content

3.1 Add Context

AI governance frameworks from Australia, China, Japan, South Korea and Singapore highlighted the need for mechanisms to enable stakeholders (e.g. regulators and public) to **identify AIGC**. Such identification methods are usually developed for **tracking and disclosing provenance** to add context to information. The prevailing method of adding context typically involves technical solutions that track the provenance at creation. This provenance is then disclosed to consumers when the content is distributed.

3.1.1 Tracking provenance at creation

Metadata and invisible watermarks are used to track the **creation and modification of content**. Content platforms can detect these provenance signals to identify content origins/modifications and display relevant labels (e.g., AI-generated) to consumers. To cope with the volume of content, these technical solutions track the provenance through various modalities[8].

- **Watermarking** are visible or invisible "stamps" that show ownership of content. Google's SynthID is one such instance of watermarking, being able to embed the watermark directly into the pixels of the image.
- **Fingerprints** represent the data in a more searchable form that allows quick matching of the data to a verified database of information. Digital Rights Management systems use such technology to verify that content is not infringing copyright.
- **Cryptographically-hashed metadata** creates a small and unique representation of the metadata ("the hash") using cryptographic hash functions. If the metadata is changed, the hash will no longer match the original. The C2PA content manifest is one such example of cryptographically-hashed metadata. Some have suggested that blockchain technologies can also be used to track and store metadata.

Given the variance and nascency of such technical solutions, some challenges remain:

[8] Laura Ellis, "Increasing Trust in Content: Media Provenance and Project Origin." *BBC* (24 October 2023).

- Who should **apply** the stamps, and **when**? Should it be the responsibility of content creators or distributors?
- How can the stamping process be **standardised**? While it is easier to stamp and label in-house produced content, detecting and labelling AIGC from third-party tools can be more challenging. The collaboration and a **standardised metadata framework** could help mitigate this issue which may also reduce future costs of compliance. One such example is the specification developed by C2PA (Coalition for Content Provenance and Authenticity)[9].
- Even without genAI, bad actors have always **intentionally and strategically circumvented fact-checking.** How can we prevent bad actors from fraudulently modifying or manipulating metadata?
- While it is relatively straightforward to add stamps to images, is it possible to apply metadata to **text, videos, and audio**?
- How can we **reduce false positives**, where images that are simply touched up with Photoshop mistakenly trigger the AI-generated content label?

Other provenance tracking methods include encouraging content creators to **disclose AI involvement** in their content creation process. For example, before posting content relating to political and social issues, content creators have to disclose to the platform their identity and whether AIGC was used. This self-disclosure method can be made more detailed if necessary. Some social media platforms also partner with **third-party fact checking** organisations to validate AIGC content.

3.1.2 Disclosing provenance at distribution

After detecting metadata and classifying the content, the next step is determining what information should be made transparent and how to label content appropriately.

From the perspective of **provenance and truth**, disclosure of the provenance of content may be relevant as a signal for truthfulness. With genAI, traditional solutions like fact-checking and editorial controls may not be as effective. Labelling AIGC allows consumers to be more circumspect about such information.

From the perspective of **truth, trust and harm**, disclosing provenance could increase trust, especially when the truthfulness of AIGC is hard to discern. Knowing whether the content is AI-generated could provide people with **relevant information and the autonomy** to evaluate its significance and make informed decisions. For example, in situations where a principal pays an agent to create a logo or marketing campaign, it is important for the principal to know if genAI was used in the creation process in case of any mis-credit issue.

[9] C2PA, "Coalition for Content Provenance and Authenticity".

Additionally, for legitimate content creators and platforms, providing labelling is a means of **transparency** that can build trust, demonstrate proactive consumer protection, and avoid potential reputational blowback. This is particularly important since some consumers remain wary of AIGC despite its growing acceptance. From a transparency perspective, this avoids making judgement calls about subjective issues such as trustworthiness, factuality or harmfulness by leaving such value claims to the user. Apart from giving consumers an **informed decision**, tracking provenance is also a good **data governance strategy** to mitigate future copyright issues.

3.1.3 Behavioural problems with adding context

Such labels are still open to interpretation. Given the volume of synthetically generated information, **over-labelling** could occur, where people might come to believe that anything unlabeled is true, while anything labelled is untrue. Consumers may not know how to meaningfully interpret information that is labelled AIGC vs unlabelled information (which can include HGC). Some consumers may interpret content labelled as AIGC untrustworthy, and therefore the content should be treated with circumspection.

However, **AIGC is not necessarily unreliable**, just like how unlabelled or HGC is not always reliable. Given the ambiguity of such labels, the uptake of genAI tools may also be affected. For instance, almost 30% of ChatGPT users reportedly said that they would use ChatGPT less if ChatGPT included watermarks[10]. Therefore, focusing solely on disclosure of provenance is neither a necessary nor sufficient indicator of mis- and dis-information.

3.1.4 Labelling trustworthiness

Alternatively, solutions can label & disclose **trustworthy sources of information** rather than provenance. One example is how SMS-es sent by the Singapore government will be sent from "gov.sg" to indicate that it is a trusted source of information[11]. Having such policies may encourage content creators to fact-check their content by using **"human-in-the-loop" approaches** before distribution regardless whether it is AIGC.

However, moving from labelling provenance which is an objective fact to **subjective notions** of trustworthiness or harmfulness of information can be problematic. Should the government, organisations or consumers decide whether some pieces or sources of information are more trustworthy or harmful than others? On one hand, **consumers should have the right** to decide subjective notions of "trustworthiness" or "harmfulness" for themselves. This approach underpins current solutions that aim to provide objective facts as context such as provenance. At the same time, purely relying on consumer choices may result in **wider negative and public consequences on the value of truth**. For instance,

[10] Wes Davis, "OpenAI Won't Watermark ChatGPT Text Because Its Users Could Get Caught." *The Verge* (5 August 2024).
[11] Singapore Government, "What You Need to Know: gov.sg SMS Sender ID." (13 June 2024).

consumers may still wish to consume and propagate misinformation for its entertainment value despite knowing that the content is untrue, which may impinge on government institutions and social media platforms' status as reliable sources of information.

Therefore, given the nascency of technical solutions labelling provenance, as well as subjectivity and ambiguity of interpretation of such labels, solutions that add context should be considered in conjunction with other solutions.

3.2 Educate

Assuming that malicious actors that aim to spread disinformation are a small minority, **educating consumers** is an important step towards the reduction and identification of misinformation. Such education should involve guidance on **using genAI responsibly when creating content**, including risks of relying on genAI caused by problems such as hallucinations. Consumers should also be trained to **identify misinformation at distribution**. For instance, the government Scamshield Bot allows consumers to report and check suspicious messages on Whatsapp[12]. In particular, both government and industry should reach out and educate vulnerable populations such as the digitally less literate. Education also feeds into solutions that add context by equipping consumers with awareness to interpret provenance or truthfulness labels meaningfully.

Other than educating consumers, **education for law enforcement** is also equally important. There are a host of genAI-enabled misinformation-related crimes such as fraud and scams, election and geopolitical disinformation campaigns, child sexual abuse material and pornography[13]. GenAI enables more subtle social engineering methods to be used to further these crimes. Law enforcement must continue to remain apprised of how genAI can be exploited by malicious actors.

3.3 Limit Speed of Transmission

Given the rapid volume and speed of generation enabled by genAI, these are solutions that try to limit distribution. In industry, these can include measures that mitigate the potential for misinformation becoming viral content. One example are methods to identify and mitigate **inauthentic behaviour** that "artificially boost the popularity of content" to deceive others[14]. Another example is **account control measures** such as banning accounts that exceed the limits of how many times a message can be forwarded. For instance, Whatsapp blocked 2 million accounts in India, with 95% exceeding the limit of the number of times that messages can be forwarded[15].

---

[12] Open Government Products, "ScamShield".
[13] Claire Meyer, "Putting Generative AI to Use for Crime: Fraud, Disinformation, Exploitation, and More." *Security Management* (17 June 2024).
[14] Meta, "Inauthentic Behaviour", *Transparency Center*.
[15] BBC, 'Whatsapp Blocks Two Million Indian Accounts' (16 July 2021).

The Digital Services Act enacted in the EU in 2023 gives consumers the **option to turn off recommendation algorithms** that are based on their personal information like race and ethnicity, which may **limit the spread** of certain types of misinformation that may be hyper personalised to certain demographics. Intermediary service providers, which include online platforms and search engines, **must cooperate with relevant national authorities** to act against illegal content and disclose information relating to recommendation algorithms and content moderation policies. Such regulatory levers allow governments to **take mitigatory action tailored to the demographics** that were targeted by misinformation, hopefully reducing the spread and harm.

3.4 Block or Correct

These are methods that block or correct misinformation, typically because the **misinformation causes certain harms to individuals or society**. These methods target the truthfulness of the content and the potential to cause harm rather than trustworthiness.

Legislation such as the Foreign Interference (Countermeasures) Act (FICA) allows the Singapore government to **takedown content, or suspend and disable the account** if there is an ongoing hostile information campaign[16]. The Protection from Online Falsehoods and Manipulation Act (POFMA) allows the Singapore government to **insert a notice to the original post with falsehoods** with a link to the government's clarification[17]. In industry, social media **content moderation standards** allow the takedown of content when content is in violation of certain guidelines such as when the where the information is likely to directly contribute to the risk of imminent physical harm or interfere with the functioning of political processes[18].

Apart from these ex-post guidelines that target misinformation only after it has been distributed, there are **LLM-specific technical strategies** that happen **ex-ante** closer to the model development stage before the information is distributed by the user. One solution is adding digital noise to AIGC so that the generated content is unable to be exploited for other AIGC. Other solutions involve techniques to reduce LLM hallucinations such as retrieval augmented generation or leveraging content filtering APIs to filter out potentially untruthful generated content.

Overall, given the complexity of the problem, it is likely that **a combination of these strategies would be used**. Using a risk-based approach, mitigation strategies can be combined and designed at relevant levels of intervention both by government and industry.

---

[16] Ministry of Home Affairs, "Introduction to Foreign Interference (Countermeasures) Act."
[17] POFMA Office, "Protection from Online Falsehoods and Manipulation Act."
[18] Meta, "Misinformation." *Transparency Center*.

3.5 Incentives for Regulation

Given the risks that misinformation poses to the public, **regulation can be justified** as seen with POFMA, FICA and the Digital Services Act. Other non-misinformation specific regulations also cover harms that could arise from misinformation, such as the Protection from Harassment Act that protects individuals from false statements of fact regardless of the medium used[19]. However, the development of tech is very **likely to outpace existing regulation** or technical solutions. GenAI content is becoming more realistic and more likely to become viral with the introduction of video generation models. **Enforcement will be increasingly ineffective** because of the difficulty of detecting such content.

While most regulatory frameworks are tech-agnostic, in light of the volume and speed of transmission of genAI misinformation, there could be sufficient justification in the public interest to **extend** current regulatory frameworks and enforcement mechanisms to also **target the creation of information** using a wider range of solutions identified above. One view is that companies can use such regulation to justify to their shareholders to invest in policies and tech R&D that combat misinformation.

However, others think that a light-handed approach such as **self-enforcement** in consultation and in sync with governments globally may be more conducive. To this end, industry has initiated projects such as the C2PA started by a consortium of tech companies. Industry also pledged to combat genAI misinformation specifically for elections-related context during the Munich Security Conference 2024[20]. Apart from collaboration between industry and government, companies within industry should also work collectively on solutions before the government regulates, making it **more costly for compliance**.

4 Next Steps

Unlike areas such as data protection, misinformation does not seem to have undergone a similar "Brussels effect". Though the EU has enacted the Digital Services Act that imposes disclosure obligations and more with extra-territorial jurisdiction, equivalent regulation outside the EU has not been uniform. Given the diversity, subjectivity, and complexity of genAI misinformation, each jurisdiction may eventually adopt **unique approaches**.

Another frame of reference could be to learn from industries that have **placed a high premium on truth and disclosure** to inform consumers who have a "right to know". For instance, advertising industry best practices mandate that a piece of content should be labelled as advertorial if the author was writing on behalf of an advertiser. The US Association of National Advertisers has updated their standard form contracts such that advertising agencies have to declare that genAI was used to generate content.

[19] Singapore Courts, "Cases Eligible for Protection From Harassment."
[20] Munich Security Conference, "AI Elections Accord."'

Another question is how **open-source genAI tools** fit into the issues and solutions identified above. Open-source tools may not be part of provenance standards that governments and industry may develop and endorse in the future. However, for AIGC to become harmful misinformation, such misinformation still needs to be distributed through popular platforms that are read by many consumers and already have existing safeguards that apply regardless of the type of technology used.

Apart from the immediate implications of genAI misinformation, there are **broader long term concerns** about the impact of AIGC on individuals, organisations, platforms, and society. For example, behavioural issues regarding the perception, consumption and response of consumers to AI-generated misinformation can be further studied in relation to existing research of human-generated misinformation (e.g. digital literacy, conspiracy theories, and social network homophily). Further, given that costs of producing quality and creative content have been driven down, how society can and should value creativity may become increasingly pertinent.

5 Conclusions

Fundamentally, AI-generated misinformation is harmful not because the misinformation is AI generated, but because **such information is untrue**. Misinformation, whether created by genAI or humans, **erodes trust** in platforms and institutions that people rely on as reliable sources of information to make **informed decisions** and **causes harm** at the individual, organisational and societal levels. However, what AI-generated misinformation does challenge is the **effectiveness** of current mechanisms to detect and **mitigate** misinformation because of increased volume and quality of content, without a corresponding increase in truthfulness.

As such, AI-generated misinformation has to be **analysed and mitigated at both the creation and distribution stage**, as well as **within the context** of existing technology, regulation, policies, and research on human-generated misinformation.

The prevailing trend to combat AI-generated misinformation are solutions that **add context** by tracking and disclosing the provenance of AI-generated content. These labels inform consumers and allow them to decide the significance of AI-generated information. However, such solutions rely on technology that is **nascent** and are **constantly rendered ineffective** by malicious actors that circumvent them. These labels are **open to interpretation** by consumers who may assume that AI-generated content is more unreliable than human generated content. Rather than focusing on provenance, providing context to indicate **trustworthy sources of information** may be another solution. Nevertheless, trustworthiness can be highly subjective.

Other solutions include methods that **block or correct** misinformation, **limit** the speed of transmission, and **educate consumers** and law enforcement on risks of genAI in generating misinformation and the responsible use of such technology. Current regulation is **harm-focused and tech-agnostic** and thus are not necessarily redundant just because of genAI. While the nature of misinformation and the possible **public risks** that could emerge from genAI misinformation may justify **extending** current regulatory frameworks, partnership between industry and government should not be neglected. Industry has **voluntarily initiated** technical and policy solutions while the government can provide regulatory and compliance **baselines, standards and guidance** in sync with other jurisdictions.

Acknowledgements

This roundtable was founded via a charitable grant from Google.org, as part of Google's Digital Futures Project, and hosted by Meta Singapore.