MISINFORMATION AND DISINFORMATION AS MANAGEMENT RISKS —THE NEED FOR MULTI-STAKEHOLDER COLLABORATION—

Kazuma Kikuchi
Industrial Research Dept., Industrial & Social Studies Div.
Mitsui & Co. Global Strategic Studies Institute

SUMMARY

- The proliferation of sophisticated misinformation and disinformation driven by advances in generative AI constitutes a major management risk that threatens corporate stock prices and credibility. The World Economic Forum has also warned that this is among the most significant risks, making management-level responses indispensable.
- False or inaccurate information can be classified into three categories based on the degree of truthfulness and malicious intent: misinformation, disinformation, and malinformation. However, the boundaries between these categories are becoming increasingly blurred.
- There are limits to an individual's ability to improve their literacy and to the effectiveness of an individual company's countermeasures. Internationally, the prevailing trend is toward multi-stakeholder initiatives that transcend public-private frameworks. Experts anticipate that enterprises, such as general trading companies with strong overseas networks, will take the lead in implementing this framework across society.

1. ADDRESSING THE INCREASING URGENCY OF MISINFORMATION AND DISINFORMATION COUNTERMEASURES

The need to address misinformation and disinformation (hereinafter referred to as mis/disinformation) is becoming increasingly urgent in corporate management. With the remarkable evolution of generative AI, it is now possible to easily generate and disseminate videos and texts that are indistinguishable from authentic content. Some are designed with the specific intent to influence viewers, resulting in economic losses that companies cannot ignore.

In its Global Risks Report, the World Economic Forum (WEF) ranked mis/disinformation as the greatest risk over the next two years in both 2024 and 2025. In the latest 2026 edition, it ranks second after geoeconomic confrontation, a clear indication that it remains a serious concern for political and business leaders worldwide.[1] In addition, Gartner, Inc. estimates that corporate spending on mis/disinformation countermeasures will exceed USD 30 billion by 2028. The firm warns that "false information poses significant financial and reputational risks to organizations."[2] Mis/disinformation is no longer a problem solely for corporate communications or risk

[1] World Economic Forum (January 14, 2026) Global risks report 2026. By age group, respondents under 40 regard mis/disinformation as the most significant risk (p.18, Figure 15).
[2] Gartner Japan (October 30, 2025), "Gartner Predicts Enterprise Spending on Battling Misinformation and Disinformation Will Surpass $30 Billion by 2028"

management divisions. It has become a major management risk capable of undermining stock prices and credibility.

This report aims to provide an overview of the harm caused by mis/disinformation and to outline current trends in countermeasures.

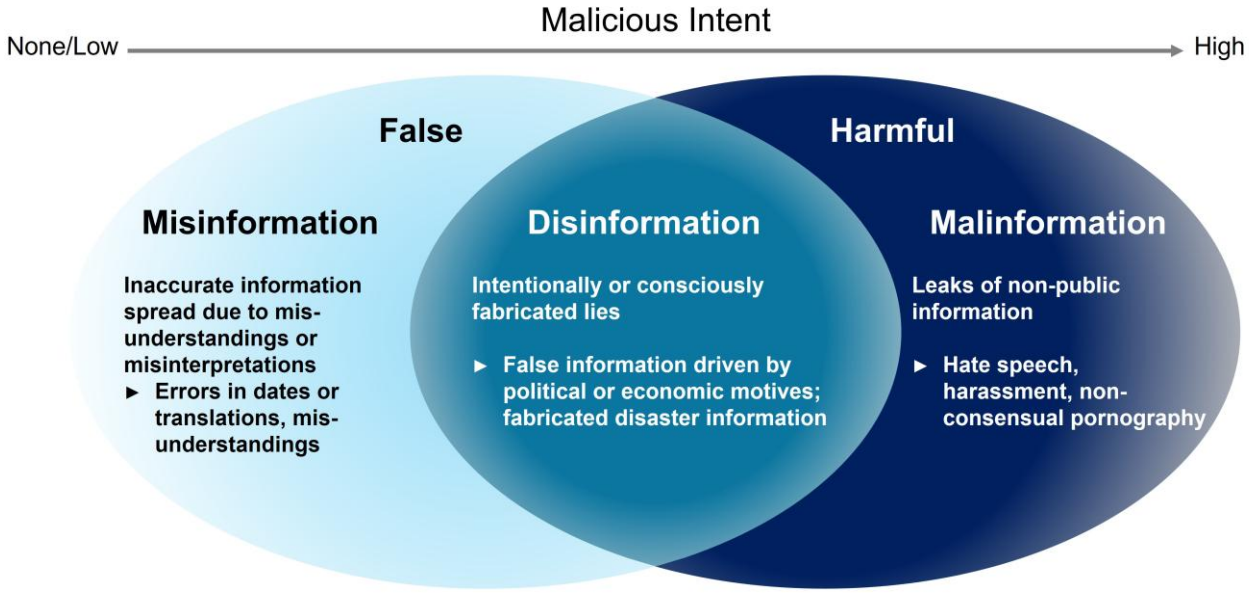
2. DEFINITIONS OF MIS/DISINFORMATION AND CASES OF HARM

2-1. WHAT IS MIS/DISINFORMATION?

While collectively referred to as mis/disinformation, judgments about the truthfulness of information can vary depending on its context, background, and the audience. Accordingly, even among experts, there is no consensus on how to define or use the terms, and no unified terminology system exists.[3] In this context, this report draws on the definition proposed by Wardle et al., which categorizes mis/disinformation based on the truthfulness of the information and the degree of malicious intent.[4] This categorization is considered useful for companies when they encounter mis/disinformation and must determine whether it is sufficient to simply correct the factual inaccuracies or whether an organization-wide response—including legal action—is required in response to what may constitute an organized attack. Wardle et al. classify mis/disinformation into the following three categories.

- **Misinformation:** Information that is false but not created with the intent to cause harm.
- **Disinformation:** Information that is false and deliberately created with the intent to harm individuals, social groups, organizations, or nations.
- **Malinformation:** Information that may be based on facts but is used with the intent to cause harm to individuals, organizations, or nations. It is often the result of leaks of confidential information.

Figure 1: Types and examples of mis/disinformation



Note: Presented as a gradient to reflect the increasingly blurry boundaries between misinformation, disinformation, and malinformation.
Source: Compiled by MGSSI based on Wardle & Derakhshan (2017); Shinichi Yamaguchi (2022), *The Complete Book of Social Media: Fake News, Flaming, and Information Bias*; and Ministry of Internal Affairs and Communications (2025), *How to Deal with the Internet: What You Should Do to Avoid Being Deceived by Dis/Misinformation, 2nd Edition*

[3] Jon Roozenbeek & Sander van der Linden (2025), *The Psychology of Misinformation*, translated by Yasuhiko Kano, Nippon Hyoronsha, p.2.
[4] Wardle, C. & Derakhshan, H (2017) Information disorder: Toward an interdisciplinary framework for research and policy making, pp.20-21

Figure 1 organizes misinformation, disinformation, and malinformation in order of the degree of malicious intent, based on the categorizations set forth by Wardle et al. Each category is represented as a gradient to reflect the current reality in which advances in AI have made it increasingly difficult to identify the source or intent behind any given piece of information, thereby blurring the boundaries between the categories.

Recently, generative AI has amplified the threat of mis/disinformation both qualitatively and quantitatively. According to NewsGuard, a US-based organization that evaluates online information, as of October 2025, there were 2,089 AI-generated news sites operating without human editorial oversight, posting articles in 16 languages, including English, Chinese, and Arabic.[5] Deepfake technology is also becoming increasingly sophisticated, and the Japanese government warns that it is now more important than ever to approach information with caution.[6] As a specific example, reports indicate that Nano Banana Pro—Google's image-generating AI released in November 2025—did not reject prompts intended to spread falsehoods relating to political messages or well-known brands, instead generating highly accurate fake images exactly as instructed.[7]

For these reasons, the proliferation of generative AI is thought to be amplifying the risks of mis/disinformation both quantitatively—through the mass generation of content—and qualitatively—through the increasing sophistication of expression.

2-2. INCREASINGLY DIVERSIFYING CASES OF HARM

Against this backdrop, the methods and scope of harm caused by mis/disinformation are becoming increasingly diverse (Figure 2).

One example is the November 2022 case involving the US pharmaceutical giant Eli Lilly, in which impersonation on social media directly affected its stock price. On Twitter (now X), an imposter account stated, "We are excited to announce insulin is free now," and the false information went viral amid public dissatisfaction over rising

Figure 2: Major cases of harm caused by mis/disinformation

Year	Company / Industry	Description	Impact	Affected Business Assets
2016	**New Balance** (US, sporting goods company)	• Comments by company executives regarding trade policy were **taken out of context and spread as "support for President Trump."** • **A far-right website unilaterally labeled the brand "the official shoe of white people."**	A boycott and loss of support, particularly among younger consumers	Brand value, customer base, and sales
2022	**Eli Lilly** (US, pharmaceutical company)	• **An imposter account posted that "insulin is now free,"** spreading false information on social media.	A 4.37% drop in stock price	Shareholder value and aggregate market value
2023	**Target** (US, retailer)	• **Fake images claiming that the company was selling satanic-themed clothing spread on social media.** • Employees were subjected to threats, etc.	A decline in sales and an increase in store operating costs (security and response)	Sales and human capital
2024	**Arup** (UK, engineering company)	• **Deepfake fraud via video calls.** Employees were deceived by **an AI-generated video impersonating a company executive with remarkable precision**, and instructed to transfer large sums of money.	USD 25 million in fraudulent transfers	Financial assets and governance (internal controls)
2025	**Japan's inbound tourism industry**	• **A rumor claiming that a massive earthquake will strike Japan in July 2025 spread in Hong Kong and other parts of Asia.** • The **source was** found to be **a 1999 Japanese manga**.	Loss of potential consumption opportunities	Sales

Source: Compiled by MGSSI based on various media reports

[5] NewsGuard (October 24, 2025) Tracking AI-enabled misinformation: Over 2000 undisclosed AI-generated news websites (and counting), plus the top false narratives generated by artificial intelligence tools.

[6] Cabinet Secretariat portal site on disinformation and related activities by foreign actors, page on avoiding being deceived by disinformation.

[7] NewsGuard Google's new AI image generator is a misinformation superspreader (accessed December 3, 2025)

insulin prices. The company's stock price later plunged by 4.37%.[8] In addition, in 2024, the Hong Kong office of the British engineering firm Arup was defrauded of USD 25 million in a scam that made use of deepfake technology.[9] The perpetrator reportedly impersonated the company's CFO and other staff members with remarkable precision during a video call, and instructed employees at the office to transfer the funds.

In addition, Japan's inbound tourism industry is an example of an entire sector suffering reputational damage. A scientifically unfounded rumor claiming that a massive earthquake will strike Japan in July 2025 spread on social media platforms throughout Asia. Estimates suggest that the loss of consumption opportunities due to canceled trips to Japan could reach as high as JPY 560 billion.[10]

At the same time, there are also cases in which swift responses to misinformation helped avoid losses. In November 2024, when a social media post accompanied by a video claiming to have found an insect in a Tirol-Choco product began to spread, the official Tirol-Choco account promptly investigated the situation and provided an explanation. By demonstrating that the insect had entered the product while it was in the poster's possession, rather than during manufacturing, the company succeeded in dispelling the misinformation.[11]

However, in the current era, in which generative AI can mass-produce highly convincing disinformation, companies that rely solely on round-the-clock monitoring by human operators, followed by swift actions to mitigate incidents as they arise, will likely face limits in both cost and speed.

3. TRENDS IN COUNTERMEASURES

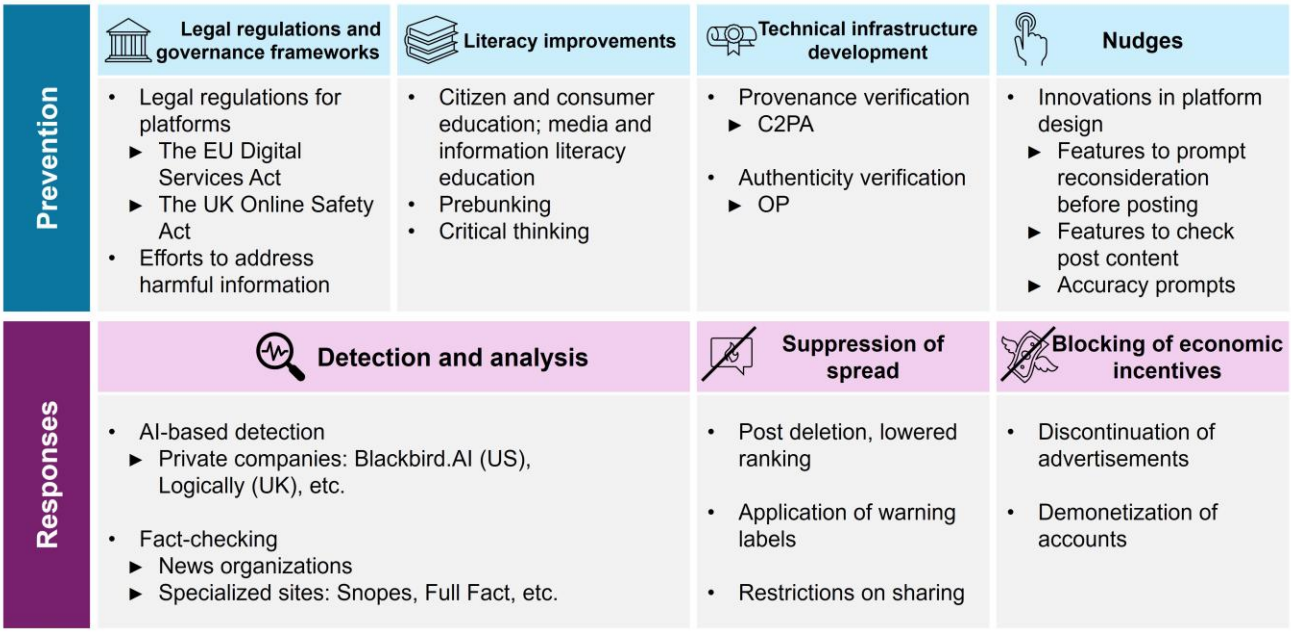
3-1. PREVENTION OF AND RESPONSES TO MIS/DISINFORMATION

Efforts to address mis/disinformation can be broadly divided into prevention measures and responses to contain incidents once they occur (Figure 3).

First, prevention measures can be grouped into four main pillars. First is the development of legal regulations and governance frameworks. Examples include the EU's Digital Services Act (DSA) and social media regulations enacted by various countries. Second are measures to bolster the audience's resilience, such as literacy education and prebunking. Measures to prebunk involve showing people examples of mis/disinformation and providing them with advance warning prior to exposure, enabling them to acquire psychological "immunity." Third is the development of technological infrastructure able to apply technology to verify the provenance of content and the authenticity of its source. Examples include the Coalition for Content Provenance and Authenticity (C2PA) and the Originator Profile (OP), discussed later. The fourth is the use of nudges, a concept from behavioral economics. A nudge is a means of encouraging people to change their behavior (in a positive direction) by providing a small prompt. Examples include platform design features that prompt users to confirm or reconsider their actions when posting on social media, such as asking, "Are you sure you want to post this?"

[8] Forbes (November 12, 2022) Fake Eli Lilly Twitter account claims insulin is free, stock falls 4.37%. As also noted in the cited article, it should be kept in mind that stock prices are influenced by a variety of factors.
[9] CNN (May 17, 2024) British engineering giant Arup revealed as $25 million deepfake scam victim.
[10] Nomura Research Institute (May 29, 2025), "Speculation About a Large-Scale Natural Disaster in July Without Scientific Basis Casts a Shadow Over Robust Demand from Inbound Visitors to Japan: Estimated Economic Losses of JPY 560 Billion," *Takahide Kiuchi's View – Insights Into World Economic Trends*.
[11] Yuzuru Kido (November 8, 2024), "Why Tirol Chocolate's Response to the 'Insect Contamination' Controversy Was So Impressive—Swift Action and the Art of Consumer Communication," *Toyo Keizai Online*.

Figure 3: Major countermeasures against mis/disinformation



Source: Compiled by MGSSI based on Jon Roozenbeek & Sander van der Linden (2025); Yamaguchi (2022); and Kozyreva, A. et al.(2024) "*Toolbox of individual-level interventions against online misinformation*"

Next, responses when working to suppress the spread of already generated mis/disinformation can be broadly divided into three approaches. The first approach—detection and analysis—makes use of AI to identify early signs of dissemination and attempts to pinpoint sources such as bots or fraudulent organizations. Representative examples include AI-powered services provided by private companies and fact-checking activities by news organizations. The second approach is to curb the spread. Measures include deleting posts containing false information, lowering their display rankings, and attaching warning labels. When applying this approach, it is necessary to prevent excessive dissemination through transparent procedures while respecting freedom of expression. The third approach—blocking economic incentives—involves cutting off advertising distribution and payment processing. The aim is to eliminate economic motivations by depriving disinformation site operators of their access to funding.

3-2. THE IMPORTANCE OF MEASURES INVOLVING MULTIPLE COMPANIES AND FUTURE PROSPECTS

Among the various countermeasures described above, efforts to develop technological infrastructure are becoming especially important. This is because the ability of individuals to improve their literacy and the effectiveness of post hoc responses by human operators are reaching their limits in the face of the mass production of mis/disinformation by generative AI. Against this backdrop, multiple stakeholders are collaborating to technically verify content provenance—indicating when, by whom, and where content was created—in an effort to restore trust in the digital space.[12]

A representative example is the C2PA, led by global tech and media companies such as Microsoft, Adobe, Intel, and the BBC.[13] This technical standard embeds provenance information, such as methods, location, and date and time, into digital content as a tamper-proof digital signature. Its advantage lies in being an open standard that does not depend on any specific platform.

[12] The 2026 edition of the WEF report also states that multi-stakeholder engagement is essential in preventing misinformation and societal polarization (p.39).

[13] C2PA | Verifying Media Content Sources

In Japan as well, research and development of OP technology is being pursued primarily by newspaper companies and advertising firms. This framework grants a certificate indicating that content on a website originates from a trustworthy source, making the reliability of online articles and advertisements more readily apparent.[14]

Of course, the application of technology alone cannot address all mis/disinformation issues. With regard to future developments in countermeasures, Professor Shinichi Yamaguchi, a specialist in econometrics at the Center for Global Communications (GLOCOM) of the International University of Japan, points out that creating a healthy information space requires not only promoting the adoption of a single technology (such as C2PA or OP), but also the establishment of a framework in which numerous actors examine issues from multiple perspectives and pursue social implementation. He also expressed hope that, while technology-focused players engage in development, general trading companies and other stakeholders with robust overseas networks will be able to play a role in standardization to facilitate social implementation. The key going forward will be whether this division of roles and collaborative framework can be concretely established through international rule-making and pilot projects. General trading companies, with their advanced ability to coordinate between governments, companies, and media outlets in multiple countries, may be able to take on a leading role in this process.

It is hard to know what to believe anymore… It is precisely because we live in such times that we face an urgent need to proliferate technologies able to serve as the foundation of trust throughout society, through engagement by multiple stakeholders.

[14] Originator Profile Technology overview