DISRUPTING THE DISORDER: A Multi-Stakeholder Discussion on Information Integrity in the Philippines

Discussion Paper on the Spread of Dis-, Mis-, and Malinformation in Social Media Platforms

By **Alma Karina "Kankan" Ramos**
With research support from Clara Quiambao

October, 2025

TABLE OF CONTENTS

EXECUTIVE
SUMMARY

EXECUTIVE SUMMARY

This paper, developed by the **FEU Public Policy Center Foundation Inc.** (FEU PPC) with the support of **TikTok PTE LTD**, synthesizes insights from roundtables and focus group discussions convened by the FEU PPC on the urgent issue of disinformation, misinformation, and malinformation in the Philippines. It is presented not as a prescriptive set of solutions but as a discussion document - meant to surface diverse perspectives, highlight challenges, and spark dialogue across sectors.

The circulation of harmful content is driven by overlapping forces: platform architectures and business models that reward virality, user behaviors shaped by cognitive biases and cultural practices, and political and economic incentives that exploit these vulnerabilities. In the Philippines, these challenges are amplified by one of the highest social media penetration rates in the world, persistent gaps in media and digital literacy, and a polarized political landscape.

KEY INSIGHTS INCLUDE:

No single cause, no single solution: Information disorders are the result of systemic dynamics involving content, creators, receivers, and platforms - all operating within a fertile digital ecosystem.

Philippine-specific vulnerabilities: Local languages, humor, parody, and political divides make harmful narratives harder to detect and easier to spread. Fact-checks often struggle to penetrate close-knit online networks such as family group chats.

Shared responsibility across sectors: Government, digital platforms, civil society, academe, media, and the communications and marketing industry each hold unique roles. While some marketing and advertising agencies and PR firms have been hired to manipulate conversations, these same industries also hold the expertise to design ethical, civic-minded campaigns that counter disinformation.
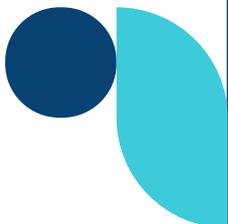
Momentum in policy experimentation: Proposed legislation such as House Bill 3860 (with support from the Digital Media Standards Coalition) illustrates efforts to find a middle ground between overbearing state regulation and opaque, profit-driven platform policies.

Beyond consumption toward creation: Strengthening Media and Information Literacy (MIL) requires not only teaching citizens to critically consume information but also to responsibly create and amplify it - recognizing that likes, shares, and posts all contribute to the information ecosystem.

> Ultimately, the path forward is not about quick fixes but about building inclusive, trust-based, and culturally grounded collaborations. By fostering cross-sector commitment, aligning incentives, and empowering Filipinos with the skills to both consume and create responsibly, stakeholders can begin to disrupt the cycle of harmful content and safeguard the integrity of the country's democratic discourse.

INTRODUCTION

INTRODUCTION

The rise of false and misleading content across social media platforms has increasingly emerged as one of the defining challenges of our digital age. In the Philippines, 90.8 million individuals (that is equivalent to 78% of the population) are active social media users (We Are Social & Meltwater 2025). This level of penetration plays a major role in shaping public opinion, trust in information, and how individuals participate in national conversations. Indeed, the Digital News Report 2025 found that Filipino concern about online disinformation is at its highest recorded level (Chua, 2025).

Social media platforms now shape public discourse more than ever, acting as a primary source of news, political messaging, and cultural narratives. However, these platforms are designed to prioritize engagement, meaning that content which sparks strong emotional reactions tends to spread more quickly and widely, regardless of accuracy, than other forms of information (Wardle & Derakhshan, 2017). This architecture amplifies misleading or harmful content, contributing to the rapid spread of disinformation, misinformation, and malinformation.

116M

PHILIPPINE POPULATION

Active Social Media Users

90.9 M

78% SOCIAL MEDIA PENETRATION

3H 43M

AVERAGE DAILY TIME SPENT ON SOCIAL MEDIA

To better understand these dynamics, this paper adopts the framework developed by Wardle and Derakhshan (2017), which distinguishes between:

Disinformation: false information shared knowingly and with intent to cause harm.

Misinformation: false information shared without intent to cause harm.

Malinformation: factual information shared to cause harm, often by stripping context or violating privacy.
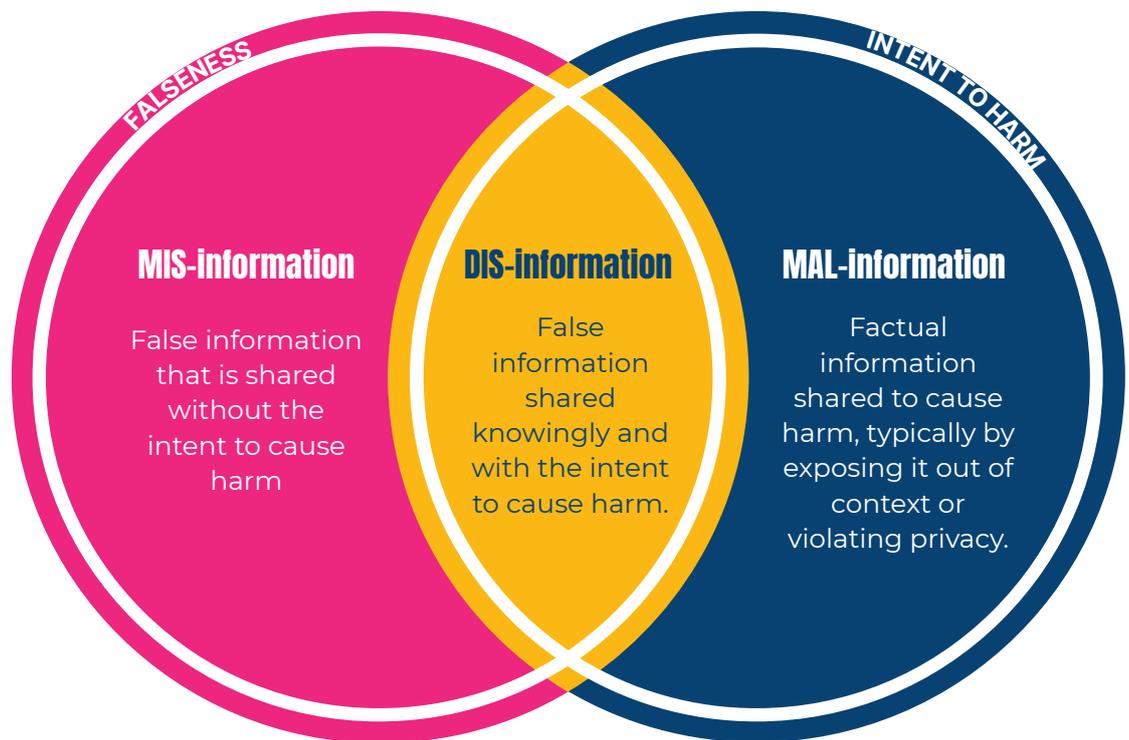
INFORMATION DISORDER
Adapted from Wardle & Derakhshan's (2017)
"Information Disorder" framework

INTRODUCTION

In recent years, these categories have played out visibly in the Philippines. During the 2016 Duterte administration, for instance, organized disinformation networks employed hierarchies of digital workers to strategically seed and amplify narratives (Ong & Cabañes, 2018; PCIJ & Cabato 2020). In the 2022 national elections, researchers observed the creation of "parallel public spheres" where influence operations reshaped political conversations online. On one side, networks amplified content that glorified the Marcos family legacy; on the other, narratives discrediting Vice President Leni Robredo spread widely through memes, parody, and misleading videos. These operations normalized historical revisionism and fragmented public debate, showing how disinformation is used not just to persuade but to restructure how Filipinos engage in democratic discourse (Media Manipulation Casebook, 2022).

The Philippine case is especially significant given its extremely high social media use and history of politicized manipulation. As highlighted by RIWI's Information Dystopia and Philippine Democracy (2022), disinformation operates not only as political strategy but also through cultural and emotional dynamics that shape trust, awareness, and civic engagement. This situates the Philippines not as an outlier but as a critical case within the broader regional and global disinformation landscape.

SCOPE

This paper aims to open dialogue on the spread of disinformation, misinformation, and malinformation among policymakers, academe, media, civil society, digital practitioners, and social media platforms. While grounded in the Philippine context, it also draws on comparative international models and platform practices.

It is important to state at the outset that this paper does not advance prescriptive recommendations. Instead, it synthesizes existing research, roundtable discussions, and focus group insights to surface key issues, binding constraints, and areas for collaboration. By doing so, it seeks to support more **responsive governance and a healthier digital information ecosystem** - one that balances information and innovation with accountability and reflects the lived realities of Filipino users.

THE RISE OF SOCIAL MEDIA AND INFORMATION ECOSYSTEMS

BACKGROUND AND CONTEXT: THE RISE OF SOCIAL MEDIA AND INFORMATION ECOSYSTEMS

The rise of social media has transformed how Filipinos access, consume, produce, and share information. From the early days of Friendster, MySpace and Multiply to the dominance of Facebook, Instagram, YouTube, and TikTok, the Philippines has consistently ranked among the most active social media user bases worldwide. By the late 2000s, the country had earned the title of the "social media capital of the world", with Filipinos among the heaviest daily users of networking platforms (FMA, 2008).

This digital connectivity has reshaped not only communication but also the public sphere. Social media has become more than a tool for personal interaction; it now functions as a primary arena for political engagement, cultural expression, and public discourse. Yet this transformation has also made the country vulnerable to the rapid spread of information disorders. Platforms, designed for engagement and virality, amplify sensational and polarizing content, often to the detriment of accuracy and trust.

The broader information environment is shaped by several defining characteristics:

Low-cost, sophisticated editing and publishing tools enable almost anyone to produce high-quality content that can mimic trusted sources.

Public and interactive consumption allows instant reactions and social validation, amplifying the spread of unverified claims.

Real-time peer-to-peer sharing lends credibility to falsehoods, as people are more likely to trust information from family, friends, or influencers.

Accelerated news cycles, intensified by mobile-first access, reduce time for reflection and verification, allowing misinformation to circulate unchecked (Filloux, 2017).

BACKGROUND AND CONTEXT: THE RISE OF SOCIAL MEDIA AND INFORMATION ECOSYSTEMS

In the Philippine context, disinformation has been systematically deployed to shape political outcomes. During the Duterte administration, investigative reports documented how hierarchies of digital workers and coordinated networks seeded deceptive narratives, normalizing disinformation as part of governance and electoral strategy (Ong & Cabañes, 2018).

These operations worked through layered disinformation infrastructures:

- **Troll farms and digital workers** produced coordinated messaging that attacked critics, spread memes, and promoted Duterte's populist image. Many "micro-influencers" were contracted to appear organic while echoing official talking points.
- **False news websites** circulated fabricated stories, often shared widely on Facebook. For example, posts claimed that Senator Leila de Lima was romantically linked to a convicted drug lord, part of an orchestrated smear campaign to discredit her investigations into Duterte's war on drugs (Rappler, 2016; Ong & Cabañes, 2018).
- **Memes and parody content** were deployed to trivialize opponents and normalize state narratives. These often masked harmful claims under humor, making them more shareable and less likely to be flagged as outright falsehoods.
- **Manipulated videos and clickbait headlines** amplified pro-government propaganda while reframing controversial events. For example, extrajudicial killings were reframed as "necessary" actions against crime, presented in sensational, populist tones.
- **Hashtag campaigns such** as #DeLimaLies or #ProtectDuterte trended to consolidate online support and drown out dissenting voices.

This system blurred the line between grassroots support and coordinated propaganda, embedding disinformation into everyday online discourse. As Ong and Cabañes argue, the strategy was not only to spread falsehoods but also to normalize them as part of the political conversation, making propaganda appear as citizen-driven expression rather than orchestrated manipulation.
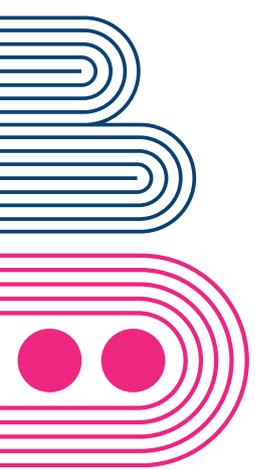
BACKGROUND AND CONTEXT: THE RISE OF SOCIAL MEDIA AND INFORMATION ECOSYSTEMS

This infrastructure persisted into the 2022 national elections, when researchers observed the creation of **"parallel public spheres"** that fragmented debate and mainstreamed historical revisionism (Media Manipulation Casebook, 2022). These parallel arenas were visible in how different candidates' supporters inhabited separate digital spaces, one amplifying positive content about Ferdinand "Bongbong" Marcos Jr. and the Marcos family legacy, while another disseminated memes and posts discrediting Vice President Leni Robredo through allegations of incompetence or electoral fraud. Specific examples included:

- **Historical revisionism** campaigns reframing the Marcos dictatorship as a "golden age," circulated through Facebook videos, TikTok clips, and YouTube vlogs that reached millions of views.

- **Memes and parody content** mocking Robredo's speeches and mannerisms, designed to go viral among youth audiences on TikTok and Twitter.

- **Coordinated hashtag pushes**, such as #KakampinkLies and #MarcosPaRin, which trended repeatedly during campaign periods, creating the illusion of widespread consensus.

- **Influencer-driven disinformation**: Research by Internews (2022) on the Political Economy of Covert Influence Operations revealed how covert payments to mid-level influencers, meme creators, and digital strategists helped sustain these operations. Influencers received fees ranging from small stipends to large-scale contracts in exchange for posting campaign-aligned content, blurring the line between organic fandom and paid propaganda.

- **Astroturfing tactics**, where fake grassroots pages and groups were set up to amplify pro-Marcos narratives and drown out critical discourse, creating what scholars call "engineered consensus."

BACKGROUND AND CONTEXT: THE RISE OF SOCIAL MEDIA AND INFORMATION ECOSYSTEMS

These practices demonstrated that disinformation in the 2022 elections was **not just cultural or ideological, it was also financialized**, with political campaigns treating online manipulation as a professionalized service. This political economy of influence operations highlights the growing need to address disinformation not only as a question of truth and ethics but also as an industry sustained by money, contracts, and organized labor.

These trends align with findings from RIWI's Information Dystopia and Philippine Democracy (2022), which highlight the cultural and emotional dimensions of disinformation and its role in declining trust in institutions and contested democratic practices. Taken together, these studies position the Philippines not as an outlier but as a regional case study of networked propaganda, with lessons that resonate globally.

Beyond the 2016 and 2022 electoral campaigns, disinformation infrastructures remain entrenched in shaping political discourse. **PCIJ's Five Ways the Duterte Influence Machine is Deceiving Filipinos** (Cabato & PCIJ 2025) documented how networks of influencers, content farms, and loyal online communities systematically manipulate narratives in favor of political figures while discrediting critics. The investigation highlighted tactics such as coordinated hashtag campaigns, seeded fake news stories, and the weaponization of parody and satire to normalize falsehoods. This underscores that influence operations in the Philippines are not episodic but continuous, evolving, and deeply embedded in the digital ecosystem.

BACKGROUND AND CONTEXT: THE RISE OF SOCIAL MEDIA AND INFORMATION ECOSYSTEMS

Hero–Victim Narratives: Resurfacing Mis- and Disinformation Around Duterte's Legacy
Rappler.com
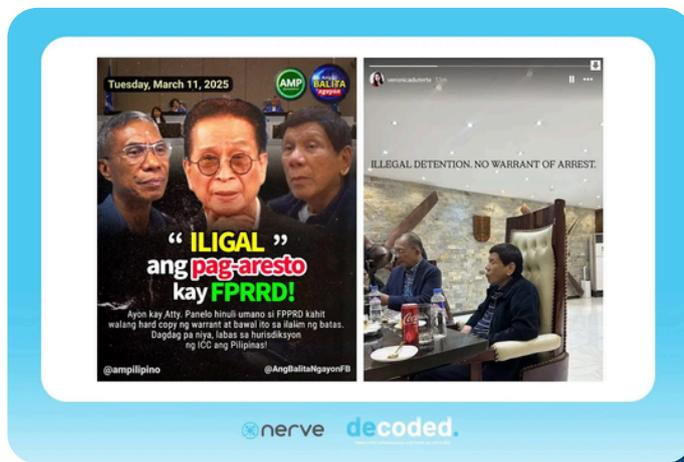
Hero–Victim Narratives: Resurfacing Mis- and Disinformation Around Duterte's Legacy
Rappler.com
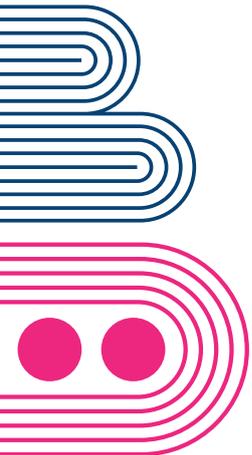
Dismissal and Distortion: False claim - The Supreme Court ruling did not declare Duterte innocent of the charges
Rappler.com

3

PHILIPPINE INTERNET LANDSCAPE

116M PHILIPPINE POPULATION

97M 83.8% INTERNET USERS

90.9 M 78% SOCIAL MEDIA USERS

142M 122% MOBILE CONNECTIONS

MOST USED SOCIAL MEDIA PLATFORMS

FACEBOOK
MESSENGER
TIKTOK
INSTAGRAM
TELEGRAM
X

0 20 40 60 80 100
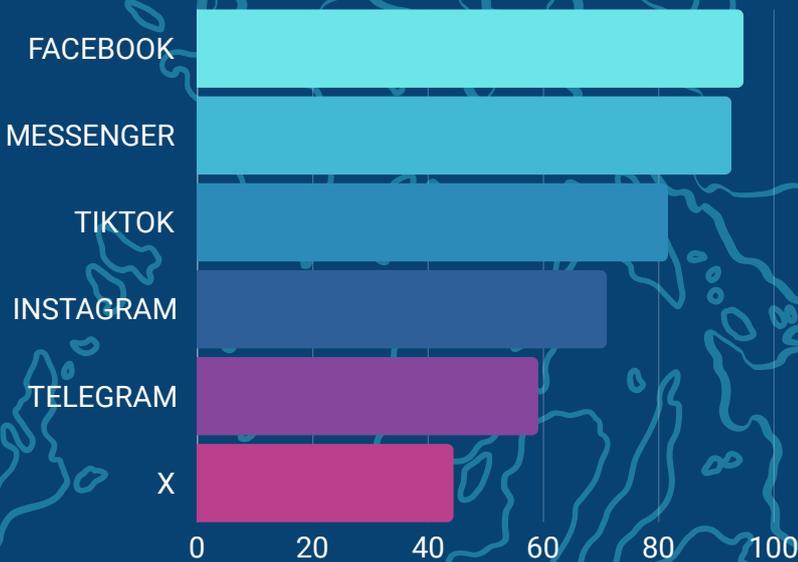
Social media platform usage in the Philippines. Source: We Are Social & MeltWater (2025)

AVERAGE TIME PER USER

TIKTOK 40H 39M

YOUTUBE 25H 26M

FACEBOOK 23H 34M

MESSENGER 16H 27M

INSTAGRAM 4H 24M

Average time per month active users spent using each platform

UNDERSTANDING DIS-, MIS-, AND MAL-INFORMATION

In the rapidly evolving digital information ecosystem, disinformation, misinformation, and malinformation, immensely impact public trust, social cohesion, and democratic processes. These phenomena are not spontaneous, they are enabled by overlapping motivations, deliberate formats, platform architecture, behavioral dynamics, and limitations in platform governance.

They operate through a dynamic system involving three core elements: **creators, messages/content, and receivers.** These elements describe how false or harmful information originates, takes form, and influences public perception. However, it is important to emphasize that **platforms act as the medium** that transfer, accelerate, and shape the interaction among all three core elements.

A THE CREATOR

Creators are individuals or entities that generate and disseminate false, misleading, or harmful information. They can be politically motivated actors, profit-driven content farms, ideologically motivated influencers, or ordinary users.

Motivations include:

Political power: To manipulate narratives, discredit opposition, or influence elections (Bradshaw & Howard, 2019). Political motivations drive the use of disinformation as a tool for power, shaping public opinion, discrediting opponents, and manipulating electoral outcomes. Coordinated online campaigns have been used in the Philippines and elsewhere to attack critics, promote political figures, and influence national narratives. In the Philippines, coordinated campaigns targeted critics like Sen. Leila de Lima (2016) and fueled parallel public spheres during the 2022 elections (Media Manipulation Casebook, 2022).

Economic gain: Financial incentives are equally powerful. The *attention economy* rewards virality—regardless of accuracy—through ad revenue, affiliate links, or paid amplification (Marwick & Lewis, 2017). Clickbait articles, fake news sites, and content farms profit from sensational content that spreads widely, especially in regions with high engagement like the Philippines. According to Internews's study on the Political Economy of Covert Influence Operations, political actors are estimated to have spent as much as ₱1.5 billion (approx. USD 27 million) outside declared campaign expenses to commission influencers for covert digital campaigning in the 2022 Philippine elections. Some campaigns also used a pay-per-post model, estimating up to ₱600 million for such postings.

The attention economy is a system where online platforms compete for user focus. Content that provokes strong emotions or clicks gets boosted by algorithms, making virality the main reward—accuracy optional.

Social or psychological drivers: Motivated by identity, grievance, or recognition. Individuals share content for validation or group belonging, amplified by cognitive biases like confirmation bias and the illusory truth effect (Lewandowsky et al., 2017). Driven by belief systems, perceived grievances, or a desire for recognition and control. Social or psychological drivers include the desire for validation, identity expression, or community belonging. People may share misleading content to reinforce group identities, gain social approval, or simply because it resonates emotionally. Cognitive biases such as confirmation bias and the illusion of truth effect amplify this behavior, often making users more susceptible to falsehoods even when factual corrections are available.

CREATORS MAY BE EITHER:

Organized: such as coordinated disinformation networks - Troll farms, political operatives, state-sponsored (local or foreign) disinformation units.

Non-organized: Influencers or ordinary users spreading falsehoods without structured intent.

These creators may work openly or covertly, often disguising their content to mimic credible sources or tapping into emotionally resonant themes to ensure virality.

B THE MESSAGES OR CONTENT

The content produced and circulated in disinformation campaigns is crafted to maximize emotional impact, shareability, and perceived authenticity.

Common formats include:

Memes and infographics that compress complex ideas into emotionally resonant visuals

Short-form videos popular on platforms like TikTok, Facebook Reels, and YouTube Shorts

Deepfakes, altered images, and manipulated video/audio clips

Clickbait headlines, fake news articles, and pseudo-news websites

Core characteristics of disinformation messages:

1 | Emotionally charged language and imagery (e.g., anger, outrage, fear, pride)

2 | Simple and easy to understand language, repetitive phrasing to facilitate recall

3 | Misleading context or selective omission of facts

4 | Appeals to authority, nationalism, or identity

UNDERSTANDING DIS-, MIS-, AND MAL-INFORMATION

These messages are made more potent by platform design; social media algorithms are optimized to promote content that generates high engagement, often at the expense of accuracy. The virality-driven logic of these systems prioritizes sensational, polarizing, or emotionally provocative content, which accelerates the spread of disinformation (Vosoughi, S., Roy, D., & Aral, S., 2018). In the Philippines, memes and TikTok-style videos were central in spreading **historical revisionism and electoral propaganda** (Ong & Tapsell, 2020).

The Duterte influence machine (PCIJ & Regine Cabato (2025) illustrates how creators range from professionalized troll farms and political operators to everyday supporters who replicate and amplify narratives. Their content spans memes, fake news articles, YouTube vlogs, livestreams, and parody posts, all crafted to stir outrage, reinforce loyalty, or discredit opponents. This combination of organized operations and organic user participation blurs the line between **top-down manipulation and bottom-up mobilization**, making it harder to trace accountability or dismantle networks.
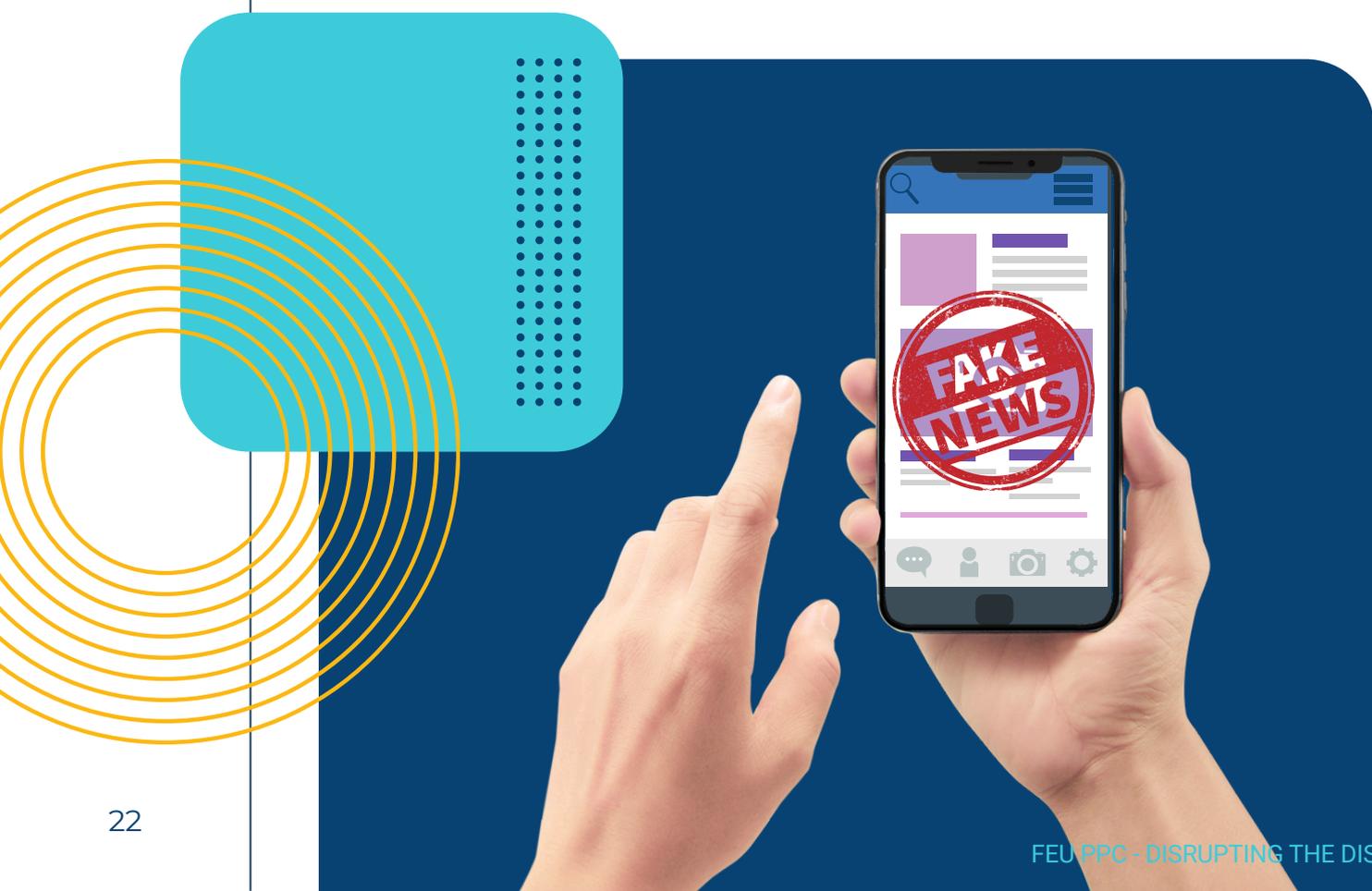
C THE RECEIVERS

Receivers are the users who encounter and engage with disinformation - either passively (e.g., viewing or reacting) or actively (e.g., sharing, commenting, defending). Their behavior plays a decisive role in whether false content dies out or spreads widely.

Receiver profiles typically include:
- Diverse socioeconomic and educational backgrounds, but with common vulnerabilities such as low media and digital literacy
- Pre-existing political or social biases that influence how information is processed and interpreted
- Heavy reliance on social networks, peer groups, and online communities for news and opinions

Psychological and behavioral factors:

Confirmation bias: The tendency to believe and share information that confirms existing beliefs (Lewandowsky, Ecker, Cook, 2017)

Illusory truth effect: Repeated exposure to false claims increases the likelihood of believing them

Information overload: Users may resort to fast and emotional decision-making instead of critical analysis

Social validation: Trust in peer-shared content over institutional sources

Speed of Sharing: Emotional content is spread quickly, often without verification due to urgency, peer pressure, or the desire to be the first to share.

These behaviors are further amplified by the echo chambers and filter bubbles created by algorithmic content curation, which limit exposure to dissenting viewpoints and reinforce belief systems.

A THE CREATOR

MOTIVATIONS: Political power, Economic Gain, Social or Psychological reasons

TYPE: Organized or non-organized
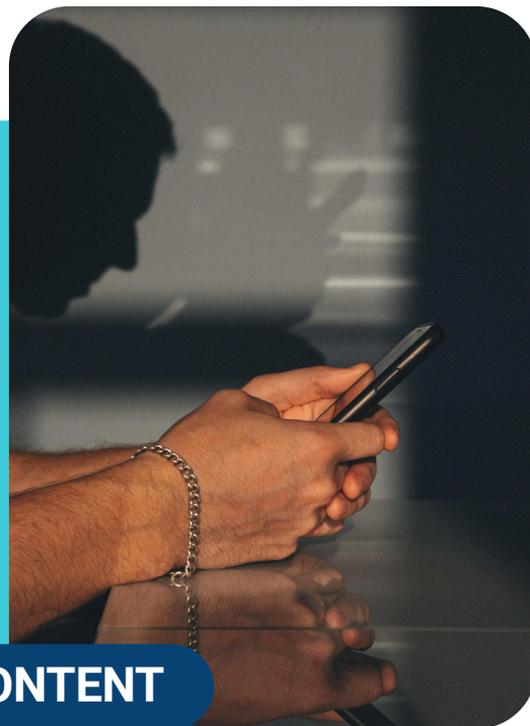
B THE MESSAGES OR CONTENT

Core characteristics of disinformation messages:

- Emotionally charged language and imagery
- Simple and easy to understand language, repetitive phrasing to facilitate recall
- Misleading context or selective omission of facts
- Appeals to authority, nationalism, or identity

C THE RECEIVERS

BEHAVIOR FACTORS: Confirmation bias, Illusory truth effect, Information overload, Social validation, Speed of sharing

The Role of Platforms as Mediums

Platforms are not passive carriers of content—they actively shape visibility, virality, and trust. In the Philippines, disinformation thrives across these key spaces:

Facebook / Meta – The country's most dominant platform, where Facebook Groups and Messenger chats act as echo chambers for political propaganda and family-circulated falsehoods.

YouTube – A hub for long-form propaganda and influencer-driven vlogs that legitimize misleading narratives through polished video formats.

TikTok – Popular with youth audiences; short-form, viral videos spread memes, historical revisionism, and emotionally charged political content.

X (Twitter) – A battleground for hashtags and fast-moving debates, often driven by bots and troll accounts to manipulate "trending" conversations.

Instagram – Lifestyle-driven but also repurposed by influencers to embed subtle political and ideological messaging.

Messaging Apps (Messenger, Viber, Telegram, WhatsApp) – Private networks where fact-checking rarely penetrates, but rumors and manipulated content circulate rapidly.

Why this matters

These platforms are the architecture of virality—through algorithms, trending lists, hashtags, and group features, they privilege sensational or polarizing posts over accurate information.

The Role of Platforms as Mediums

Platforms are the infrastructure where creators, content, and receivers interact at scale.

SHAPING VISIBILITY

The algorithms that most platforms use value engagement over accuracy (Vosoughi, Roy, & Aral 2018). They prioritize content that maximizes engagement, which can privilege sensational or polarizing posts over accurate information.

SETTING THE RULES

Platforms choose their own policies for moderation, content removal, and labeling, which determine what remains visible.

PROVIDING THE ARCHITECTURE OF VIRALITY

Features like trending hashtags, recommendation feeds, and group functionalities facilitate the rapid spread of content.

EMBEDDING INCENTIVES

Engagement-driven business models often reward creators who produce attention-grabbing (even if misleading) content.

In this sense, platforms are not neutral carriers but active environments where the mechanics of information disorder play out. They may amplify the influence of creators, shape the form and reach of messages, and condition the behavior of receivers.

This framing makes clear that while the three elements (creators, content, receivers) remain analytically useful, platforms are the critical ecosystem layer that connects and intensifies their interactions.

Platform Moderation Challenges

Platforms themselves face significant obstacles in managing disinformation at scale:

- **Moderation at scale:** Platforms like Meta report removing billions of fake accounts, but automated systems often fail to detect more nuanced or localized content (Meta, 2023).
- **Language and cultural nuance:** Content in local dialects or languages, or the use of humor, sarcasm, or coded language can evade detection by AI tools and moderators unfamiliar with local contexts.
- **Adaptability of bad actors:** Coordinated disinformation networks quickly evolve, creating new accounts, modifying tactics, and finding loopholes in moderation systems.

These challenges are particularly acute in the Philippines, where high social media penetration, low media literacy, and political polarization create fertile ground for information manipulation.

The interplay between disinformation creators, emotionally resonant and misleading content, and cognitively biased receivers is central to understanding why harmful content spreads so rapidly. Without systemic interventions, ranging from improved platform accountability to widespread digital literacy, the information ecosystem remains vulnerable to coordinated manipulation and viral falsehoods.

AI-Driven Disinformation
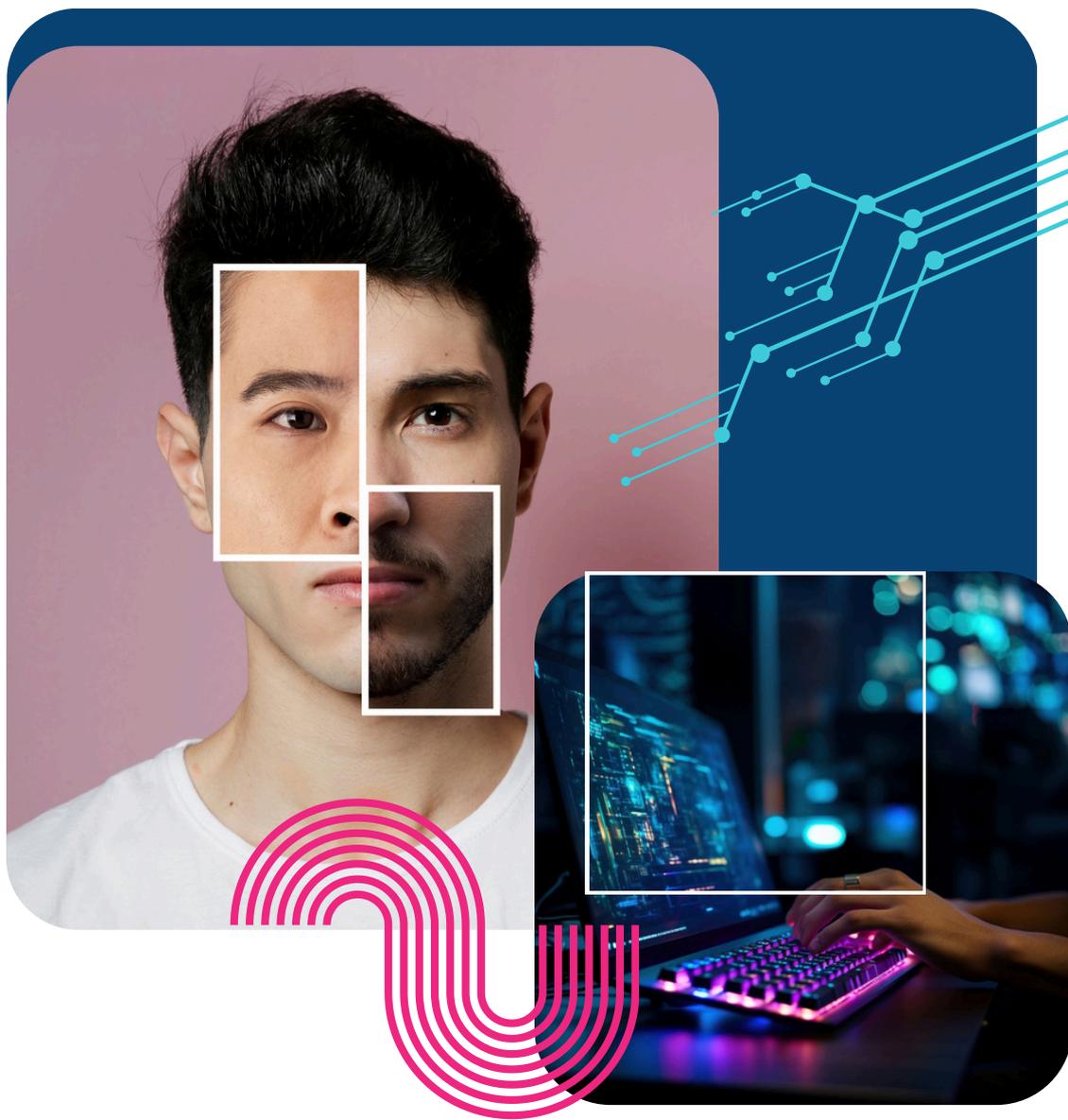
AI has already been weaponized in Philippine elections. During the 2025 midterms, #RespectMyPost documented face-swapped images, AI-generated profiles, and synthetic videos circulating online. These blurred the line between authentic and manipulated activity, complicating watchdog efforts and raising questions about detection, regulation, and transparency.

This underscores that AI is not just a future risk, it is a present challenge that must be addressed with urgency.

INFORMATION DISORDER CYCLE

1. **CREATION AND PRODUCTION**
2. **DISTRIBUTION**
3. **AMPLIFICATION**

This is the phase where harmful content originates. It may be crafted by political actors, troll farms, influencers, marketing agencies, or even ordinary users. Motivations vary: some seek political gain, others financial profit, ideological influence, or social disruption.

- Disinformation in this phase often includes fabricated data, deepfakes, altered images, or false testimonials designed to mimic credible sources or evoke emotional reactions.

- Malinformation, such as doxxing or weaponized leaks, relies on truthful information distorted to cause harm.

- Tools such as generative AI, cheap editing software, and bot networks have made producing such content more accessible and scalable than ever before (Wardle, & Derakhshan, 2017).

Example: The 2016 campaign against Senator Leila de Lima featured coordinated smear tactics involving fake news articles and social media manipulation to discredit her politically (Ong, & Cabañes, 2018).

INFORMATION DISORDER CYCLE

1 **CREATION AND PRODUCTION**

2 **DISTRIBUTION**

3 **AMPLIFICATION**

Once created, content is seeded into the ecosystem via multiple channels: social media, messaging apps, forums, and comments on websites. Distribution methods include:

- Use of influencers or "bridge figures" who spread content beyond its original echo chamber *(Donovan, 2020)*.

- Use of "non-persona" accounts

- Strategic targeting of vulnerable communities or demographics with tailored messaging.

- Platforms such as Facebook and YouTube use engagement-based ranking algorithms that prioritize shareable (but not necessarily truthful) content *(Tufekci, 2018)*.

In the Philippine context, the reliance on Facebook as a primary news source, coupled with the high smartphone penetration rate, increases both the speed and breadth of disinformation dissemination *(We Are Social & Hootsuite., 2024)*.

INFORMATION DISORDER CYCLE

1 CREATION AND PRODUCTION

2 DISTRIBUTION

3 AMPLIFICATION

Amplification occurs when content reaches a wide audience — either organically through user sharing, or via algorithmic boosts due to high engagement. This stage often:

- Reinforces cognitive biases, as users are more likely to believe and share information that aligns with their pre-existing beliefs (Kahneman, 2011).

- Benefits from repetition and visibility, which can create the illusion of truth, a phenomenon known as the illusory truth effect (Fazio, Brashier, Payne, & Marsh, 2015).

- Invokes coverage by mainstream media, which can unintentionally legitimize falsehoods by reporting on them without sufficient context or debunking. (Lewandowsky, Ecker, & Cook, 2017)

Why This Framework Matters

Understanding these phases is essential in designing effective interventions:

① CREATION

Legal accountability, regulation of troll farm financiers.

Creation can be addressed through better accountability for content producers and clearer legal definitions.

② DISTRIBUTION

Stronger platform ad transparency and local language moderation.

Distribution can be mitigated with stronger platform policies, especially for political ads and misinformation.

③ AMPLIFICATION

Digital literacy, fact-checking, and responsible reporting.

Amplification requires public education, digital literacy, and responsible media reporting.

Equally important is responsible content creation and amplification. As stakeholders noted, social media has turned everyone into a producer. Educating users not just to consume critically but also to create and share responsibly is a crucial, under explored element of Media and Information Literacy.

BARRIERS AND GAPS IN ADDRESSING INFORMATION DISORDERS

BARRIERS AND GAPS IN ADDRESSING INFORMATION DISORDERS

Even with ongoing efforts to combat information disorders in the Philippines, we still face significant challenges across legal, technical, socio-cultural, and context-specific dimensions. These barriers illustrate why disinformation remains deeply entrenched in our digital ecosystem.

LEGAL AND REGULATORY BARRIERS

The Philippines has introduced laws such as the **Cybercrime Prevention Act of 2012 (RA 10175)** and the **SIM Registration Act (RA 11934)** to address online harms. However, challenges in implementation and interpretation persist. In Disini v. Secretary of Justice (2014), the Supreme Court upheld the cyber libel provision of RA 10175, sparking concern among free expression advocates that the law could silence legitimate speech and discourage investigative journalism (Columbia Global Freedom of Expression, n.d.).

Recent proposals, such as **House Bill 3860 (Digital Media Integrity and Platform Accountability Act of 2025)**, illustrate evolving debates. The bill, filed by Rep. Antonino Roman III, seeks to establish a **Digital Media Standards Coalition** as a co-regulatory body. Supporters see it as a step toward accountability; critics warn of feasibility issues, risks of politicization, and potential misuse. These tensions underscore the trade-offs inherent in regulation: protecting the public versus safeguarding free expression.

TECHNICAL AND PLATFORM LIMITATIONS

The scale of online activity in the Philippines (nearly 91 million users - 78% of the population) makes effective moderation a daunting task. Automated systems struggle with **Filipino, Taglish, and local languages and dialects,** leaving much harmful content undetected (East Asia Forum, 2024). Platforms rely on uneven enforcement: while Facebook partners with fact-checkers, TikTok and YouTube were flooded with political propaganda during the 2022 elections, with looser oversight (Ong & Cabañes, 2018; SCMP, 2025).

New forms of manipulation are also emerging. In the **2025 midterms**, Democracy.Net.PH's #RespectMyPost campaign documented disinformation using **AI-generated profiles, face-swapped images**, **and synthetic videos,** which blurred the line between authentic and manipulated content. Without stronger transparency and local detection systems, such tactics threaten to overwhelm existing safeguards.

Another challenge is **platform monetization**. Disinformation spreads not only because it is engaging but also because it is profitable. Stakeholders argue that platforms must be prevented from **financially rewarding content proven to be false or harmful**, whether through ad revenue, algorithmic boosts, or sponsorship.

SOCIO-CULTURAL FACTORS

The way Filipinos share information is deeply personal and relational. Family group chats and peer networks often hold more weight than official sources. As a result, fact-checks may not reach their intended audiences or may even backfire, reinforcing false beliefs.

Surveys highlight persistent vulnerabilities: in 2021, 51% of Filipino adults admitted difficulty identifying misinformation on social media (GMA News, 2022). **The Philippine Information Agency (2025)** underscores that MIL interventions are particularly urgent for older adults and rural communities. Yet current MIL efforts often emphasize responsible consumption without equal focus on **responsible content creation and amplification**. Given that likes, shares, and comments make every user a content producer, this imbalance leaves a critical gap in literacy efforts.

Cultural values such as *pakikisama* (harmony) and *utang na loob* (debt of gratitude) are also exploited in information operations. Ong and Cabañes (2018) describe how disinformation networks employ influencers and micro-workers who leverage these values to normalize propaganda.

PHILIPPINE-SPECIFIC CHALLENGES

Disinformation in the Philippines often takes culturally embedded forms like parody, memes, and humor, these are content that can evade fact-checking yet normalize falsehoods through repetition. Linguistic diversity further hinders detection, as slang and idioms in regional dialects evade algorithmic tools.

Platforms such as Facebook groups and Messenger chats create **hyper-local echo chambers** that are difficult to monitor but highly influential. Meanwhile, the political economy of covert influence operations (Internews, 2022) reveals how influencers are contracted to shape narratives at scale, showing how money fuels disinformation alongside ideology.

Adding to the challenge is the reality that many Filipinos still struggle to distinguish between what's true and what's not online. According to the 2021 Social Weather Stations survey, 51% of adults struggled to identify misinformation on social media, highlighting deeper gaps in digital literacy and everyday critical thinking (GMA News, 2022).

Finally, the very architecture of platforms designed for virality favors volume and emotional engagement. In a country where social media penetration is among the highest globally, this accelerates the spread of misleading content beyond the reach of fact-checkers or corrective reporting.

BARRIERS AND GAPS IN ADDRESSING INFORMATION DISORDERS

Recent investigations further highlight how political actors can both benefit from and fall victim to the same disinformation infrastructures. PCIJ's Five Ways the Duterte Influence Machine is Deceiving Filipinos (Cabato & PCIJ, 2025) showed how coordinated networks of influencers, content farms, and loyal communities continue to manipulate narratives long after electoral campaigns. These findings reveal the **dual role of politicians as targets and beneficiaries**, which complicates regulatory responses. Lawmakers may call for stronger controls when disinformation is deployed against them, but remain silent or even complicit when it advantages their own political interests. This duality erodes public trust in government-led reforms and illustrates how disinformation is **structurally embedded in the incentives of Philippine politics.**

In sum, despite steps taken, major gaps remain. Legal frameworks risk being punitive or misused, platforms struggle with moderation at scale and resist transparency, and users remain vulnerable due to cultural habits and limited digital literacy. The Philippine information ecosystem, shaped by its linguistic diversity, cultural values, and political economy, requires solutions that are not only reactive but culturally grounded, context-specific, and multi-sectoral.

With this, it's important to understand how Filipinos create, share, and engage with content, which is key in designing solutions that are **CULTURALLY GROUNDED** to not only counter disinformation, misinformation, and malinformation, but also strengthen trust in information.
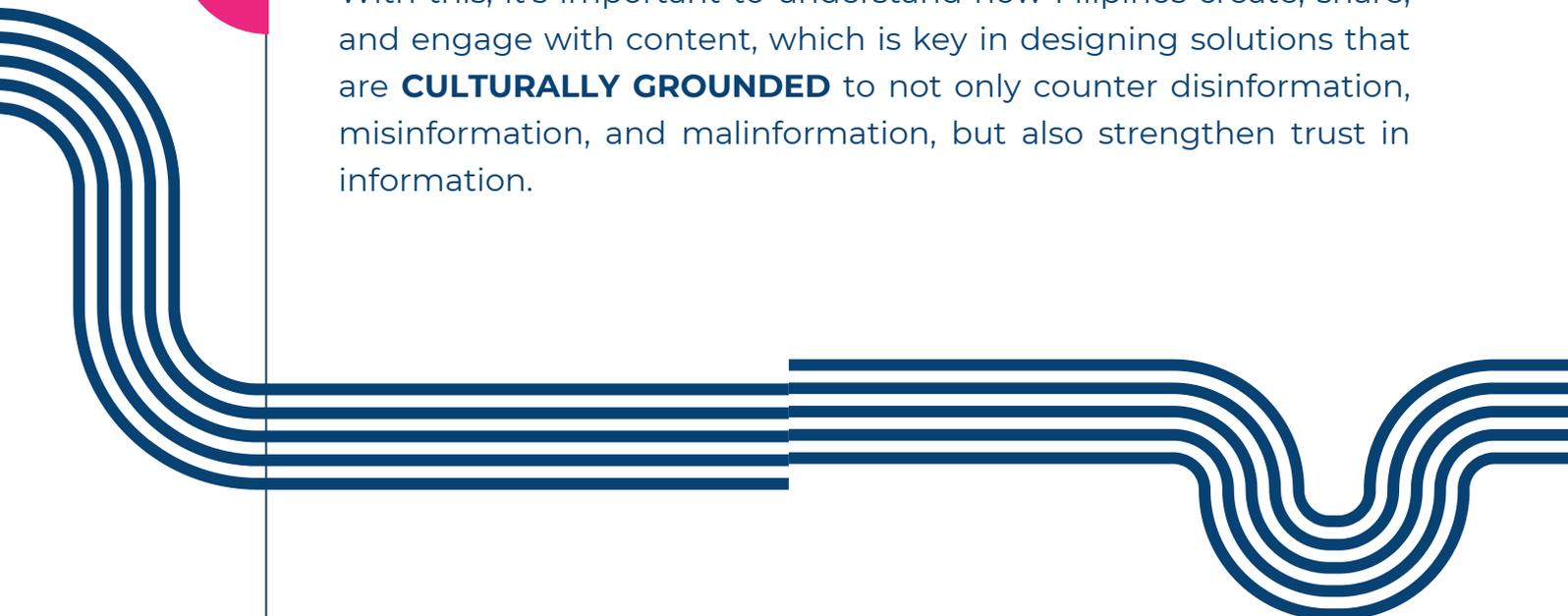
STAKEHOLDER PERSPECTIVES

Effectively tackling information disorders means that no single group holds the answers. Each sector approaches the problem from a different lens, shaped by its goals, limitations, and incentives. Understanding these diverse perspectives is key to crafting sustainable, long-term solutions.

The Government Agencies and Legislators

Government actors face the difficult task of legislating digital spaces while upholding constitutional guarantees of free expression. Efforts to regulate harmful content are often framed around **national security and electoral integrity,** but these raise concerns about political overreach, selective enforcement, and due process.

Fragmentation compounds the challenge: responsibility is spread across agencies like the DICT, NPC, COMELEC, and DOJ, none of which have sufficient technical or staffing capacity. Public distrust of political motives further complicates governance, as legislators themselves have both been **targets and beneficiaries of disinformation networks** (Ong, Tapsell, & Curato, 2019).

The Social Media Platforms

Platforms balance a three-way tension: **moderation vs. free expression vs. business incentives**. Campaigns like TikTok's #ThinkTwice (TikTok Newsroom, 2025) demonstrates attempts at user education. Yet, decisions remain guided by commercial priorities, and enforcement in the Philippines is inconsistent.

A critical gap lies in **localization**. Global moderation frameworks are often poorly adapted to Filipino languages, humor, and cultural nuances. As a result, harmful content is under-detected, while community norms of sharing remain unchecked (GMA News, 2025)

The Academe and Researchers

Academics provide evidence-based insights by mapping propaganda networks, tracing narrative flows, and highlighting intersections between politics, platforms, and influencer economies (Ong & Tapsell, 2020; ISEAS, 2023). However, researchers face **limited access to platform data** and difficulty in translating findings into enforceable policy.

At the same time, Philippine academia is beginning to explore **PR and advertising's role in shaping disinformation ecosystems,** filling a gap in the current literature and pointing to new fronts for accountability.

Civil Society and Advocacy Groups

Civil society plays a watchdog role, emphasizing **transparency, accountability, and public empowerment**. Coalitions like #FactsFirstPH bring together media organizations, NGOs, and fact-checkers for collective verification.

Yet these groups operate under pressure: **resource limitations, political harassment, and difficulties in reaching audiences beyond echo chambers** undermine their impact. Their work is vital but requires stronger partnerships and protection mechanisms (Media Practice Support, 2023).

The Media Organizations

Philippine newsrooms remain central to fact-checking and investigative journalism. Many have embraced **digital forensics and cross-newsroom collaborations** to counter false narratives.

However, their independence is threatened by shrinking advertising revenues, online harassment, accusations of bias, and punitive measures like cyber libel laws (Cabato, 2024). These constraints make sustaining robust journalism difficult at a time when accurate reporting is most needed.

The Communications and Marketing Industry

PR, advertising, and digital marketing agencies are critical but ambivalent actors in the information disorder landscape. On one hand, they are hired to craft campaigns that seed narratives, manipulate conversations, or legitimize covert influence operations. Internews (2022) documents how PR and influencer networks were contracted in the 2022 elections, producing data-driven estimates of the costs of hiring influencers to amplify political messages. This study highlights the political economy of disinformation in the Philippines, where professionalized communications services fuel covert propaganda.

On the other hand, the same industry has the skills and reach to be a powerful ally in counter-disinformation campaigns. By engaging marketing professionals in developing ethical standards, transparency practices, and voluntary codes of conduct, the sector can help shift incentives away from covert manipulation and toward civic empowerment.

The General Public

Filipino users are not passive consumers; they are **active interpreters and amplifiers of information**. Sharing is influenced by trust in family, friends, and influencers, as well as cultural values like *pakikisama* (harmony) and *utang na loob* (gratitude).

Low levels of digital literacy and high emotional engagement drive vulnerability to false content. Surveys show that **over half of Filipino adults struggle to identify misinformation** (GMA New, 2022). Yet, the public also represents the **greatest potential for resilience** if equipped with Media and Information Literacy (MIL) that covers not only responsible consumption but also responsible creation and amplification of content.

MULTI-SECTORAL SOLUTIONS
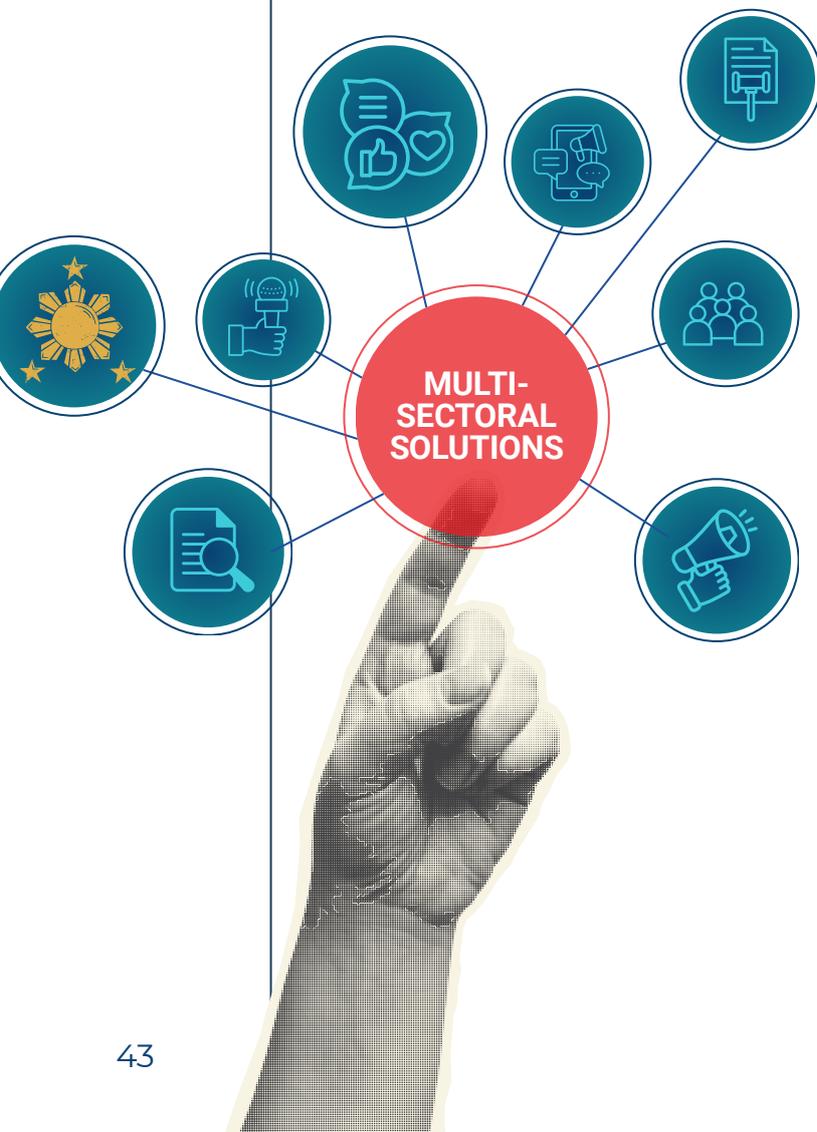
> " *Taken together, these perspectives reveal that information disorder is deeply woven into **Philippine political, cultural, and economic life**. Addressing it requires moving beyond isolated interventions toward **multi-sectoral solutions** that acknowledge stakeholder constraints, engage the communication industry, and rebalance incentives in the digital ecosystem.*

CURRENT AND EMERGING PUBLIC-PRIVATE SECTOR COLLABORATION

CURRENT AND EMERGING PUBLIC-PRIVATE SECTOR COLLABORATION

Addressing information disorders requires collaborative engagements from both public and private stakeholders. This section examines ongoing initiatives, synthesizes findings from focused group discussions (FGDs), and outlines opportunities for deeper and more strategic collaboration.

ONGOING COLLABORATIONS IN THE PHILIPPINES

Fact-Checking Partnerships

The **#FactsFirstPH** coalition, formed by media outlets, civil society groups, and academic institutions, mobilized over 100 organizations to publish more than 1,500 fact checks during election cycles (Meedan, 2022; Interaksyon, 2022). These partnerships have improved information verification, but sustaining them requires long-term funding and systemic integration into platform workflows.

Educational Campaigns

TikTok Philippines launched the #ThinkTwice campaign, collaborating with creators and youth groups to encourage critical thinking and responsible content sharing (TikTok Newsroom, 2025).

Google Philippines introduced Be Internet Awesome in the Philippines, teaching children safe and ethical internet use, with modules on digital health, literacy, and ethics (GMA Network, 2021).

CURRENT AND EMERGING PUBLIC-PRIVATE SECTOR COLLABORATION

> *These initiatives demonstrate progress, yet disinformation tactics are evolving, becoming more localized, more visual, and increasingly AI-driven. One-off responses are no longer enough; what is needed is **sustained, transparent, and trust-based collaboration** between government, platforms, educators, industry, and communities.*

Crisis Response Protocols

Platforms like Meta and TikTok have also established dedicated coordination mechanisms with key government agencies such as the Commission on Elections (COMELEC) and the Department of Health (DOH) to mitigate spikes in harmful misinformation during high-stakes events proactively:

Meta (Facebook): During the COVID-19 pandemic, Meta expanded its content moderation policies in the Philippines to specifically cover false or misleading claims related to the coronavirus, vaccines, and public health guidance. It leveraged its global human rights policy frameworks to support takedown decisions and collaborated with local partners to address rapidly evolving health disinformation (Meta Philippines HRIA, 2023). In the lead-up to the 2022 general election, Meta also activated an Elections Operations Center in partnership with the Philippine Commission on Elections (COMELEC) to monitor and address election-related misinformation and coordinated networks in real-time (Meta, 2022).

TikTok Philippines launched an Elections Center in collaboration with COMELEC, NAMFREL, and LENTE. This platform feature provided verified election content and included mechanisms for removing posts that violated civic integrity standards (TikTok Newsroom, 2025).

Reflections from Focus Group Discussions and Roundtables

The following insights from two focused group discussions (FGDs), with representatives from different social media platforms and a media and information literacy roundtable, shed light on the current gaps, challenges, and opportunities that lie ahead.

From the FGD with Social Media Platforms:

Policymakers often **lack a nuanced understanding** of how disinformation, misinformation, and malinformation manifest across different platforms. Participants noted the differences in platform architecture, user behavior, and moderation capacities that influence how information disorders evolve.

Concerns were raised over **politically motivated content takedown** requests. While platforms implement content moderation policies, the perception of partisan enforcement can undermine public trust, highlighting the need for platform transparency and due process safeguards.

Participants shared that social media should **not be regulated in the same manner as traditional broadcast or print media** (e.g., through franchise-style licensing). Instead, regulatory approaches must recognize the dynamic and interactive nature of these digital platforms.

Effective public policy must be **grounded in a clear understanding of internal platform systems** such as algorithmic recommendation engines, community standards enforcement, and business models that prioritize engagement.

CURRENT AND EMERGING PUBLIC-PRIVATE SECTOR COLLABORATION

International models were referenced, such as Thailand's digital identity verification initiatives, Malaysia's self-regulation efforts through industry councils, and the European Union's Digital Services Act emphasizing platform accountability and user rights.

Structural challenges continue, including **fragmented policy coordination across agencies**, conflicting mandates among government units, and insufficient technical capacity to develop agile and forward-looking digital regulations.

Government-led regulation of misinformation—where the government imposes its own definition of misinformation and consequently orders takedowns—may present **substantial threats to upholding free and democratic speech in the country**, while rendering social media platforms vulnerable to partisan interests. It may be worthwhile to consider an alternative regulatory framework crafted and implemented by an independent council or body, which may be composed of multiple relevant stakeholders not captured by vested political agendas.
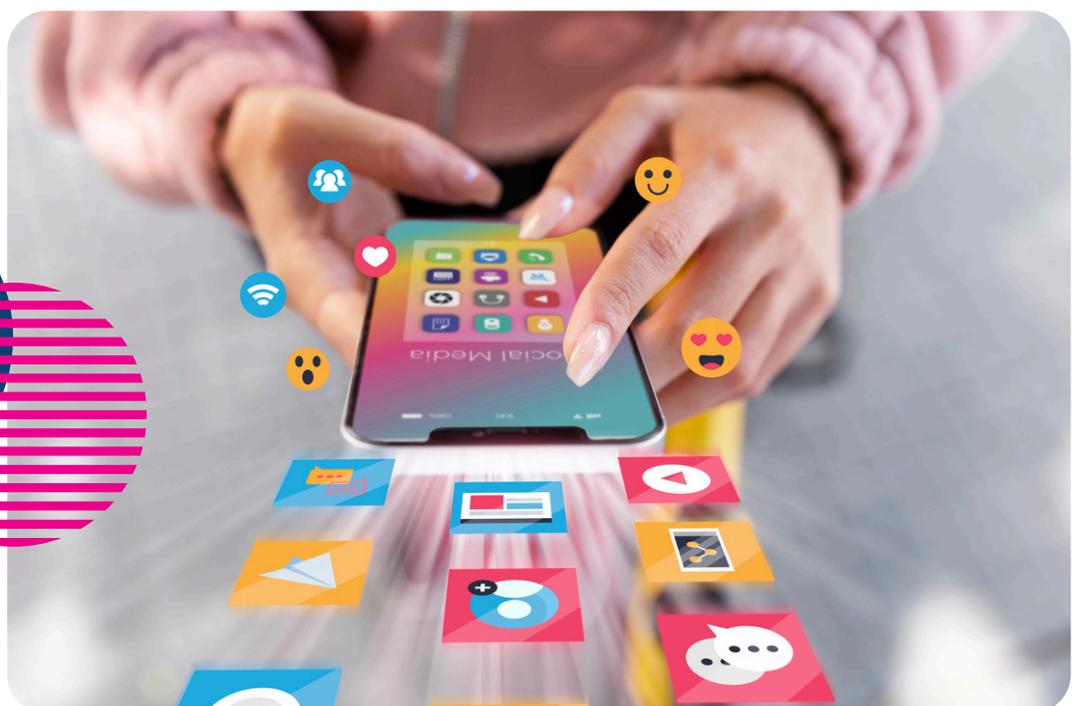
From the Media and Information Literacy (MIL) Roundtable:

Participants discussed the overwhelming volume of online content, which leads to overload and limits users' ability to assess its credibility. Many users, especially youth and marginalized communities, lack access to tools and support that build critical thinking skills in digital environments.

Algorithms that prioritize virality often amplify polarizing or emotionally charged content. Participants emphasized that this system design contributes to the rapid spread of false or misleading information regardless of user intent.

There were calls for social media platforms to also take greater responsibility in ensuring safe digital environments by investing in localized content moderation, improving transparency around algorithmic curation, and enabling user controls such as content filters and reporting tools.

Policy recommendations included requiring platforms to register with the Securities and Exchange Commission (SEC), mandating funding for fact-checking initiatives, supporting the establishment of a self-regulating organization, and making algorithmic practices more transparent.

Educational Front: Strengthening Media and Information Literacy (MIL)

Participants underscored MIL as a long-term resilience strategy, with emphasis not just on consumption, but also responsible content creation and amplification, recognizing that every post, like, and share constitutes participation in the information ecosystem.

Key Areas of Focus:

MIL should extend beyond schools into barangays, communities, and even platforms themselves. The curricula should be co-developed by educators, civil society, and platforms, aligning with real-world digital experiences.

> **Additional Resource Allocation:** Greater investment in MIL programs is needed, ensuring schools and communities have access to quality training materials, teachers, and technology.

> **Clear Leadership and Coordination:** A lead agency should be designated to coordinate MIL initiatives across government, education, and civil society, avoiding fragmented or duplicative efforts.

> **Equity in Education:** Programs must prioritize underserved and marginalized communities so that media literacy is accessible to all, not just urban or privileged populations.

Key Areas of Focus:

Shifts in Pedagogy: Teaching approaches should move toward student-centered learning that empowers learners to think critically, question information, and apply skills in real-world contexts.

Integration Across Levels: MIL should be mainstreamed across all levels of education, from primary school through higher education and integrated into various subject areas rather than treated as a stand-alone module.

Harnessing Technology: Digital tools, online platforms, and interactive learning methods should be leveraged to deliver MIL in engaging ways that mirror how young people consume content.

Responsible Content Creation and Amplification: Participants stressed that MIL must go beyond teaching responsible consumption. Because social media gives everyone a voice, transforms every user into a potential producer of content (through posts, comments, likes, and shares) learners need guidance on ethical content creation, responsible amplification, and understanding the consequences of online participation. This ensures students not only learn to discern information, but also contribute positively to the digital ecosystem.

GLOBAL MODELS & FRAMEWORKS

UNESCO Global Standards for MIL

UNESCO's guidelines emphasize dual competencies in access, analysis, evaluation, creation, and sharing of information underscores the importance of active participation in media ecosystems. They define MIL as a set of life skills relevant to digital citizenship and encourage national adaptation of these standards.

www.unesco.org/en/media-information-literacy

Finland's National Media Literacy Policy

Finland exemplifies systemic integration of MIL across early childhood, primary, and secondary education. Media literacy is taught through interdisciplinary methods in subjects like language, social studies, and digital design, coordinated by the National Audiovisual Institute (KAVI). Citizens are also encouraged to produce media content ethically, and widespread public awareness campaigns reinforce these values.

www.oph.fi/en/education-and-qualifications/multiliteracy-and-media-literacy

CURRENT AND EMERGING PUBLIC-PRIVATE SECTOR COLLABORATION

Expanding the Collaborative Lens

Stakeholders emphasized that collaboration must also extend to the communication and marketing industry, which often designs and deploys campaigns that manipulate conversations online. Research shows that PR and advertising agencies were contracted during the 2022 elections to coordinate influencer-driven disinformation campaigns (Internews, 2022). Industry actors must therefore be engaged in developing ethical standards, voluntary codes of conduct, and accountability frameworks, ensuring that the tools of persuasion are not weaponized against democratic processes.

Why This Matters

Both FGDs and roundtables made clear: successful collaboration requires **trust, transparency, and long-term commitment**. This includes not only government, platforms, and civil society, but also industry professionals and communities. The challenge is not simply to react to each new wave of disinformation but to **institutionalize partnerships** that build a healthier and more resilient digital public sphere.

MIL that embraces both consumption and creation reflects the realities of today's media landscape where individuals are both audiences and publishers. UNESCO's global framework provides a foundational template for responsible citizenship in the digital age, while Finland demonstrates how these principles can be embedded system-wide, producing resilient, media-literate communities.
Integrating insights from these models strengthens the credibility of MIL recommendations and aligns them with internationally validated best practices.

INSTITUTIONAL AND POLICY RESPONSES
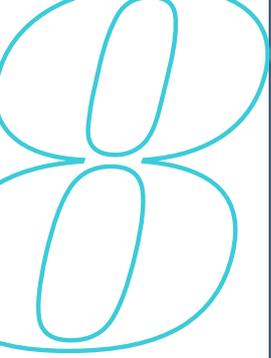
INSTITUTIONAL AND POLICY RESPONSES

Addressing information disorders requires both domestic reforms and alignment with global best practices. The Philippines has made some legislative and institutional efforts, but gaps remain in enforcement, coordination, and adaptability to emerging threats such as AI-driven disinformation.

A Philippine Context
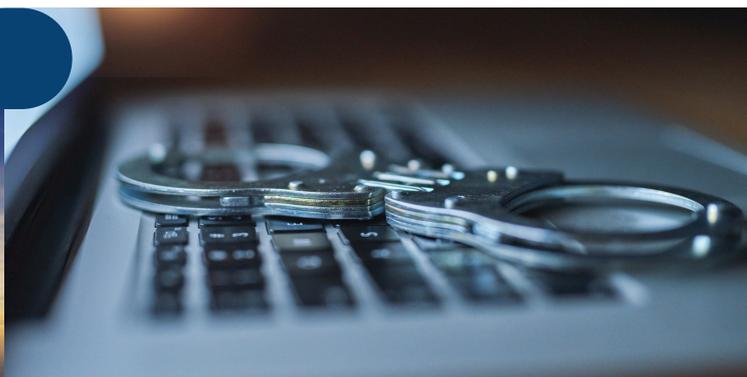
The Philippines has introduced several laws and proposals to regulate online behavior and safeguard digital integrity:

Cybercrime Prevention Act of 2012 (RA 10175): Criminalizes a range of online offenses, including cyber libel. While intended to curb online harms, the law has been criticized for its potential misuse to silence journalists and activists (Disini v. Secretary of Justice, 2014; Columbia Global Freedom of Expression, n.d.).

SIM Registration Act (RA 11934): Requires all SIM cards to be registered, aiming to reduce anonymity in online activity. Implementation has faced challenges, particularly with fake IDs and fraudulent registrations.

Pending Anti-Fake News Bills: Since 2017, multiple bills have been filed, but many focus on punitive measures that could undermine freedom of expression. Overview: Various bills have been introduced in Congress to penalize the intentional spread of false information. While these proposals demonstrate urgency, critics argue that vague definitions and punitive measures may undermine press freedom and be weaponized against critics. (Center for Media Freedom and Responsibility, (CMFR 2022)
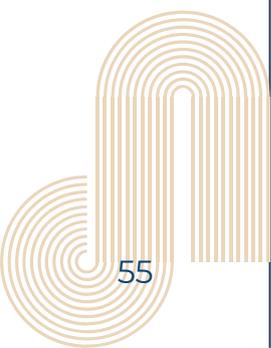
A Philippine Context

House Bill 3860 – Digital Media Integrity and Platform Accountability Act (2025): Filed by Rep. Antonino Roman III, this bill introduces the Digital Media Standards Coalition (DMSC) as a non-governmental, non-partisan, and independently governed body. Inspired by existing Philippine self-regulation models such as the Ad Standards Council (ASC) in advertising and the Kapisanan ng mga Brodkaster ng Pilipinas (KBP) in broadcasting, the DMSC aims to act as a moral compass for digital governance.

The DMSC's role would be to provide ethical advisories, issue reputational sanctions, and conduct ethical audits of platforms, consistent with data privacy protections. Importantly, its powers exclude criminal or administrative penalties, ensuring it cannot censor or silence voices. Instead, platforms would be required to establish a legal presence in the Philippines and cooperate in good faith with DMSC processes. Non-compliance could affect their regulatory standing with government agencies, ensuring accountability without overstepping constitutional boundaries.

By design, **HB 3860 seeks a middle ground between state overreach and unchecked, profit-driven platform governance.** It positions the DMSC as a values-driven coalition that can guide platforms and government in ethically addressing disinformation, misinformation, and malinformation.

A Philippine Context

Gaps Between Policy and Practice

Despite these legislative frameworks, disinformation practices in the Philippines often outpace institutional responses:

- In the 2025 midterm elections, the #RespectMyPost initiative documented coordinated disinformation campaigns such as the "Duterte Kidnapping" hoax and allegations of overseas election fraud. These relied on AI-generated profiles, face-swapped images, and synthetic videos, exposing institutional blind spots in detecting and addressing technologically sophisticated manipulation.

A People's Republic of China-aligned actor published several videos amplifying the Duterte kidnapping narrative. The title cards, written in Filipino, prominently featured the word "kidnapping"

- Internews (2022) mapped the political economy of covert influence operations, estimating the costs of hiring influencers and PR agencies during the 2022 elections. This study highlighted how professionalized communications services play a central role in amplifying disinformation, showing the need for both regulatory oversight and industry accountability.

These cases underline how disinformation evolves faster than current policy mechanisms, leaving regulators and watchdogs playing catch-up.

B Global Examples

Various countries around the world have taken both stricter and more cooperative measures to control harmful online content..

European Union – Digital Services Act (DSA)
Introduced by: European Commission (led by Executive Vice-President Margrethe Vestager and Commissioner Thierry Breton)
Adopted: 2022; entered into force in 2023

Key Provisions: Mandates transparency in content moderation
Requires platforms to assess systemic risks (e.g., disinformation, election interference)

Applies more stringent rules for "very large online platforms" (VLOPs)
Significance: The DSA is widely regarded as a gold standard for comprehensive, rights-based digital regulation. (European Commission, 2023)

Singapore – Protection from Online Falsehoods and Manipulation Act (POFMA)
Passed by: Singapore Parliament in 2019

Key Sponsor: Ministry of Law, under the leadership of Minister K. Shanmugam

Overview: This law empowers the government to issue correction directives or takedown orders for content deemed false and harmful to public interest. While praised for swift implementation, concerns have been raised about overreach and political misuse. (Ministry of Law Singapore, 2019)

B   Global Examples

Australia – Voluntary Code of Practice on Disinformation and Misinformation
Developed by: Digital Industry Group Inc. (DIGI), representing Google, Meta, TikTok, Twitter (now X), and Microsoft
Endorsed by: Australian Communications and Media Authority (ACMA) in 2021

Overview: Developed with major platforms, the code outlines best practices on content moderation, transparency, and partnership with fact-checkers. Its voluntary nature, however, has led to calls for stronger regulatory enforcement. (Australian Communications and Media Authority, 2021)

Together, these models represent a spectrum from state-driven regulation to voluntary industry self-regulation, with the Philippines exploring hybrid approaches such as the proposed DMSC under HB 3860.

C   Civil Society and Platform Efforts

Fact-Checking Collaborations: #FactsFirstPH mobilized over 100 organizations during election cycles, strengthening verification ecosystems.

Transparency Reports: Platforms like Meta and TikTok have begun publishing transparency data, but gaps remain in local adaptation and accessibility.

AI Moderation and Labeling: Platforms are experimenting with content labeling (e.g., "Made with AI"), but coverage of Filipino languages and contexts is limited.

Voluntary Codes of Conduct: Stakeholders have proposed developing Philippine-specific voluntary codes, involving platforms, academe, and civil society, to standardize ethical practices.

Emerging Issues and Trade-offs
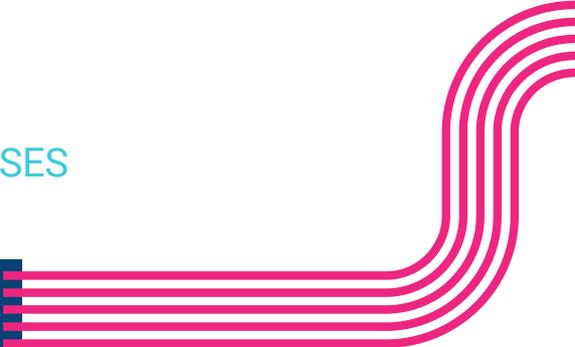
Responsible Monetization: Platforms still profit from viral, high-engagement content—even when harmful. Restricting ad revenue or sponsorships for flagged disinformation could realign incentives.

Fragmented Oversight: Oversight remains split among DICT, COMELEC, NPC, and DOJ, each constrained by technical and resource limitations.

Local Language Gaps: Philippine disinformation often uses humor, parody, memes, and dialect-specific slang, which evade global moderation rules.

Regulatory Trade-offs: Government-heavy solutions risk politicization, while voluntary models risk ineffectiveness. HB 3860 illustrates this tension—offering a hybrid "middle ground" model, but with open questions around independence, enforcement, and sustainability.

Institutional responses in the Philippines remain still developing, fragmented, and reactive, even as disinformation actors adopt increasingly sophisticated strategies. Global examples show pathways forward, but local adaptation is essential. HB 3860's proposed Digital Media Standards Coalition represents a potential innovation in co-regulation—balancing freedom and accountability, ethical guidance and enforcement, and state oversight with industry-led responsibility. Its success, however, will depend on clear mandates, transparency safeguards, and active participation from civil society, academe, and the communication industry.

HOUSE BILL 3860

Policy Snapshot: Digital Media Standards Coalition

? What is it?

- A bill filed by Rep. Antonino Roman III in 2025.
- Establishes the Digital Media Standards Coalition (DMSC), a non-governmental, non-partisan, independently governed body.
- Functions as an ethical authority for digital content governance, especially disinformation, misinformation, and malinformation.

⚙ How does it work?

- Provides ethical advisories and reputational sanctions (not criminal penalties).
- Conducts ethical audits of platforms, with safeguards under the Data Privacy Act.
- Platforms must establish a legal presence in the Philippines and cooperate in good faith with DMSC processes.
- Non-cooperation may affect their regulatory standing before state agencies.

★ Why it matters

- Offers a "middle ground" between government overreach and opaque, profit-driven platform policies.
- Inspired by Philippine self-regulation bodies: Ad Standards Council (ASC) and Kapisanan ng mga Brodkaster ng Pilipinas (KBP).
- Promotes transparency, ethical responsibility, and independent oversight.

💡 Opportunities

- Builds accountability without censorship.
- Encourages platforms to align with ethical standards.
- Establishes a permanent structure for co-regulation.

⚠ Risks & Open Questions

- Independence: How to ensure DMSC is not captured by political or commercial interests?
- Enforcement: Without criminal powers, will advisories and reputational sanctions be enough?
- Sustainability: Requires long-term funding, stakeholder buy-in, and public trust.
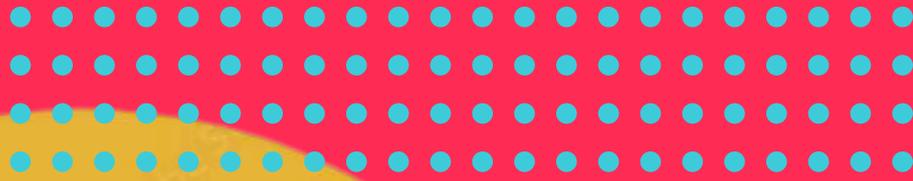
POLICY CONVERSATIONS AND POINTS OF EXPLORATION

POLICY CONVERSATIONS AND POINTS OF EXPLORATION

The conversations that took place during the FEU Public Policy Center's roundtables and focus group discussions offered a valuable window into how different sectors are thinking about the challenge of disinformation, misinformation, and malinformation. While no single solution emerged, the discussions revealed a wide range of ideas shaped by each group's experiences, responsibilities, and insights.

This section brings together those ideas and is meant to help spark further dialogue and action. Some of the proposals center on improving policy and regulation. Others focus on practical steps that platforms, industry, or communities can take.

For Government

Stakeholders underscored the need for the government to play a more proactive role in safeguarding the public from harmful content.

- **Integrating Media and Digital Literacy:** Participants stressed embedding media and digital literacy into the national education curriculum, co-developed with educators, civil society, platforms, and communications practitioners to reflect real-world online use. MIL should also extend beyond classrooms to reach teachers, parents, and lifelong learners.

- **Independent Oversight:** Proposals included creating oversight bodies to audit platforms, strengthen accountability, and uphold democratic values through transparent reporting.

- **Evidence-Based Policymaking:** There was interest in legislation supporting open data sharing between platforms, researchers, and regulators, with safeguards for privacy and security. Instead of blanket open-data laws, participants proposed developing open data standards through a self-regulatory organization (SRO), in coordination with the National Privacy Commission, ensuring anonymization and aggregation in line with the Philippine Data Privacy Act.

For Social Media Platforms

Platforms were urged to move beyond global templates and adopt context-aware moderation tailored to Philippine political, linguistic, and cultural realities.

- **Investing in Localization:** Hire Filipino and regional language teams, enhance community reporting, and provide users with stronger filtering tools.
- **Algorithmic Transparency:** Stakeholders called for disclosure on how content is promoted or suppressed, particularly for political, health, and news content.
- **Collaborating with Civil Society:** Build awareness campaigns, amplify credible voices, and co-create interventions with local groups.

For Civil Society and the Academic Sector

Civil society and academe were recognized as bridges between communities, government, and platforms.

- **Grassroots Campaigns:** Promote hyperlocal, participatory media literacy programs through schools, barangays, and youth organizations.
- **Multi-Sector Dialogues:** Establish sustained spaces for trust-building and collaboration beyond one-off events.
- **Locally Grounded Research:** Support research disaggregated by language, geography, and demographic profile to design effective interventions.
- **Self-Regulatory Organization (SRO):** Have SRO composed of regulators, industry leaders, and independent experts. This can be lighter alternative to heavy government regulation, modeled after ASC (Ad Standards Council) or KBP (Kapisanan ng mga Brodkaster ng Pilipinas) of the marketing and advertising industry.

For the PR, Marketing, and Communications Industry

Participants stressed that the communications industry must be included in policy conversations. Agencies and consultancies often provide the infrastructure for influence operations, from orchestrating covert influencer campaigns to running data-driven advertising strategies that can blur the line between persuasion and manipulation.

- **Accountability and Standards:** The industry should be encouraged (or mandated) to adopt ethical standards and voluntary codes of conduct, similar to the Ad Standards Council (ASC) model. This could include guidelines on client disclosure, limits on political campaign contracts, and bans on deceptive practices.

- **Transparency in Campaigns:** Agencies working on political or issue-based campaigns should disclose funding sources and avoid covert amplification strategies.

- **Positive Engagement:** At the same time, PR and marketing professionals could play a constructive role in counter-disinformation efforts by applying their expertise in messaging, narrative design, and amplification to promote civic education, fact-based campaigns, and digital literacy initiatives.

- **Professional Education:** The industry should integrate ethical digital communications into training programs, encouraging practitioners to recognize their role in either enabling or countering disinformation.

*As dialogue continues, these points of exploration can inform future collaborations and help shape interventions that are **inclusive, culturally grounded, and responsive to the needs of Filipinos online**. Crucially, expanding the conversation to include PR, marketing, and communications professionals acknowledges that information disorders are not only about platforms and users but also about the industries that profit from shaping digital narratives. By recognizing their dual role as both enablers and potential allies, policymakers and stakeholders can better align incentives toward building a healthier information ecosystem.*

CONCLUSION

CONCLUSION
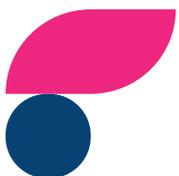
The spread of disinformation, misinformation, and malinformation poses profound risks to Philippine democracy, social trust, and the integrity of public discourse. These information disorders are not the result of a single factor but rather the outcome of **intertwined dynamics**: platform design and business models that reward virality, user behaviors shaped by cognitive biases and cultural practices, and political and economic incentives that exploit these vulnerabilities.

The **Philippine context** heightens these challenges. With one of the highest social media usage rates in the world, a fragmented political environment, and persistent gaps in digital literacy, harmful content finds fertile ground to circulate widely. Platform moderation struggles with linguistic diversity and cultural nuance, while policymakers face the difficult task of balancing regulation with democratic freedoms. Adding to this complexity, the communication, marketing, and PR industries often play dual roles: they can be hired to orchestrate covert influence campaigns, yet they also possess the expertise to design ethical, civic-minded interventions.

Despite these challenges, the discussions convened by the FEU Public Policy Center reveal a growing recognition of **shared responsibility.** Government, platforms, civil society, academe, the communication industry, media, and the general public all have distinct but complementary roles to play in shaping a healthier information ecosystem. Emerging initiatives like fact-checking networks, educational campaigns, public-private partnerships, and global regulatory models offer important starting points, but they must evolve into **sustained, transparent, and inclusive collaborations.**

CONCLUSION

Importantly, there is growing momentum around **innovative governance experiments**, such as Rep. Antonino Roman III's proposed Digital Media Standards Coalition (HB 3860), which seeks a middle ground between state overreach and opaque platform self-regulation. This illustrates how policymakers are beginning to explore hybrid models that combine accountability, transparency, and independence.

THERE ARE NO QUICK FIXES TO THE DISORDER OF INFORMATION

"

What is required is a continuous commitment to dialogue, trust-building, and collective action. By grounding efforts in cultural realities, strengthening accountability frameworks, engaging industries that shape narratives, and empowering citizens with the skills to critically consume and responsibly create content, stakeholders can begin to disrupt the cycle of harmful information.

This paper does not prescribe specific solutions; instead, it highlights the urgency of working together to protect the integrity of our digital public sphere and to safeguard the democratic values that depend on it.

ABOUT THE AUTHOR

Kankan Ramos is a marketing and communications leader who brings creativity, strategy, and purpose together. As Chief Operating Officer of InvictusPH, she drives transformative, collaborative partnerships between brands and communities, always guided by her belief that meaningful impact begins with a can-do mindset. Beyond the boardroom, she is a passionate educator, teaching Digital and Social Media Marketing at Ateneo de Manila University and St. Theresa's College QC, where she champions her advocacy for the responsible use of social media and inspires the next generation to become thoughtful, ethical digital citizens.

Her influence extends into the cultural and community sectors as a board member of Alliance Française de Manille, where she helps foster cross-cultural dialogue and artistic exchange, and as a board member of the St. Theresa's College QC Alumnae Association, where she contributes to initiatives that empower women and support educational opportunities. A practicing artist with multiple solo and group exhibitions in the Philippines and abroad, she seamlessly integrates her artistry with her leadership in marketing and communications.

Through every role she takes on—executive, mentor, advocate, and artist—Kankan empowers others to find their voice, embrace possibility, and build communities rooted in creativity, responsibility, and collaboration. She is also an admin of Borahae From Manila, a thriving ARMY (BTS fan) community of over 30,000 members.

Clara Quiambao is a Mass Communications graduate major in Broadcast and currently heads social media at Invictus PH. Her work examines the intersections of media, technology, and public discourse in the Philippine context, while also considering global trends and platform dynamics.

CRITIQUES FROM REFEREES

RACHEL E. KHAN, DPA

Dr. Rachel E. Khan is a Professor of Journalism at the University of the Philippines (UP) College of Media and Communication. She is concurrently a co-coordinator of Tsek.ph, the pioneer fact-checking coalition of media, academe and civil society in the country. She serves as the Editor-in-Chief of the Asian Congress for Media and Communication (ACMC) Journal and a Board member of the Pacific Media Journal. Prior to the academe, she was a senior reporter for BusinessWorld. She holds an M.S. in Journalism with a New Media concentration from Columbia University, New York under the Fulbright program; and, a Doctorate in Public Administration from UP NCPAG. She also has a certificate in AI Journalism from the London School of Economics. She has authored several books and journal articles on journalism, media literacy, and the internet.

The discussion paper, initiated by the FEU Public Policy Center, is a valuable and well-structured contribution to the complex discourse surrounding disinformation, misinformation, and malinformation in the country. The paper's primary strength lies in its ability to serve multiple functions: a comprehensive explainer, a successful aggregator of existing academic knowledge, and a source of practical, multi-stakeholder policy suggestions.

Clarity and Utility: An Essential Explainer

For anyone new to the gravity and complexity of the Philippine information ecosystem—whether a foreign policy analyst, an investor, or a journalist—this document functions as an excellent, self-contained explainer. The paper clearly and methodically breaks down the problem even as it defines the nuances between disinformation, misinformation, and malinformation, ensuring that the reader understands not just what is false, but why it is being spread.

CRITIQUES FROM REFEREES

By detailing the drivers—political, economic, and psychological—it frames the issue as a systemic dysfunction rather than a simple moral failure. The explicit mention of the Philippines' high social media penetration and deeply polarized political climate establishes the unique intensity of the problem, making the paper immediately accessible and highly useful for those seeking a foundational understanding of this national crisis.

Echoing and Advancing Existing Research

A key value of this discussion paper is its successful aggregation of insights that have been surfaced by previous academic and journalistic studies. It effectively synthesizes the known challenges, including:

> The role of troll farms and coordinated inauthentic behavior in content creation;
>
> The problematic nature of platform architectures that incentivize virality over truth;
>
> The vulnerability stemming from low media literacy and pervasive cognitive biases like confirmation bias.

However, the paper moves beyond mere repetition by incorporating original, high-value insights derived from the FEU Policy Center's focus group discussions and roundtables. These insights introduce crucial, context-specific factors that often escape analysis:

> The exploitation of Filipino socio-cultural values like pakikisama and utang na loob, the ability of disinformation to evade detection through the use of local languages, humor, and parody, and the difficulty fact-checks face in penetrating close-knit online networks like family group chats. This ground-up perspective validates the existing research while significantly enriching it with practical, on-the-ground context.

CRITIQUES FROM REFEREES

Policy Suggestions and Shared Responsibility

Finally, the paper makes a compelling case for policy that recognizes the complexity of the disorder. By concluding that there is "no single cause and no single solution," it rightly steers the conversation toward shared responsibility.

The policy suggestions embedded within the analysis are practical and pointed:

Elevating MIL: Emphasizing that Media and Information Literacy must evolve "beyond consumption" to address responsible creation and amplification is a critical, actionable suggestion.

Platform Accountability: Demanding increased transparency regarding algorithmic curation and the localization of content moderation directly addresses platform-level failures.

Engaging the Marketing Industry: The suggestion to engage the marketing and communication sector—which possesses the design and persuasive expertise used to spread disinformation—to instead build ethical, civic-minded counter-campaigns is a highly strategic and useful policy suggestion.

This approach—which calls for collaboration between government, platforms, civil society, and the private sector—provides a useful roadmap for moving past the polarizing debates of state censorship versus pure platform self-regulation. It offers tangible areas where collaborative action can lead to rational progress in addressing this problem.

CRITIQUES FROM REFEREES

NOEMI LARDIZABAL-DADO

Noemi Lardizabal-Dado, known as @momblogger on social media, has been an active presence online since 1995. She is a co-convenor for the Consortium on Democracy and Disinformation and a member of #FactsFirstPH. She also serves as a columnist for the Manila Times and maintains several blogs, including the award- winning **aboutmyrecovery.com** and citizen advocacy site, **blogwatch.tv**

The paper is timely and rooted in Philippine realities such as heavy social media use, polarized discourse, and regular election cycles. As a discussion draft, it does what it set out to do. It consolidates research, meeting notes, and community perspectives to identify key issues, limits, and potential areas for collaboration. Read in that light, the range of options is appropriate. Its value lies in mapping the terrain, not prescribing a step-by-step plan. To help readers move from assessment to potential routes without turning this into a prescriptive plan, the draft could add a short section that shows how any chosen option might be tested in Year 1. That section can name key questions, temporary leads, rough timelines, and the evidence that shows it is worth scaling.

Accuracy. The wording is precise, and the past patterns cited match known events in the local context. Because the document's role is to synthesize rather than prescribe, tie example claims to specific, verifiable data points where available. Use concrete data: dates of notable campaigns, number of items removed, removal numbers during election windows, language coverage for Filipino, Taglish, and major regional languages, and any measured error rates. When you can, add a short note in parentheses with the latest year or cycle. That keeps the summary up to date without turning it into a live policy memo. If exact numbers are not available, note the data gaps so later work knows what to request from platforms or agencies.

CRITIQUES FROM REFEREES

Analysis. The analysis is solid. The paper identifies the forces at play, from how platforms reward engagement and financial pressures to the trust networks that shape how content spreads within families and peer groups. It also notes limits that cut across agencies and sectors, including overlapping roles, low transparency, and staffing and skills gaps. Given the non-prescriptive purpose, one way to add clarity without endorsing a path is to set a few simple tests that different groups can adopt. For example, if moderation in Filipino, Taglish, and major regional languages makes fewer mistakes (fewer false positives and false negatives), rumors should spread for a shorter time during high-risk periods than in the past. As another example, if Media and Information Literacy (MIL) expands from critical consumption to responsible creation and amplification, then self-reported forwarding of doubtful content in pilot communities should fall over a set period. These tests do not direct policy. They identify what evidence to watch and give room for multiple groups to run comparable trials.

HB 3860 themes. House Bill 3860 appears in the paper as one option among several, and it is reasonable to keep that neutral stance. With that in mind, include common themes from the discussions so the paper reflects what was debated. Definitions should be tighter and more precise, especially for malinformation, to avoid chilling legitimate speech and advocacy. Any intervention should pass a rights test that is clear, legal, necessary, and proportionate, with user education in media and digital literacy and some transparency about algorithms in scope. Many stakeholders are more comfortable with a non-state coalition whose findings stay advisory, paired with quick, independent appeals with access to evidence and clear timelines. Concerns remain about a blanket rule that every platform keep a local office, which may burden smaller or non-commercial services, and about any path that could turn advisories into penalties by other means. There are also concerns about overlapping mandates, the risk of capture if funding is not independent, and privacy risks in sharing data and technical details during investigations. Stating these as themes fits the paper's role and gives future drafters a clearer list of design choices to test.

Two simple steps.

First, for localization, pose questions with evidence that can be verified rather than directives. Which measures would show progress for Filipino and regional languages? Possibilities include coverage rates, reviewer hours on Philippine content, error rates by language, and a yearly third-party audit method. If platforms cannot share numbers, consider stand-ins, public-interest studies, or sampling methods that still shed light.

Second, for Media and Information Literacy, suggest a small, measurable pilot without locking anyone in. A Create Responsibly module in a few barangays and public schools, with ready-made scripts for intervening in family group chats and simple verification steps for short-video posts, could be run by any willing group and measured before and after. These are examples of tests and pilots that others can adopt or adapt. They stay within the paper's role to inform rather than to prescribe.

Overall view. The paper identifies the problem space, recurring limits, and available options, including House Bill 3860, without endorsing a single path. Adding a short section that names the questions to answer, the basic data to gather, and the types of pilots worth trying would make the document more useful to policymakers and practitioners who will take the next steps. This keeps the document within its stated limits while making it more actionable for readers who need to turn ideas into learning-focused pilots.

"

People feel information disorder close to home, often in family group chats late at night. Listening first and mapping the landscape is a good starting point. What people will trust next are small proofs that something works, even if it is just in one city, one semester, or one election window. If the paper directs readers to the evidence that matters and to pilots that test it, the conversation can stay open, diverse, and rights-respecting, while moving from talk to learning and from learning to improvement. That is a fair role for a paper that chose to inform rather than to prescribe.

MARK S. GORRICETA

Atty. Mark S. Gorriceta is the Chairman and Managing Partner of Gorriceta Africa Cauton & Saavedra. He is widely regarded as the Philippines' leading legal authority in fintech, blockchain, digital assets, and emerging technologies, and serves as counsel to major players in the country's digital economy. He has been consistently recognized among the Philippines' Top 100 Lawyers.

We note that several of our previous inputs and recommendations have been incorporated into the latest version of the discussion paper either explicitly or in part, reflecting alignment with our proposed approaches.

Under Part V. Barriers and Gaps in Addressing Information Disorders, the paper acknowledges Filipino social values such as pakikisama and utang na loob, as well as the role of social validation in information sharing. However, these factors are not yet fully explored in terms of their impact on the spread of misinformation. Further elaboration on how these values influence online behaviors would provide deeper insight into culturally grounded drivers of information disorder.

Moreover, the Discussion Paper could more directly address root causes, particularly gaps in critical thinking skills stemming from weaknesses in the education system. While digital literacy and critical thinking are mentioned, a stronger emphasis on developing these skills from an early stage would help Filipinos engage more discerningly with information, thereby reducing susceptibility to misinformation.

Under Part VI. Stakeholder Perspectives, Media Organizations such as newsrooms are cited as central to fact-checking and investigative journalism. However, the rise of digital and alternated media have greatly eroded traditional media organization's viewership/readership base, which further dilutes its effectivity as public fact-checkers. The Discussion Paper could further explore how media organizations have adapted and pivoted to new distribution channels to meet the demands of the social media age.

Under Part VII. Current And Emerging Public-Private Sector Collaboration highlights the need of strengthening Media and Information Literacy (MIL) with cited examples of ongoing collaboration largely driven by the private sector (particularly social media networks such as Meta and TikTok). On the matter of additional resource allocation for MIL campaigns, the Discussion Paper could also explore government-funded or sponsored MIL campaigns/programs such as EU's Action Plan against Disinformation, the U.S. Homeland Security's Disinformation and Governance Board, and Canada's Countering Disinformation: A Guidebook for Public Servants. Government instrumentalities must spearhead MIL campaigns, particularly for matters within their jurisdiction (e.g. DOH for health affairs, DepEd for educational affairs, etc.). structural interventions. Further refinements, particularly broadening case studies, expanding socio-cultural analysis, and including user-facing interventions such as neutral disclaimers would strengthen its comprehensiveness and policy relevance.

Under Part VIII. Institutional And Policy Responses, the paper highlights House Bill 3860, the Digital Media Integrity and Platform Accountability Act of 2025. This bill introduces the Digital Media Standards Coalition (DMSC), a non-governmental, non-partisan, and independently governed body inspired by existing Philippine self-regulation models, such as the Ad Standards Council (ASC) and the Kapisanan ng mga Brodkaster ng Pilipinas (KBP). The paper also acknowledges the role of a self-regulatory organization (SRO) with government participation, providing an oversight mechanism for tech platforms. These inclusions align with prior recommendations and demonstrate an effort to foster cross-sector collaboration.

Further refinements could include promoting open data-sharing standards developed by the SRO in coordination with the National Privacy Commission to ensure compliance with the Philippine Data Privacy Act.

Under Part IX. Policy Conversations and Points of Exploration, the paper discusses initiatives such as investing in localization, enhancing community reporting, and promoting algorithmic transparency in social media platforms. In relation to collaborations with fact-checking organizations, such as #FactsFirstPH, the recommendation regarding the automatic placement of neutral disclaimers or warnings on political, health, and news-related content should be incorporated. This measure remains crucial as a user-facing intervention, providing immediate prompts for critical assessment of content, and complementing structural approaches such as localization, algorithmic transparency, and fact-checking partnerships.

CRITIQUES FROM REFEREES

Regulatory Barriers and Enforcement Challenges

Although the paper references the Cybercrime Prevention Act, SIM Registration Act, and House Bill 3860, the discussion of barriers remains limited. The paper does not sufficiently explore the potential chilling effects of these laws or current enforcement challenges, nor does it provide recommendations for improvement. While the addition of Technical and Platform Limitations (under Part V) is a positive development, the focus remains largely on election-related misinformation. Expanding this discussion to include other sectors, such as public health or national security, would provide a more comprehensive analysis.

The section on platform monetization notes stakeholder concerns about financial incentives for harmful content but could benefit from more detailed suggestions on how platforms might implement robust fact-checking mechanisms and mitigate the spread of misinformation while balancing user engagement.

Focus Beyond Politics

Finally, much of the paper's discussion, including stakeholder perspectives and gaps between policy and practice, remains centered on politics and electoral events. Broadening the scope to include other government-related news and public affairs would enhance the paper's applicability and reduce the perception of political selectivity.

CONCLUSION

Overall, the Discussion Paper demonstrates significant progress in incorporating prior recommendations and addressing systemic aspects of information disorder. The paper appropriately focuses on solutions over punitive measures, emphasizing cross-sector collaboration, digital literacy, and structural interventions. Further refinements, particularly broadening case studies, expanding socio-cultural analysis, and including user-facing interventions such as neutral disclaimers would strengthen its comprehensiveness and policy relevance.

82

ROWENA CAPULONG REYES, PHD

Rowena Capulong Reyes, Ph.D., is the Vice president for Corporate Affairs of Far Eastern University. She is the immediate past Dean of the Institute of Arts in the same university. She also served as manager of the Marketing and Communication of FEU until May 2019.

Dr. Reyes also holds the rank of professor at FEU, handling select communication courses; and gives lectures on leadership, corporate ethics, and crisis communication. Her published articles cover spatial governance, associational life, social media, and communication studies in Scopus-indexed and peer-reviewed journals. She is the lead convenor of FRAMEwork Asia-Pacific Communication Conference and was lead convener of the first FEU Global

An accredited PR counselor, Reyes is a partner and managing director of SEC registered, 5 BC Integrated Marketing Communication, a full-service marketing agency.

Reyes is the country representative for Luzon, Philippines of Asia Media Information and Communication Center (AMIC) and executive committee member of Asia Pacific Communication Alliance (APCA). She is the Ex-Officio of the Philippine Association of Communication Educators (PACE). She held two terms as president from 2017 to 2021 and the position of PRO from 2013 to 2017.

She is a member of the 2024 Metro Manila Film Festival (MMFF) Executive Committee and serves as chairperson of the MMFF Education Committee. She served as chair of the 38th to 40th MMFF board of jurors and sat as a member of the executive and special working committees from 2012 until 2016.

Dr. Reyes was also appointed as dean of the College of Liberal Arts and Sciences of Colegio de San Juan de Letran from 2011 to 2014. She served as the founding director of the Institute of Communication and set up four "specialist" communication programs in 2009.

She has a BA in Communication Arts from the University of Sto. Tomas, MA in Communication Management from the Asian Institute of Journalism and Communication, and a PhD. in Development Studies from the University of Sto. Tomas.

CRITIQUES FROM REFEREES

The discussion paper Disrupting the Disorder: A Multi-Stakeholder Discussion on Information Integrity in the Philippines (Ramos, 2025) arrives at a critical moment in our digital democracy. It captures the intertwined nature of technology, culture, and human behavior in shaping the country's information ecosystem. What distinguishes it is its balanced perspective—recognizing that the Filipino's relationship with information is not purely cognitive but deeply emotional, social, and cultural.

TOWARD AN EXAG-INFORMATION FRAMEWORK

While the study effectively draws from Wardle and Derakhshan's (2017) typology of dis-, mis-, and malinformation, my ongoing research explores what I call the Exaggerated-information Framework or Exag-information, a concept I am developing to describe the exaggeration of a simple truth into an absolute claim, stripped of nuance and presented as hidden knowledge.

Where misinformation misleads unintentionally and disinformation deceives deliberately, Exag-information distorts by amplification. It begins with a factual kernel, something partially true, and inflates it to evoke emotional resonance and moral certainty. This makes it especially persuasive in digital cultures that reward authenticity and emotion over evidence.

I see Exag-information as a missing link in the information disorder spectrum, explaining how plausible content, such as "natural health cures," "anti-mainstream revelations," or "patriotic retellings of history" can outpace outright lies. It blends emotional appeal with epistemic manipulation, creating narratives that feel true even when they mislead.

The Human and Cultural Infrastructure of Disinformation

Ramos (2025) accurately describes disinformation as an ecosystem rather than an event. Platforms, algorithms, and troll farms form the machinery, but the true infrastructure is cultural. Filipino online behavior, shaped by *pakikisama* and relational trust, allows emotional narratives to travel faster than factual corrections. In online exchanges, Filipinos often hesitate to contradict friends or relatives out of courtesy or fear of offending, choosing instead to go with the flow to preserve good relations. Trust is extended not to verified sources but to familiar voices such as classmates, co-workers, influencers, or family members whose posts feel more genuine than institutional statements. In this environment, Exag-Information thrives because it exploits the very traits that define Filipino sociability, humor, loyalty, and a shared sense of discovery. This reinforces why interventions must go beyond fact-checking or regulation; they must engage our communicative values and relational identities. Studies have shown that false or exaggerated content thrives in intimate, trust-based digital spaces like family group chats or fandom communities (Ong & Cabañes, 2018; RIWI, 2022).

From Literacy to Ethical Accountability

One of the paper's most meaningful contributions is its reframing of Media and Information Literacy (MIL) as encompassing not only critical consumption but also responsible creation. Every Filipino online is both an audience and a publisher. As UNESCO (2021) suggests, literacy today must include the ability to create, evaluate, and share ethically.

The proposed Digital Media Standards Coalition (HB 3860) (House of Representatives, 2025) could serve as an important framework, ethical rather than punitive, that is anchored on transparency and inclusion. Yet it should also consider the overt forms of distortion represented by Exag-information. Regulation must evolve from identifying falsehoods to recognizing exaggeration as communicative harm.

CRITIQUES FROM REFEREES

Reclaiming Trust and the Ethics of Imagination

The information crisis is not only about data integrity but about trust. Facts alone cannot rebuild democratic discourse if audiences no longer trust the motives of individuals and institutions presenting them. Exag-Information thrives in this vacuum, offering emotional certainty and a sense of belonging when institutional trust falters. By dramatizing truth and appealing to shared emotions such as pride, humor, or outrage, it gives people the comfort of conviction amid uncertainty, allowing what feels true to replace what is verified.

The challenge for communicators, educators, and policymakers is to make truth not only accurate but meaningful, to restore empathy, context, and imagination in the practice of communication.

> **Meaningful communication begins when facts are placed within human stories, when data is connected to lived experience, and when audiences are invited to see themselves in the message. Imagination enables this process by allowing communicators to envision perspectives beyond their own, transforming truth from information into understanding.**

As I continue to refine the **Exag-Information Framework**, I see it not as an alternative to existing models but as an ethical lens, one that reminds us that even truths, when exaggerated, can become instruments of manipulation.

In this sense, the FEU PPC paper does more than diagnose an ecosystem in crisis; it calls on us to cultivate a new moral ecology of information, one that recognizes how exaggeration distorts empathy and accountability. Countering Exag-Information therefore requires nurturing discernment, relational ethics, and shared responsibility for how truth is told and amplified. That is how we truly begin to disrupt the disorder.

DR. EMMANUEL S. DE DIOS

Dr. Emmanuel S. de Dios is a Professor of Economics at the University of the Philippines School of Economics (UPSE), a position he has held since 1989. He is also the President of Human Development Network (Philippines) since July 2012. He has been an Independent Director of ABS-CBN Corp. since April 23, 2013. He was the Dean of the UPSE from 2007 to 2010 and a trustee of Pulse Asia Philippines, Inc. since 2008.

Dr. de Dios received his AB Economics degree from the Ateneo de Manila University, cum laude in 1978 and his Ph.D. in Economics from UP in 1987.

He pursued post-doctoral studies in the Univeritat Konstanz from 1987 to 1988. He is the author of various books, monographs, articles and reviews in the field of economics.

REFERENCES

REFERENCES

Section II

- We Are Social & Meltwater (2025). *Digital 2025: Philippines*. https://datareportal.com/reports/digital-2025-philippines
- Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking.* Council of Europe. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html
- Ong, J.C., & Cabañes, J.V.A. (2018). *Architects of Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines.* https://scholarworks.umass.edu/items/01d06f54-c7f2-4103-96c1-168a16f9028b
- RIWI (2022). *Information Dystopia and Philippine Democracy.* https://riwi.com/reports
- Media Manipulation Casebook (2022). *Parallel Public Spheres: Influence Operations in the 2022 Philippine Elections.* https://mediamanipulation.org/case-studies/parallel-public-spheres
- Internews (2022). *Political Economy of Covert Influence Operations in the 2022 Philippine Elections.* https://informationsaveslives.org/resources

Section III

- Arugay, A. A., & Quimpo, N. G. (2022). *Democratic Decline and Disinformation in the Philippines.* Journal of Current Southeast Asian Affairs.
- Filloux, F. (2017). "Moore's Law and the Disinformation Explosion." *The Monday Note.* https://mondaynote.com
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* Yale University Press.
- Rappler. (2022). *Disinformation and the 2022 Philippine Elections.* https://rappler.com
- Ressa, M. (2021). *How to Stand Up to a Dictator.* HarperCollins.
- Tufekci, Z. (2015). *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency.* Colorado Technology Law Journal.
- Chua, Y. T. (2025, June 17). *Filipino concern over online disinfo hits record high—Digital News Report 2025.* Tsek.ph. https://www.tsek.ph/filipino-concern-over-online-disinfo-hits-record-high-digital-news-report-2025/
- *Information Dystopia and Philippine Democracy,*, & Khan, R. E. (2023). *Creating safe communication spaces amidst the disinformation quandary.* In *SDG18 Communication for All* (Vol. 1, pp. 199–233). Springer. https://doi.org/10.1007/978-3-031-19142-8_8
- *Cabato, R., & Philippine Center for Investigative Journalism (PCIJ). (2025, August 7). Five ways the Duterte influence machine is deceiving Filipinos. PCIJ. https://pcij.org/2025/08/07/five-ways-the-duterte-influence-machine-is-deceiving-filipinos*
- *We Are Social & Meltwater. (2025). Digital 2025: The Philippines. (DataReportal).*

REFERENCES

Section IV

- Bradshaw, S., & Howard, P. N. (2019). *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation.* Oxford Internet Institute.
- Marwick, A., & Lewis, R. (2017). *Media Manipulation and Disinformation Online.* Data & Society Research Institute.
- Vosoughi, S., Roy, D., & Aral, S. (2018). *The spread of true and false news online.* Science, 359(6380), 1146-1151.
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). *Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era.* Journal of Applied Research in Memory and Cognition, 6(4), 353–369.

Section V

- Columbia Global Freedom of Expression Initiative. (n.d.). *Disini v. The Secretary of Justice (G.R. No. 203335).* https://globalfreedomofexpression.columbia.edu/cases/disini-v-the-secretary-of-justice/
- Republic Act No. 10175. (2012). *Cybercrime Prevention Act of 2012.* https://www.officialgazette.gov.ph/2012/09/12/republic-act-no-10175/
- Republic Act No. 11934. (2022). *SIM Registration Act of 2022.* https://www.officialgazette.gov.ph/2022/10/10/republic-act-no-11934/
- East Asia Forum. (2024, April 11). *Tempering the Philippines' AI disinformation storm.* https://eastasiaforum.org/2024/10/16/tempering-the-philippines-ai-disinformation-storm/
- Democracy.Net.PH (2025). *#RespectMyPost: The Use of Information Operations to Undermine Philippine Elections.* Democracy.Net.PH report.

Section VI

- Cabato, R. (2024). *Navigating malign influence campaigns in the Philippines: A guide for journalists and newsrooms.* Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/navigating-malign-influence-campaigns-philippines-guide-journalists-and-newsrooms
- Khan, R.E., & Chua, Y.T. (2023). *Countering disinformation-tools and initiatives in the Philippines: Tsek.ph case study.* https://www.mediasupport.org/wp-content/uploads/2023/06/Countering-disinformation-tools-and-initiatives-in-the-Philippines-May2023.pdf
- Ong, J. C., Tapsell, R., & Curato, N. (2019). *Tracking digital disinformation in the 2019 Philippine midterm election.* New Mandala. https://www.newmandala.org/wp-content/uploads/2019/08/Digital-Disinformation-2019-Midterms.pdf

REFERENCES

Section VII

- Commission on Elections (COMELEC). (2022). *COMELEC partnership with platforms during the 2022 national elections.* Referenced in Meta and TikTok newsroom reports.
- GMA Network. (2021, November 24). *Google PH launches program promoting online safety among children.* GMA Lifestyle. https://www.gmanetwork.com/lifestyle/home/82985/google-ph-launches-program-promoting-online-safety-among-children/story
- Interaksyon. (2022, January 26). *#FactsFirstPH to Filipinos: The battle for facts needs you.* Interaksyon. https://interaksyon.philstar.com/trends-spotlights/2022/01/26/209259/factsfirstph-to-filipinos-the-battle-for-facts-needs-you/
- Meedan. (2022, January 26). *The battle for facts in the Philippines: Launching #FactsFirstPH.* Meedan. https://meedan.com/post/the-battle-for-facts-in-the-philippines-launching-factsfirstph-f
- Meta. (2022, January 19). *Protecting election integrity in the Philippines.* Meta Newsroom. https://about.fb.com/news/2022/01/protecting-election-integrity-in-the-philippines/
- Meta. (2023). *Human Rights Impact Assessment (HRIA) – Philippines.* Meta Transparency Center. https://transparency.fb.com/
- TikTok Newsroom. (2025, April 18). *TikTok partners with creators to encourage Filipinos to #ThinkTwice.* TikTok Newsroom. https://newsroom.tiktok.com/fil-ph/tiktok-partners-with-creators-to-encourage-filipinos-to-thinktwice
- UNESCO. (2021). *Media and information literacy curriculum for teachers.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000377069
- National Audiovisual Institute (KAVI). (2022). *Finland's National Media Literacy Policy.* KAVI. https://kavi.fi/en/media-education/

Section VIII

- House of Representatives (2025). *House Bill 3860: Digital Media Integrity and Platform Accountability Act.* Explanatory Note.
- Center for Media Freedom and Responsibility (CMFR). (2022). "Why Anti-Fake News Laws are Problematic." https://cmfr-phil.org
- European Commission. (2023). *Digital Services Act.* https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package
- Ministry of Law Singapore. (2019). *Protection from Online Falsehoods and Manipulation Act (POFMA).* https://www.pofmaoffice.gov.sg
- Australian Communications and Media Authority. (2021). *Disinformation and Misinformation Code.* https://www.acma.gov.au

REFERENCES

Section XII

- Bradshaw, S., & Howard, P. N. (2019). The global disinformation order: 2019 global inventory of organised social media manipulation. Oxford Internet Institute.
- Chua, Y. T. (2025, June 17). Filipino concern over online disinformation hits record high—Digital News Report 2025. Tsek.ph. https://www.tsek.ph
- Filloux, F. (2017). Moore's law and the disinformation explosion. The Monday Note. https://mondaynote.com
- House of Representatives. (2025). House Bill 3860: Digital Media Integrity and Platform Accountability Act.
- Internews. (2022). Political economy of covert influence operations in the 2022 Philippine elections. https://informationsaveslives.org/resources
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. Journal of Applied Research in Memory and Cognition, 6(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008
- Ong, J. C., & Cabañes, J. V. A. (2018). Architects of networked disinformation: Behind the scenes of troll accounts and fake news production in the Philippines. University of Massachusetts Amherst.
- Ramos, A. K. (2025). Disrupting the disorder: A multi-stakeholder discussion on information integrity in the Philippines. Far Eastern University Public Policy Center.
- RIWI. (2022). Information dystopia and Philippine democracy. https://riwi.com/reports
- UNESCO. (2021). Media and information literacy curriculum for teachers. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000377069
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
- Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe.

DISRUPTING THE DISORDER: A Multi-Stakeholder Discussion on Information Integrity in the Philippines

Discussion Paper on the Spread of Dis-, Mis-, and Malinformation in Social Media Platforms

By **Alma Karina "Kankan" Ramos**
With research support from Clara Quiambao

October, 2025