Taiwan Fellowship Report

Fake Factory: Disinformation and Learning Machines in Taiwan

Florian Schneider, Leiden University

Abstract: This report explores how Artificial Intelligence (AI) interacts with controversial information flows (CIF) in Taiwan. Taiwan's unique geopolitical status makes it an epicentre for disinformation and foreign influence campaigns. The study is based on extensive interviews with stakeholders across fact-checking, policy-making, and digital literacy sectors, utilizing a discourse-theoretical approach. Findings reveal that CIF are aggravated by entrenched polarisation and the manipulative design of corporate social media, facilitating a decoupling of realities within society. While generative technologies enable Foreign Information Manipulation and Interference (FIMI) through cheap content production and AI bias, the core manipulative patterns remain consistent. Taiwan counters these challenges using a 'whole-of-society' approach, focusing on coalition building, supporting civil society fact-checkers, prioritizing interpersonal solutions, and advancing AI literacy to build democratic resilience. While this approach has generally been effective, it nevertheless struggles with several challenges, most notable political partisanship, the attention economy created by multinational platform corporations, and the lack of funding for civil society initiatives. The report concludes with important lessons, most notably the need to avoid hyperbole and 'othering' while favouring a calm, deliberate approach that emphasises personal interactions and empathy.[1]

[1] This study was made possible with the support of a half-year Taiwan Fellowship, awarded by the Centre for Chinese Studies at the National Central Library, financed by the Ministry of Foreign Affairs of the Republic of China, and hosted by the Political Science department of National Taiwan University. The Leiden University Institute for Area Studies arranged the teaching relief to make this fieldwork possible, and I am particularly grateful to Vincent Chang for financing that teaching relief and to Vincent Brussee for stepping in as such a competent replacement. Most importantly, my gratitude goes to the interviewees who were so generous with their time, and to the many people who spoke to me off the record. I have done my best to assure their voices, and their interests, are reflected fairly in this report, but the interpretations of what they told me, along with any errors or omissions, are of course entirely my own.

Table of Contents

1. Introduction: Struggles over Truth in Digital Societies

In 2018, Typhoon Jebi hits the Kansai region in Japan. The subsequent flooding strands thousands at Kansai International Airport (KIA), including several Taiwanese travellers. On social media, questionable information about the disaster makes the rounds. One particularly viral allegation: Taipei's head diplomat in Osaka, Su Chii-cherng, is supposedly not helping his stranded compatriots. The vitriol about Su's alleged incompetence grows exponentially, culminating in reports that the General Consulate of the People's Republic of China (PRC) had succeeded where the Taipei representation had failed: to rescue the beleaguered Taiwanese. On 15 September, Su hangs himself in his home. In his suicide note, he explains how the news had pushed him over the brink. Only later does the Japanese government establish that the stories had been fabricated.[2]

In late 2023, Taiwanese social media is ablaze with concerns over the safety of Taiwanese women. The alleged cause of the problem: Indian immigrants. False information had spread on social media that the Taiwanese government was allowing 100,000 Indian men to enter the island. The pro-China news outlet *China Times* picked up the story. Further stereotyping and racist remarks followed online, claiming that Indian immigrants were turning Taiwan into a 'sexual-assault island'. The authorities responded by pointing out that the false online statements were 'strikingly similar in phrasing, suggesting automized use of fake accounts or paid commenters'. The timing was also suspicious: the xenophobic speculations came only two months before Taiwan's 2024 general election, raising concerns about foreign information manipulation.[3]

We live in 'computational knowledge societies' (Berry 2011, 176), which offer access to abundant, networked streams of information. And yet, these societies struggle to deliver on earlier promises of improved participation and democratic decision-making (see Shirky 2008). Earlier hopes that advanced information and communication technologies (ICT) would create a 'see-for-yourself' culture (Benkler 2006, 218), in which citizens can reliably assess political claims and act on them, have been confounded by reality. Instead, ICT have aggravated, accelerated, and obscured the challenges that democratic societies face today. They regularly filter information in ways that reduce serendipity (Pariser 2012), lead users into the walled gardens of social media echo-chambers (Diaz Ruiz & Nilsson 2023, Terren & Borge 2021), amplify negative emotions like anger and anxiety (Doroshenko & Tu 2023, Sampson et al. 2018), polarize political attitudes (Qureshi & Bhatt 2024), and fuel populist scepticism of authoritative information sources like mainstream news media or government (Strömbäck et al. 2022).

Contested information flows and their risks

One particular concern is the accelerated spread of controversial and contested information. In what follows, I will refer to this phenomenon as controversial information flows (CIF). This includes the

[2] For news reports and analyses of this case, see e.g. Huang (2018), Lin (2018), and Becerra (2022).
[3] The case was reported by Thomson (2023) and Feng (2024); for a commentary, see Pathak (2023).

unintentional sharing of false information (misinformation), but it also incorporates the strategic use of falsehoods (disinformation) and of misleading facts (malinformation), sometimes by individuals, sometimes by highly organised groups such as political and religious organisations, corporate actors, governments, and militaries (see Appendix 1 for a glossary of definitions).

Granted, concerns over the validity of information are not new to the much-evoked information age (see Chirovici 2015 and Rid 2020). However, our hyper-networked societies have introduced an unprecedented degree of uncertainty, leading to the sense that we now live in an era of 'post truth' in which false information and conspiracy theories run rampant (MacKenzie & Bhatt 2020, Rainie et al. 2017, Tandoc et al. 2018, and Waisbord 2018). As societies around the world try to come to grips with CIF, bad-faith actors in business and politics are cashing in on the confusion.[4] The stakes are high. The scope of such efforts creates an unreliable information environment, undermining the ability to create legitimate, meaningful political discourse. As innovations in ICT supercharge the information flows of our societies, they also threaten to spell a future of distrust and aggression, eroding the very foundations of participatory politics.

Networked societies in Asia, long at the forefront of technological innovation, are well acquainted with the risks. This is particularly the case in Taiwan. While Taiwan's political status may seem unique, its democratic politics, liberal media system, and wide-spread adoption of social media invite comparisons with societies elsewhere. What is more, and as the examples that opened this report illustrate, Taiwan has been an epicentre of CIF, so much so that one former regulator I spoke to called it a 'digital wildfire' (I13). Indeed, dealing with CIF has become part of everyday life for most Taiwanese. Taiwan's public spheres are today inundated with questionable information about anything from mundane consumer topics and cultural preferences to less innocent issues affecting health and safety (Lin & Wu 2019). Throughout, politics loom large: studies have repeatedly shown how information manipulation has impacted elections (Doublethink Lab 2024a, 2024b, Lau 2024, Wang 2020) or shaped important political debates, for instance on same-sex marriage (Rich et al. 2018), on questions of how to handle foreign social media platforms (AI Lab 2024), or on the relationship between Taiwan and foreign powers (Kuo & Wu 2022, Yu 2023).

Stakeholders on both sides of the proverbial parliamentary aisle are eminently aware of the risks to democratic politics (see my own previous research in Schneider 2019). Concerns over information 'warfare' are widespread (Lin & Wu 2019) among officials, non-governmental organizations, journalists, analysts, and activists. And while these concerns have been a feature of Taiwanese politics for many years,

[4] The number of recent studies has mushroomed to an extent that makes it unwieldy to reference the field in its entirety. For a collection of works, see the contributions in Arcos et al. (2023). Noteworthy stand-alone studies include, for the USA, Allcott & Gentzkow (2017), Lazer et al. (2018), and Spohr (2017, 155-156); for the UK, see Bastos & Mercea (2019) and Spohr (2017, 156-157); for continental Europe, see Fletcher et al. (2018) and Schäfer & Schadauer (2019). Waisbord & Amado (2017) have studied Latin American cases, and my own work has looked at East Asian examples (Schneider 2023).

they take on a novel dimension as new ICT reshape how political communication works. Reflecting risks and anxieties elsewhere, Taiwan's public spheres now contend with the power of hyper-accelerated manipulation campaigns that utilize machine learning and automated decision-making systems, in short: artificial intelligence (AI). Indeed, as Taiwan's Doublethink Lab (2024a) found, in a study of the 2024 elections in Taiwan, 'video content produced by AI and deepfakes have become more popular during this election, including virtual anchors with AI-generated narrations to read scripts, or clip collections with AI-generated dubbing'.[5] These advances in generative technologies lower the barriers of entry for media content creation, speed up the production process of such content, and increase the risk that our media ecologies will become overwhelmed by the resulting 'AI slop'.[6]

As disconcerting as these new challenges may be, Taiwan's civil society has a strong tradition of building digital and in-real-life systems for addressing them, and for strengthening the resilience of democratic participation (Ai 2019, Ho 2020, Lee 2020, Thornton 2025). In the past, this has included highly innovative, bespoke counter strategies that try to generate empathy and use self-deprecating humour and memes (Schneider 2019). As the risks of information manipulation and uncertainty change in the face of new AI technologies, and as those risks grow across the world, democratic societies can learn much from the Taiwanese experience.

A note on methodology

While this study is informed by an extensive literature review and dozens of informal conversations with Taiwanese citizens from all walks of life, the core of the analysis relies on qualitative stakeholder interviews, conducted in the spring of 2025. In the interest of safety and fairness, I have anonymized all participants and refer to them only by shorthand (i.e. interviewee 1 is I01, etc). This setup follows the ethics and data management guidelines of Leiden University and has been approved by the university's ethics and data authorities.

The interviews involved a total of 16 practitioners and experts on the struggle against disinformation in Taiwan. The interviewees included civil society actors, specifically in the realms of fact-checking, literacy training, and civic tech design, but also policymakers, administrators, and analysts. Some interviewees wore several proverbial 'hats' (for instance academics who had also served in government, or activists who volunteered for civic organisations while working in business or tech). The people I spoke to also had diverse political allegiances across Taiwan's 'green/blue' party spectrum as well as a mix of gender

[5] In the original: '本次選舉期間 AI 與深偽製作的影片內容變得普及，包括虛擬主播配上 AI 生成旁白來讀稿，或是片段合輯加上 AI 生成的旁白配音。順應著 AI 技術的發達，影片的製作成本及時間降低，製作影片內容來進行資訊操作逐漸成為主流手法。' All translations in this report are my own.

[6] For a tongue-in-check but nevertheless harrowing deep-dive into AI slop, I can recommend John Oliver's report, in which he presciently sums up the problem: 'AI Slop can be somewhat lucrative for its creators, massively lucrative for the platforms that use it to drive engagement, and worryingly corrosive to the general concept of objective reality' (Last Week Tonight 2025, 25:39).

and age backgrounds. Together, these interviews provide a crosscut of experiences from Taiwan's information society.

The interviews were semi-structured and open-ended. They were conducted in either English or Mandarin, lasting between one and two hours, and they were recorded with the interviewees' permission for later analysis. To confirm the accuracy of the interviews, the relevant research notes were later shared with the respective interviewees.

This analysis of the interviews was meant to establish and verify important facts about CIF in Taiwan, but it was also interested in the discourse that emerged from the conversations. In discourse theory, the issue at stake is how people talk about certain issues. The idea is that such talk reveals underlying structures of reasoning. As humans, we reason our way through our world, and to do this we draw from the available ideas, concepts, relations, and explanations that are available to us culturally. Much of this happens unconsciously, but it still affects what happens next: how we interpret our problems, what decisions we make, how we intervene in our societies. Taking a discourse approach to the topic of disinformation and AI in Taiwan makes it possible to explore the rationales from which stakeholders draw to make sense of this topic, and to extrapolate from this the potential consequences of these ways of understanding. As we will see, there is much to be learned from the way people in Taiwan reason about this profoundly important topic.

In short, this report asks how AI is changing the nature of CIF in Taiwan. It analyses how stakeholders deal with digital struggles over political truths, and what those struggles reveal about best-practice strategies for immunizing democratic societies against AI-accelerated information risks. In what follows, I first present the challenges that stakeholders identified in our conversations about CIF. The subsequent chapter explores counterstrategies, specifically regulatory and socio-technological efforts. The report concludes with key lessons.

2. Taiwan's Digital Wildfires

How do relevant stakeholders in Taiwan conceptualize the vulnerabilities of their society? This section analyses the risks that they discussed, and how they situated those risks within geopolitical concerns, specifically the issue of foreign information and manipulation interference (FIMI). It also explores what relevance stakeholders attributed to the boons and banes of machine learning.

2.1 Common Vulnerabilities

A central question I put to stakeholders in Taiwan was how they diagnosed the problem of contested information. Did they recognize patterns? What did they see as underlying causes of the problem? Acknowledging the complexity of the issue, the respondents stressed the need for nuance. They repeatedly pointed out how the problem connects in murky but important ways with wider societal challenges. It is not a stand-alone issue. A recurring theme across the interviews was that the 'digital

wildfires' of false or manipulative information find fertile kindling in two wider concerns that should be recognizable to readers elsewhere as well: polarisation and the widespread use of corporate social media.

Popular and elite polarisation

With its two large political camps, and with media organisations that reproduce a sense of diametrically opposed interests between these camps, polarisation has been a long-standing issue in Taiwanese politics.[7] When it came to CIF, my interviewees consistently pointed to this issue as an aggravating factor. One fact-checker even described it as Taiwan's greatest vulnerability (I01), pointing to deep-seated but often misleading prejudices between the political camps, especially with regards to their views on China. The interviewee argued that, when surveyed, most supporters of the Democratic Progressive Party (DPP) vastly overestimated how strongly supporters of the opposing Kuomintang (KMT) were in favour of unification. Similarly, KMT supporters overestimated how strongly voters in the green camp supported independence. In reality, so the interviewee argued, most people in Taiwan supported the status quo,[8] even if this was not always clear to voters.

Interviewees widely shared the sense that polarisation was a fundamental risk factor (e.g. by I01, I02, I03, I04, I05, I06, I07, I08, I09, I10, I11, I12, and I15). One interviewee called polarisation 'a huge issue' (I07), explaining that it fed off of the increasing social cleavages in Taiwan and empowered antagonists to drive a wedge into Taiwanese society. Others similarly stressed that 'this kind of polarisation targets people who are suffering from lack of opportunities, for example lack of education, higher salary expectations' (I02). As one analyst summed up the challenge: 'I gotta be honest with you, I still feel the problem we need to solve is polarisation and trust, and people being able to talk to each other' (I12).

Polarisation among voters and the general public was certainly a concern, but so was *elite* polarisation, specifically the unwillingness of politicians to compromise with their political rivals (I10). As one interviewee put it (I15): 'Political polarisation has poisoned our society on almost every issue, and on almost every public policy. It has gotten so far that, if the TPP [Taiwan People's Party] or KMT want to push for a certain policy, the DPP will say no, even if that's a long-time policy of their own. And if the DPP wants to push for some kind of policy, the KMT will say no. It's not about principles. The only consistent principle is "I'm not going to buy your argument or any of your proposals".'

This polarisation has fuelled populist scepticism of the fact checkers themselves. Numerous interviewees had stories to share of how fact-check organisations like the Taiwan FactCheck Centre, MyGoPen, or CoFacts, as well as the think tank Doublethink Lab, had come under fire from highly partisan media users

[7] For an introduction to Taiwan's politics, see Fell (2012) and Riggers (2014) For discussions of Taiwan's media and its complicated relation with Taiwanese identity issues, see Hsu (2014, ch.3).
[8] Taiwanese identity and political allegiance are subjects of extensive survey research. For examples, see Qi & Lin (2021) and Yang et al. (2021). Huang & Kuo (2022), Hsiao & Yu (2020), and Wang (2019) have specifically analysed polarisation in Taiwan, and for a survey of how polarisation and perceptions of disinformation relate, see Hsu (2024).

(for an overview of organisations, see Appendix 2). This frequently included accusations that the fact-checkers and analysts were secretly beholden to one of Taiwan's political factions. Some of these organisations now received almost daily aggressive messages disparaging their work and accusing them of bias (I08). Interviewees explained that such accusations came from both pan-green (DPP-aligned) and pan-blue (KMT-aligned) supporters, as well as the relatively young 'white' camp of TPP supporters, depending on whose position had most recently been debunked (I02, I08). However, allegations about green influences were particularly common. They generally took the form of online criticism and toxic behaviour on social media, but in at least one case right-wing activists went as far as protesting at the offices of a fact-checking organisation in the hopes of seeing the populist politician Han Kuo-yu exonerated from accusations that he had been spreading false information (I10).[9]

Interviewees frequently offered examples that illustrated how right-wing populism was particularly guilty of CIF; that said, they also echoed scholarly assessments that the issue was not limited to any particular political worldview or creed.[10] One interviewee explicitly criticized green politicians for leveraging disinformation to promote their agenda, while in the same move delegitimating their critics as the supposed source of disinformation. The interviewee particularly criticised the use of hyperbole to heighten anxieties and pre-empt any meaningful political dialogue across party lines: 'They [DPP politicians] are saying, "if we can't get rid of the blue and white parties in congress... as quickly as possible… then our kids will become Xinjiang ren". That's the propaganda. Our kids will become Xinjiang ren and there will be organ harvesting' (I15).[11] The interviewee went on to describe their own political confrontations with specific DPP administrators: 'I always ask: "why don't you give me the evidence". But it's not about evidence. It's about propaganda.' Another critic of green politics (I16) was similarly concerned: 'They [the DPP] treat digital power and the internet and technology as a way to do communication, to broadcast their ideology (…) a not-so-good word is propaganda – ideology broadcasting.' While acknowledging that right-wing populists often spread lies, they continued to ouline how that observation had itself become political ammunition in the hands of partisan authorities: 'the government, they have the power to designate who is fraudulent. And according to their political thinking... maybe a lot of their supporters think the KMT party is a scam. The KMT is a fraud. So some of [the KMT's] local representatives, of the city council, their accounts have been banned, their accounts are being pulled down.'

[9] For a study of populism in Taiwan, and specifically of Han Kuo-yu, see Krumbein (2023).
[10] See, for instance, Chang et al. (2021).
[11] Literally 'people in Xinjiang', a reference primarily to the various Muslim ethnic groups that live in the PRC's western autonomous region of the same name. The present report is not a study of PRC policy in Xinjiang, but it is worth acknowledging how this hot-button topic has fuelled CIF on both sides of the issue, with PRC officials and supporters reframing documented practices of re-education and cultural erasure as morally justified de-extremification initiatives while their more radical critics opportunistically compare these politics to the Holocaust; the gratuitous trope of organ harvesting has been particularly popular with Falun-Gong-aligned propaganda channels. For a scholarly critique of PRC policy in that region, see Tobin (2021). For a pro-Chinese defense, see Slawotsky (2021).

In general, then, Taiwan's information environment is struggling with a potent combination of uncertainty and deep-routed antagonisms that create fertile ground for suspicion and anxiety, themselves sources of further unethical communication practices. The result is a vicious cycle of, on the one hand, lies and obfuscations, and on the other hand dismissal of any truth claims *as* lies and obfuscations. This leads to a cascading series of problems, in which CIF 1) reinforce entrenched positions (regardless of their truthfulness); 2) empower political camps to delegitimate each other's truth claims by labelling them as 'disinformation' or 'fake news'; 3) hamstring potentially valid criticism of such claims e.g. by fact-checkers; and consequently 4) fuel cynicism about political and journalistic processes in general.[12]

In such an environment, it becomes nearly impossible to return to any semblance of objective reality. I have called the effect 'reality decoupling' (Schneider 2023); Taiwanese actors have coined the term 'artificial multiverse' (Doublethink Lab 2024a) to describe this outcome. The general sense is that 'disinformation is everywhere' (I10), that 'there's too much information, and we don't have a way to determine which [pieces of information] to trust' (I09), and that 'it's very difficult to get back to the facts. There's no facts, the facts have become too complicated (...). It has become a belief and trust issue now' (I07). Across the board, and regardless of political allegiances, my conversation partners pointed to a common factor that aggravated these issues of trust deficits and information overloads: the wide-spread reliance on corporate social media.

Social media and the 'temperature' of debate

Throughout my conversations with Taiwenese stakeholders, ICT stood front-and-centre. As one interviewee put it, 'even before the internet, we had rumours, but because of the internet, spreading those rumours is much easier and quicker and more effective than before. So we are concerned, because – it's kind of terrible – it just keeps evolving and it never stops' (I10). Another interviewee was explicit that 'the fundamental problem is social media. It's kind of sad to say, because it's not going away, but it's the fundamental problem' (I12).

This widely-shared sentiment marks a change in attitudes over the past decade, from generally optimistic views of ICT to more ambiguous, and at times unequivocally bleak, assessments today. Reflecting on Taiwan's vibrant history of democracy and social movements (see Appendix 4), one interviewee familiar with Taiwan's social change activism and its progressive hacker community discussed this shift (I01). They outlined how, starting in the mid-2010s, many civically-minded activists were powerfully deploying social and digital technologies for their activist purposes, most notably the g0v ('gov-zero') hacker community, and most prominently during the 2014 Sunflower Movement. However, that movement's 'very beautiful vision' of ICT-enabled participatory politics quickly soured. Sentiments arguably tipped into disappointment around 2018. At that time, KMT politician Han Kuo-yu emerged onto Taiwan's political scene as a major populist force, celebrated on social media by many blue-collar conservatives as a

[12] Fabian Schäfer (2023, in German) has examined political cynicism in the context of digital information flows.

'national saviour'. As the interviewee described it, Han's popularity, along with reactionary voter behaviour in the 2018 local elections and referenda, had been informed by 'huge disinformation'.[13] Many reformers were dismayed by what they viewed as having 'lost a huge battle', and disillusionment set in as democratic backsliding and online toxic behaviours drove home that 'social media was not designed for democracy or deliberation; it was designed for popularity'.

The challenge for many who today try to intervene in social media information flows, such as fact-checkers and civic organisations, is that corporate social media are working against them on several levels. Three issues stood out in my conversations with Taiwanese professionals: emotional manipulation, surveillance, and what scholars call 'homophily': the human tendency to associate primarily with similar others.

As the interviewees stressed, social media algorithms are designed to generate virality by 'upping the temperature' (I02) on social and political debates. Malicious actors, in turn, are adapt at manipulating those mechanics to fan negative emotions like fear, hatred, and anxiety. The second issue was that today's app ecology allows exploitative data-gathering and surveillance activities that can be turned against users, in ways that are largely opaque. As one digital literacy advocate put it (I14): 'They [suspicious apps] can even look at you. They can do whatever they want. They know your schedule. They can see all your pictures, they can see your videos, where you go, things like that. (...) So when we show this to people, they start to realise they are selling themselves to someone else for a free app. So, in Taiwan, we have a saying that the most expensive thing is "free".' The risk, so the interviewee, was that such questionable data-gathering practices empowered malicious actors in commerce and politics to take advantage of hapless social media users.

Finally, practitioners and analysts in Taiwan voiced concerns that commercial social technologies exposed users to 'filter bubbles' and pushed them into 'echo chambers' – a view that aligned with academic assessments that social media facilitate 'homophily' (Pariser 2012, Ruiz & Nilsson 2023, Terren & Borge 2021). Much like the algorithmic curation of high-temperature emotions, these walled gardens lent themselves both to unwitting and maliciously designed CIF: 'We're all in our own echo chambers, we all have our cognitive vulnerabilities, and social media... it's this massive supercharged weapon for exploiting our cognitive vulnerabilities', said one interviewee (I12). Ironically, those who avoided certain platforms for fear of manipulation could not escape this problem: one interviewee explicitly discussed how China-sceptics who boycotted TikTok 'don't see what's going on' in that ecosystem (I01).

[13] The most high-profile issue was the failed referendum to legalize gay marriage, which the ruling DPP then ignored, further fuelling populist perceptions that the green camp's avowed goal to promote participation was mere window-dressing for power politics. Supporters of the referendum in turn stressed that the public debate had been hijacked by foreign information manipulation, forcing the DPP's hand. For a discussion of this referendum, see Rich et al. (2018).

Asked which demographics they saw as most vulnerable to the adverse effects of social media, the general sense was that all ages, educational levels, and socio-economic groups were susceptible to digital CIF. While interviewees agreed that elderly people were particularly challenged by advances in digital technology, they also acknowledged that senior citizens were often aware of their limitations and sought help, e.g. by asking younger relatives for a reality-check or by enrolling in digital literacy workshops. A group that was in many ways more at risk was young media users, as their over-confidence with digital technologies frequently proved deceptive. Recounting several horrific cases of cyberbullying, harmful pranks, and deadly social media challenges, one interviewee summed up the problem (I14): 'younger people, that's more difficult [than with older generations]. Because they are the technology generation, so they think they know everything. But no, actually no. They know how to use it [their device], but they don't understand about safety of information and technology at all.'

In that context, several interviewees voiced concerns about natively Chinese apps like RedNote (Xiaohongshu) or TikTok's domestic version Douyin.[14] While the general sense was that most Taiwanese users were capable of using these platforms with sufficient scepticism and care, young adults often lacked the experience and media literacy to fully appreciate the biases inherent in such social media platforms, so the argument. 'They feel that Douyin is much more interesting, the content is more interesting than on the international version', said one interviewee (I10). 'With TikTok, at least you can say that it's more transparent. Douyin is a black box. You never know what's inside, you just know that the Chinese government is watching it and manipulating it. So this is a worry for us.'

I will return to questions of media literacy below. For now, it is worth concluding that Taiwanese stakeholders viewed social media as a potent source of division, especially when interactions on these platforms amplified polarising views of existing social cleavages. Many saw the root problem of CIF in the active manipulation of information within an environment that had already been eroded by continuous information overload. One analyst (I07) explained that 'there's too much information, you don't know what's correct, what's not correct. It's very difficult to make any judgement.' When asked about foreign interference into Taiwan's political debates, they argued that, in such a context, 'if I were to one day launch a real attack, then you'll just collapse on your own.'

2.2 Geopolitics and Foreign Information Manipulation and Interference

A central concern for many interviewees was then also the manipulation of Taiwan's information and media environment at the hands of bad-faith actors. It is worth stressing, however, that especially practitioners in the fact-check community where either hesitant to speculate about the actors behind CIF, or they were explicit that the origins of CIF were not relevant to their work. As we will see below, some even saw attributions of origin as counter-productive, since allegations as to who might be manipulating

[14] For research into TikTok, see Cervi et al. (2023) on political contents and Primig et al. (2023) on disinformation.

debates risked stoking the flames of polarisation further, potentially causing the fact-check efforts to backfire.

These hesitations contrasted with the assessment of analysts, especially those who worked on national security and cyber-affairs, who stressed the risks posed by foreign information and manipulation interference (FIMI). They generally painted a nuanced picture of the layered CIF ecosystem in Taiwan and the actors that manipulated it. As one regulator put it, 'for almost a decade now, I've observed that Taiwan has been the country most heavily attacked by foreign forces. Of all the countries under attack, Taiwan is number one' (I13).

Examples of such interference ran the gamut of commercial to political initiatives. Malicious actors included influencers and content farms that cashed in on the viral potential of emotional content. They also included politically-motivated actors, for example the anti-Chinese religious order Falun Gong (with the *Epoch Times* and *New Tang Dynasty TV* media outlets being frequent sources of disinformation), alt-right American MAGA supporters and neo-conservative evangelists (which were often involved in populist agitation against liberal values like LGBTQ+ rights), and foreign states like the Russian Federation. In fact, the case of Indian immigrants that opened this report was rapidly spread by Russian outlets, and closely followed the Russian FIMI playbook, specifically the pattern of trying to use online accounts to mobilize local communities for offline protest. This has led analysts to suspect Russian state involvement in this case (I12). However, despite this diversity of actors, the most commonly-named forces behind information manipulation that the interviewees named were the Chinese Communist Party (CCP) and its People's Liberation Army (PLA).

Chinese actors in Taiwan's CIF

PRC-originated FIMI has been well-documented, and yet neither the exact scope nor the effectiveness of such interventions into Taiwan's information environment are straight-forward.[15] While the CCP considers control of communication a crucial part of its politics, and a moral obligation, its strategic communication approach is complex: it encompasses both domestic and international actions, in both overt or covert ways (see Brady 2008, Zhang 2011). Adding to this complexity is the fact that a large range of Chinese actors engage in discussions in or about Taiwan, with often diverging interests and strategies. These actors include CCP organisations, the state, state-owned news enterprises, and the military,[16] both at the central level, and across China's regional and local bureaucracies. Official actors

[15] For studies of PRC FIMI in Taiwan, see Doublethink Lab (2021, 2024a, 2024b), DSET (2024), and Lee (2024). For assessments, see Chang et al. (2021), Huang (2023), Hung & Hung (2020), Rauchfleisch et al. (2023), and Tseng & Chen (2020). Bui (2025) has covered recent findings and debates.

[16] Specifically on the PLA's cybersecurity approach, see Ventre (2014, 55-80). The PLA's Strategic Support Force (PLASSF, see Kania & Costello 2021) takes a 'three warfares' (三战) approach to the international information environment, combining public opinion warfare, psychological warfare, and lawfare to secure PRC interest. Its Base 311 in Fujian Province along with its commercial arm, the China Huayi Broadcasting Corporation (Beauchamp-Mustafaga & Drun 2021), are responsible for implementing these approaches against Taiwan.

often unapologetically engage in explicit public relations and propaganda activities, including on social media platforms that are blocked in China itself. This involves fake followers to exaggerate the impact of these official outlets: 'the People's Daily on Facebook has 84 million followers', explained one analyst (I04). 'The New York Times has something like 20 million, and Chinese people can't even use Facebook. How is that possible? Like, Xinhua also has 99 million, and that's been going on for a long time. So right from the start, they knew "we have to be present on these websites and then we also have to look like we have a lot of followers".'

In that sense, official media outlets are part of an overt attempt to 'tell China's stories well', but they are also flanked, and bolstered, by covert operations, including botnets, sock-puppets, and bought accounts. An important part of these covert efforts to shape public opinion is fake grassroot activity, or 'astroturfing', named after the fake lawn on sports fields.[17] Such activities are part of what the analytics company Graphika has dubbed 'spamouflage': an attempt to seed political influence within a wider barrage of seemingly innocuous human-interest content.[18] 'These are accounts with a mix of photos, just like random scenery or beautiful individuals', explained one interviewee (I04). 'And then suddenly there'll be like a bunch of propaganda.' Another analyst gave examples of seemingly mundane accounts that posted about 'daily English' vocabulary or 'daily jokes', but that then attempted to smuggle political manipulations into their content: 'They [PRC propagandists] hire a lot of influencers or buy existing fan pages and pretend that these are Taiwanese citizens, and then they turn that towards political content... they pretend they are normal citizens; they are just like you!' (I01).

In some instances, such efforts could be linked directly to Chinese officials, for instance when suspicious activities coincided with the office hours of military and propaganda personnel: 'so, 9-to-5, because they're either operated by government officials or, like, content farms. So these people... it's their job' (I04). In other cases, the connection to the Party and state was more tenuous. PRC efforts were not as unified as observers sometimes assume, explained one interviewee (I05): 'every unit or institute has their own KPI [key performance indicator] to fulfil'. This led to inconsistent activities and narratives. What is more, officials also recruited social media influencers into their communicative approach. It can be lucrative business to create social-media content that appeals to Chinese nationalist worldviews. Some such creators may well be paid directly by the CCP; however, others likely benefit from Chinese support in less tangible ways, for instance by receiving privileged access to interesting sites around China, or by seeing their social media presence on Chinese platforms boosted algorithmically by Chinese companies.

[17] For a seminal analysis, see Han (2018).
[18] The original discussion of this practice appeared in Nimmo et al. (2019).

The result is what one interviewer called 'entrepreneurial propaganda'.[19] These influencers 'can make money from spreading CCP propaganda', as the interviewee explained (I12):

> There's an economy and a structure that has been set up, at the hand of the CCP, for building, maintaining, and spreading – through, firstly, subsidies, so these direct payments – certain things that get said; and through opening up donations from the PRC to people for saying these things. So you get WeChat pay logos that sit there, and then people go "click". So you can make money off of the *Xiaofenghong* [the 'Little Pinks', i.e. PRC nationalists] by saying this stuff to Taiwanese.

In some instances, influencers may well act out of their own calculations within this propaganda economy; in other cases, they may inadvertently have been roped into it. However, where the activities of influencers demonstrably trail the propaganda efforts of Chinese state media, analysts are more confident to speak of a concerted official effort to shift public opinion: 'we can say, for example, that Chinese state media were the first to present story X, and then these people [social media influencers] picked up story X within two minutes, for example. That's not going to be natural' (I12).

Beyond direct collaborations with influencers, the Chinese authorities also rely indirectly on nationalist citizens who take their opinions and feelings online, or carry them abroad in their travels. The activities of such ostensible supporters do not always fit PRC interests, for instance when activities turn toxic, threaten to turn against the party itself, or accidentally empower critics of the PRC.[20] However, the party seems willing to accept such tensions, as private citizens overall serve an important role in representing China: '...they [the CPP] see people-to-people exchanges as an integral part of [their public diplomacy]. They think that if they can persuade or educate their people about the righteousness of Chinese ideas, then those people will do their own personal diplomacy for China', argued one analyst (I05).

As these discussions illustrate, Chinese communication efforts in and about Taiwan remain complex, and the stakeholders I spoke to then also stressed that it would be misleading to think of these efforts solely in terms of outright falsehoods: 'The content can be disinformation, but it can also be unverifiable things like conspiracy theories. It can also be true. You can also have, on the behavioural side, a network of fake accounts pushing true content... So disinformation is only a subset of the problem' (I12). This, however, is not to say that the actors behind FIMI in Taiwan were not pursuing specific goals, using identifiable tactics.

FIMI strategies and tactics

As the interviewees described it, the reality of FIMI in Taiwan had shifted over the years. PRC actors in particular were increasingly more willing to lean into the messiness of CIF. Rather than dogmatically pushing for unification, they had realized that such blatant efforts could back-fire. 'In the early days,'

[19] Influencer propaganda has inspired a growing field of scholarship, specifically on China, though it is not limited to the PRC. See e.g. Xu & Qu (2025), Xu & Schneider (2025), and Xu & Yang (2025).
[20] For an example, see Wasserstrom (2025). I have discussed similar dynamics in Schneider (2018).

explained one interviewee (I10), 'they tried to get you to be pro-China. Now, they don't care whether you are pro-China or not. They want you to be anti-US, anti-Japan, so that you will be isolated and they can create chaos. That's what they want. So the whole strategy, over the last decade, we've seen that changing'.

Analysts in Taiwan then also linked foreign interference efforts back to the vulnerabilities of Taiwan's polarised political environment. To seed discontent, foreign actors try to 'find controversial topics in society and then amplify them, usually the side that is against the DPP. They almost always do that. Very occasionally, they will create their own conspiracy theories. But in most cases, the KMT criticizes something that the DPP does, and then there will be all these fake accounts that we believe are connected to the PRC that are amplifying these narratives' (I04). In a similar vein, another analyst argued that 'they try to increase political struggle, to help create, to foster political struggle in Taiwan. So disinformation serves that purpose' (I05). A third explained that the goal was to break the connections between Taiwanese people 'by seeding suspicion and trying to separate people (...), to try and make people have fights on the internet, to disconnect people' (I02).

Overall, these FIMI campaigns were adept at using the steps of disinformation that Graphika analyst Ben Nimmo called the '4Ds': to dismiss, distort, distract, and dismay (see Appendix 3). Specifically for Taiwan, recurring narrative patterns about domestic politics were that 1) Taiwan's social problems were getting out of hand while 2) Taiwan's leaders were corrupt and 3) Taiwan's democracy was consequently unable to address these issues. On international politics, the pattern was to argue that: 1) Taiwan was internationally isolated, 2) the US and its military could not be relied on to assist Taiwan, 3) Taiwan's own military was weak and incompetent, and that in contrast 4) the PLA was strong and competent. Ultimately, the goal then seemed to be to shake Taiwanese citizens' self-confidence, and their trust in domestic institutions, while anchoring the idea that 'there is no way that Taiwan can resist the threat or invasion from China' (I06) and that 'standing up against the PRC doesn't really make sense... there's no point in resisting' (I04).

It is again difficult to assess to what extent such narratives are convincing; they may very well be compelling only to those who are already convinced. In terms of electoral behaviour, past experiences suggest that even the suspicion of PRC influence can backfire by tainting a political actor's chances at success.[21] A greater risk, so the analysts I spoke to argued, was that confusion over facts might temporarily hinder important decision-making, for instance during natural disasters or a potential military engagement. They gave examples of precision deepfakes of Volodymyr Zelenskyy during Russia's

[21] The 2016 presidential elections, for instance, stood in the shadow of the highly publicized silencing of Taiwanese K-pop star Chou Tzu-yu for having been caught on camera holding an ROC flag; while the case probably did not affect the overall outcome of that election, it may have pushed a number of reticent voters to the polls in support of the DPP's candidate Tsai Ying-wen. See BBC (2016) and Kuhn (2016) for reporting on those elections, and Sullivan & Smyth (2016) for a scholarly assessment.

invasion of Ukraine (I06), and they discussed how false representations of leaders and public figures could 'seed confusion and doubt' with 'very dangerous consequences' (I11). As one interviewee put it, by the time the false information has been debunked, 'the damage may already be done... maybe 15 minutes later, they find out that it's fake news, but that 15-minute gap matters' (I05).

However, there is one other risk, and it only emerged indirectly from my conversation with Taiwanese stakeholders. Somewhat ironically, it may well be that the relatable discussions about having to fend off outside forces ultimately empowers precisely such forces. For instance, my respondents frequently spoke of China or the CCP as 'the enemy', using the language of war to make sense of the manipulation efforts they were documenting. This is certainly understandable, and it may in some regards be useful: Taiwan is demonstrably confronted with military-level adversarial tactics that conceptualize information flows as part of a 'cognitive warfare' strategy. Response frameworks to such strategies then likewise often use the language of military conflict to capture the adversarial nature of the phenomenon, for example conceptualising disinformation as engagements between 'red' attackers and 'blue' defenders, or adopting terminology such as 'kill chain'.[22] Using these frameworks is arguably an effective way to reverse-engineers what the perpetrators behind disinformation are doing, which can in turn be useful for anticipating and countering such tactics.

On the other hand, using the language of warfare poses two risks: the first is that it may tacitly perpetuate the premise that our information societies are structured antagonistically, even when no militaries are involved. This impression benefits the agents behind disinformation and FIMI. Rather than dialling back anxieties to create spaces for calm deliberations and responses, it injects furtherer anxieties into a situation that is already inundated with emotion. In the Taiwanese case, this is especially evident when discussions turn to China, but it applies to other societies as well. Implying that democratic societies are under siege by hostile forces unwittingly opens the situation up to even greater opportunities for manipulation. The second risk is that the language of warfare may antagonise media users who genuinely hold contested ideas. This in turn may push them to align themselves with actual adversaries, when dialogue and empathy would have been better equipped at compelling them to rejoin a reasoned discussion.

To be clear: the analysts I spoke to were unanimous of the view that a critical assessment of, and response to, PRC-induced FIMI should not be understood as a criticism of pro-Chinese sentiments in general. They strongly felt that genuinely-held political views were part of a functioning democratic free-speech environment. And yet, analysts were often caught in a dilemma: on the one hand, they felt the need to inform the public of clear and present threats, on the other hand they had to articulate their threat-

[22] Arguably the most extensive approach of this kind is the Disinformation Analysis Risk Management (DISARM) framework (see DISARM Foundation, n/d), itself built on a cybersecurity approach called Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK, see MITRE 2025). The term 'kill chain' stems from military doctrine and was ported to cybersecurity contexts by US defence contractor Lockheed Martin (see the White Paper by Hutchins et al. 2011).

perceptions in a way that did not feed back into the already worrying degrees of polarisation and anxiety in Taiwan. As we will see, there is an important lesson to be learned from this Catch-22. But first, it is worth exploring how CIF are interacting with advances in ICT, specifically with AI technologies.

2.3. AI and Generative Technologies

A particular focus of my conversations with Taiwanese stakeholders was the way in which technological advances, specifically in the field of AI, were impacting both CIF and strategies to counter CIF. The general picture was that AI played an ambiguous role. On the one hand, AI-generated content demonstrably contributed to CIF (Chen 2024, DSET 2024, Hung et al. 2024, Wang 2024). On the other hand, interviewees told me that the underlying patterns of CIF (and FIMI) were not changing fundamentally. Several interviewees pointed out, for instance, that – at least at the time of our conversations – influence operations from China were not particularly sophisticated in their generative AI use. 'They're not yet very mature, allowing relatively easy detection', said one analyst (I05). Anther explained (I04): 'The people who are implementing these influence operations, they studied languages and so on. They're not necessarily very technically savvy. Or these content farms: they're just people writing things online, but they don't know how to set up this whole system.' A third interviewee (I09) agreed: 'The technology is there, but probably it is not necessary, because, well, the low tech is still useful. So why bother with high tech?'

Analysts and practitioners also stressed that even when generative AI had been used to create manipulative contents, this did not substantially alter what they saw as effective counter strategies: 'the way we mitigate generative images and generative videos', explained one fact-checker (I03), 'is very similar to the way we try to mitigate cheep fakes'. Another interviewee (I12) similarly maintained that generative AI had not changed the core challenge: 'it's annoying, but yeah, we're still facing the same problems we were facing before generative AI, and we should just be solving those problems instead of worrying too much about AI.'

Simultaneously, there was a sense that AI-development was taking place at a rapid pace, and that 'things are just moving too fast' (I10) to arrive at a clear verdict about AI's role in Taiwan's CIF ecology. While AI might not have shifted the needle on CIF in Taiwan yet, three concerns nevertheless stood out as risk factors. These were: 1) the malicious use of AI for advanced surveillance, simulation, and prediction (AI analytics), 2) the speed and scalability of production / dissemination, and the resulting high-volume but low-quality contents (AI slop), and 3) the deceptiveness inherent in using AI chatbots for everyday tasks (AI bias).

AI analytics: misuse and overreach

As mentioned above, interviewees were concerned about the surveillance capacities of 'free' consumer applications, and that worry was compounded by AI's potential to enhance the surveillance and analytical

capacities of bad-faith actors. One analyst (I04) pointed out how actors who were not restricted by ethical or legal restrictions could potentially insert AI-powered computational processes into every aspect of their surveillance and control efforts. While use of AI for such purposes in China was still limited,[23] the interviewee voiced concerns about the ambition to create a technocratic system of governance on the mainland:

> That's when it starts to get really scary, because then you can just monitor tens of thousands of people at the same time and look at all their posts, analysing them according to certain requirements, and then send emails to the secret police to have them arrested almost automatically. And they're not using Chinese AI to make that [system]. They're using ChatGPT at literally every step of the process, so: developing the code, making the business pitch that they're giving to their superiors...

The concern then also went beyond the implications for PRC cyber-governance, as the ability to scrape and analyse transnational social media networks at speed and scale offered unprecedented analytical opportunities. One analysist (I09) pointed to the possibilities of digital simulation, specifically the creation of so-called 'AI twins' of entire societies. 'You can create a digital twin of a real person and, using this digital twin, you can manipulate or simulate things that happen to this real person. And now we can create a digital twin society, right? Not just the individual, but the population.' The interviewee cited Xiamen University as a PRC institution that was using data from the social media app RedNote (*Xiaohongshu*) to create a 'digital twin of Taiwanese society', potentially enabling the government to then 'test or to see how Taiwanese respond to an emergency, a war, or whatever' – a prospect the interviewee found 'kind of scary'.

In that sense, AI was seen by many of the people I spoke to as a suite of analytical and generative tools that bad-faith actors were either already experimenting with or that they were on the brink of integrating into their efforts, potentially with wide-ranging consequences.

AI slop: cheap content production

One specific AI use-case was the generation of fake text, audio, and video. While the verisimilitude of AI-generated media contents was often quite low, leading to images and videos that are 'so weird or crazy that nobody would think that they're real' (I04), the cheap and easy access to automated media production capabilities was a concern to many of the professionals I spoke to. With generative AI, 'it's easier for people to launch all this fraud and these disinformation campaigns than in the past. So that's quite worrying' (I07).

A particular issue was the ability of generative AI to create variations of the same piece of information, making it harder to track false information online and establish whether the creators were genuine people

[23] For comprehensive and nuanced overviews of the ambitions and realities of China's cyber-governance, see the contributions in Creemers et al. (2025) and Yu & Creemers (2025).

or not. 'At some point', said one analyst (I04), 'it's very difficult to tell that this is a bot.' A fact-checker (I10) explained the problem further:

> Before AI, you'd probably see posts that are all the same text, with just a few words changed. And then these posts would be posted everywhere, but it's the same content. But with AI, they're able to produce different versions, and very quickly. And it's hard for us to catch, because the wording has been changed, or the meaning has been changed, or they even target different groups, choosing different paths, like targeting the pet lovers, targeting the food lovers, so that's become quite difficult for us.

The challenge extended from text to multi-media contents as well. One interviewee (I01) discussed how producing e.g. 100 videos had previously required a huge team, and many months of editing, but that the same could now be accomplished by 'just one guy' in a short amount of time, allowing such contents to be 'mass-produced'. A fact-checker (I08) described a similar challenge: 'It's really cheap right now. So you can publish hundreds of fake videos just with a few clicks, you know? So it's very easy for everyone to publish some videos'. They went on to outline the scope of the problem: 'There's so so many videos, we just can't debunk them one by one. We can debunk one or two, but then there's 99 more on YouTube. We just can't kill them.'

At the time of writing, most examples that interviewees cited fell into the category of AI slop: cheaply generated and easily identifiable fake contents. However, the sense that the generative AI models were improving, and that imitations of images and voices might in the future be generated with higher accuracy and fidelity, was clearly a concern.[24] That concern was exacerbated by the ways in which manipulative practices interacted with the use of generative AI chatbots in everyday life.

AI bias: deceptive chatbots

The most obvious concern about AI use among my interviewees was that Taiwanese citizens might turn to mainland-Chinese services for everyday questions, but without fully realizing the inner workings and 'biases created by chatbots like DeepSeek' (I10). The risks of relying on DeepSeek had led the Taiwanese government to prohibit its use among officials (I16), but its adoption by general users remained a concern. 'China has DeepSeek, and it's really powerful', reflected one interviewee (I08). 'So many people will use that, because they think it's really good at Chinese... But it's more pro-China. You'll ask it questions about Xi Jinping, you'll get Xi Jinping answers. But you can't prohibit people from using it. You have to tell them that the results behind it are from China.'

[24] That this concern is not entirely unfounded is confirmed by studies of deepfake technologies, especially the practice of improving fake images by pinning two neural networks against each other as one generates content from the training set (generator) and the other detects flaws in the outcome (discriminator), leading to an iterative process of continuous improvement (Walorska 2020). This approach, known as 'Generative Adversarial Networks', or GANs, is used both in the creation and the detection of deepfakes, leading to a virtual 'arms race' (Bontridder & Poullet 2021) between actors who spread such contents and those trying to debunk them (Wang et al. 2025).

Beyond these obvious examples of mainland-Chinese services, the professionals I interviewed were keenly aware that popular genAI chatbots generally suffered from biases. They frequently pointed out how the services built in Silicon Valley had been created using biased training data. They worried that the average user in Taiwan would not be fully aware of the implications: 'Once the technology becomes more mature, then we're afraid it can have a greater impact on Taiwanese society, because those large language models, their source in Mandarin is about 98.6% from mainland China. Only about 1.4% originates from Taiwan', explained one analyst (I05). This meant that important debates or social concepts were coloured by mainland worldviews, sometimes inadvertently, e.g. when American systems like ChatGPT reproduce PRC cultural perspectives. One policymaker (I16) explained:

> ...if you ask ChatGPT any questions about democracy in Chinese, you will get an ideology about democracy from mainland China. Because the two Chinese characters, they're written the same way in simplified and traditional... I had a discussion with some of the big companies, about ChatGPT and Gemini, and they told me 'it's true, if you want to know the true value of democracy, you have to use [the word] "democracy" in English' But how do our students know?

In addition, interviewees pointed out that the US-models were susceptible to manipulations by malicious actors, for instance when such actors created false websites to 'poison' AI algorithms with faulty data. 'All of this is going to end up in the training data', said one analyst while discussing Russian FIMI operations (I04). 'At some point you're going to have these really niche issues about, I don't know, certain Russian dissidents, and then when you ask a question [on ChatGPT], all it knows is this Russian narrative because it read it on some shady website.'

Biases in big data are of course a general problem that has been hotly debated for decades, e.g. in the context of how search engines like Google or the algorithms of social media platforms reproduced certain worldviews, often as a side-effect of their design and business models.[25] With AI, however, these effects are magnified by two innovations inherent to genAI: the first is the fact that chatbots are not merely *curators* like search engines that guide users to external content, they are *creators* of the content itself. Secondly, the way tech corporations like Anthropic, Google, Microsoft, or OpenAI have deliberately designed their human-machine interfaces as conversational bots creates a false sense of reliability and accountability. Evidence has been accruing that, for many users of genAI chatbots, it is not clear that 1) they are *not directly* querying an LLM but rather an intermediary interface that amends their prompts to generate the results in compelling human language, and that 2) the LLM is merely a predictive machine, *not* an agent with human qualities like a mind.[26] Scholars like Simone Natale (2021, 2) then also go as far as to call AI an inherently 'deceitful medium':

[25] Seminal early discussions of search engine biases are Halavais (2009) and König & Rasch (2014). For more recent studies on algorithmic biases, see Kitchin (2017) and the contributions in Jarke et al. (2024).
[26] The human tendency to anthropomorphize algorithms has been well-established since the 1960s. However, the degree to which AI chatbots inspire such projection of human qualities, and the ways corporate actors intentionally

> Even in the absence of deliberate misrepresentation, AI technologies entail forms of deception that are perhaps less evident and straightforward but deeply impact societies. We should regard deception not just as a possible way to employ AI but as a constitutive element of these technologies. Deception is as central to AI's functioning as the circuits, software, and data that make it run.

I quote Natale here in full because the professionals I interviewed, while not necessarily making quite such a radical case, or using the language of 'deceitful media', were clearly uneasy about the interaction between deliberate CIF manipulations and the built-in potential of AI chatbots to mislead users.

Several interviewees then also advocated the construction of a home-grown, sovereign AI system for Taiwan, 'to make sure that simplified Chinese and mainland Chinese ideology do not override Taiwanese values and traditional Chinese ways of thinking in AI machines' (I16). The idea was to create a modest LLM trained on Mandarin sources in the traditional character script, using materials from Taiwanese. To this end, the National Science and Technology Council (NSTC) under the Executive Yuan started work on the 'Trustworthy AI Dialogue Engine' (TAIDE) in 2022 (see also Robertson 2024). This LLM, built on top of open-source models by Meta (Llama) and Google (Gemma), uses licensed, open-source data to create a chatbot.[27] The goal was fairly modest, as joining the 'AI race' was not realistic for smaller actors, explained the experts I spoke to. Instead, as one interviewee put it, 'the purpose of developing our own system is not to compete with the big companies, but to at least have something to fall back on... otherwise you will purely rely on others' (I09). However, the project seemed to have stalled in 2024, with interviewees criticising that very little information had been released to the public (I11) and worrying that TAIDE was 'in sleep mode' (I15).

One issue was that the data-gathering approach that fuelled TAIDE was slow and expensive. A proposal to scrape data widely for fair use, and to only exclude materials if copyright holders objected, was rejected due to concerns over legal and political fallout, as one interviewee explained (I15). Instead, the project required individual negotiations with each copyright holder, which had bogged down progress. Funding was the other issue: 'It's a resource problem' (I08). One opposition politician put it bluntly (I16):

> The TAIDE project, they need money, they need data, they need talent. We have some in Taiwan, but they don't have money, they don't have enough talent to work for them, and: they don't have data. So you can see today... Lama 4 has four trillion parameters, 4 trillion! And TAIDE, the version our government released, has 8 billion. 8 billion versus 4 trillion. And DeepSeek R1, they have 673 billion.

A major concern, as the interviewees described it, was that government priorities had shifted. One interviewee outlined how the newly appointed head of the NSTC was focusing on semiconductors rather

rely on this tendency to sell their genAI products, have been a matter of grave social concern; for discussions, see e.g. Claypool (2023), Daley & Johnston (2025), and – in Dutch – Grosfeld (2025).

[27] The project and its milestones are explained by NSTC (2025). Demos are available at https://huggingface.co/taide (last accessed 30 November 2025).

than sovereign AI (I15); another pointed to political partisanship as again getting in the way of supporting the project sufficiently (I09):

> It's doable. I think it's doable. But still, you need to pass the legislation, and that's the difficult part... passing normal legislation in our legislative yuan is kind of impossible, nowadays... it's doable if we have a reasonable mind. We will find the right solution to fix the problem. But I have no control over the legislative yuan.

Ultimately, many interviewees were sceptical that a small-scale model like this, created under conditions of tight funding and shifting government priorities, could fulfil the role that the developers had envisioned for it. As much as the model was not meant to compete with commercial service, those services remained more attractive than what the government could offer. As one interviewee put it: 'if you're not ahead... you lose' (I01).

3. Taiwan's Reponses to Contested Information Flows

As one of my colleagues has stressed in her analysis of disinformation responses (Thornton 2025), the Taiwanese attempt to grapple with CIF is best described as a 'whole-of-society' approach.[28] This approach involves, to quote the UN's definition, 'civil society as well as the public and private sectors in the joint pursuit of common solutions to complex problems', embracing 'both formal and informal institutions in seeking a generalized agreement across society about policy goals and the means to achieve them' (Kjellen & Wong 2023, 173). While my informants did not directly reference this terminology, the strategies they described were generally in the spirit of such a comprehensive, multi-stakeholder approach. The interviewees frequently stressed how the Taiwanese approach relied on a broad societal coalition of actors that analysed, debunked, and informed the public about CIF in Taiwan, and how those actors leveraged a combination of social and technological strategies to achieve their ends. Here, I will summarize how the interviewees discussed policy and government action as well as a societal activities.

3.1 State Efforts: Between a Rock and a Hard Place

Throughout our conversations, stakeholders familiar with Taiwan's regulatory environment stressed that a response to CIF, and especially to FIMI, needed to involve legislative and government action, with one legislator summing up the general sentiment that 'cyber security is national security' (I16). Indeed, Taiwan has a sophisticated government framework, designed by several offices of the Executive Yuan in the late 2010s.[29] However, the interviewees painted a mixed picture of how successful the ROC's approach had been.

[28] The 'whole-of-society' idea was probably first articulated at a crisis management workshop at the Asia-Pacific Center for Security Studies in 2010 (APCSS 2010). It has since been adopted specifically to counter 'hybrid threats' like disinformation, by organisations ranging from the US military (Maddox et al. 2021) to the EU (Wigell et al. 2021).
[29] Kao (2021) has covered the government approach in detail.

Policy and regulation

When it came to legislation, interviewees pointed to the ROC's anti-fraud act of 2024 as a small but important legal step for regulating harmful CIF. The act criminalizes financial, telecommunications, and digital manipulations.[30] In combination with Taiwan's existing criminal code, this means that creating and releasing manipulative contents such as deepfakes is punishable by up to seven years in prison (I09). Others argued that, with the Lai administration designating China a 'rival', potential PRC activities in Taiwan's information sphere fell under national security legislation, making the use of conspiracy theories and other malicious CIF to intentionally promote PRC interests punishable as a criminal offense (I11). Especially in cases where Taiwanese actors had been paid to spread harmful information, such cases could potentially fall under the jurisdiction of national security regulations (I15).

Not everyone I spoke to was positive about how the government was using this legislative framework. One opposition legislator (I16) harshly criticized the anti-fraud act, arguing that it allowed the government to shut down social media accounts without concrete evidence of maleficence and without due process. He felt that government agencies like the Ministry of Health and Welfare, the Ministry of Digital Affairs, or the Financial Service Council applied an excessively broad list of 'daily keywords we use in anything' to have social media accounts closed down on platforms like Meta, and without giving the users an avenue for potential redress. The interviewee described their own experience as a public figure on social media: 'I just get up one morning, and I find my account disabled, and it says it's disabled because I'm suspected of fraud and scamming... I can do nothing. There's no telephone number, there's no email, there's no way to report back, there's no way to protect my rights.'

While not all interviewees were this negative about regulatory practices in Taiwan, several voiced unease when it came to regulating user content. They explained that Taiwan's history of martial law made concerns over free speech regulation particularly sensitive (I15), but also that liberal democratic societies generally struggled with the tension between safeguarding core values like freedom of expression while having to mitigate the risks that such freedoms generate: 'Democratic societies... they're stuck between a rock and a hard place', as one interviewee put it (I13). Several stakeholders then also argued that a sensible regulatory approach should focus less on content and more explicitly on behaviours and the actors involved,[31] specifically when it came to establishing ill intent or foreign financial support (I15). As one analyst put it (I12), 'if you want to spread some pro-China or pro-Russian content, that's alright, but if you're using an army of inauthentic social media accounts to help you do so, then we'll step in and try to respond'.

[30] The full act is available as ROC (2024).
[31] The idea that approaches should distinguish actors, behaviours, and contents lies at the heart of the 'Disinformation ABC' formulated by François (2019).

Those who supported stronger government action on cyber-issues like hate-speech, bullying, and online manipulation were generally disheartened by the inability of Taiwanese politicians to arrive at a balanced bipartisan approach. One interviewee (I13) described the attempt by the National Communications Commission (NCC) to create a regulatory framework modelled on the EU's Digital Services Act,[32] which would have covered malicious information and pornography. That bill fell through in 2022 because, as the interviewee described it, the KMT accused the DPP of using it to enforce 'internet martial law'. Another interviewee recounted how the draft bill generated such an outcry from civil society and opposition politicians about the dangers of censorship that the government effectively 'gave up legislating the communication environment' entirely (I01). One legal expert voiced regret over this outcome, arguing that the NCC had been 'too scared to face the critics'. Overall, there was again a sense that Taiwan's polarised political environment, and the hyperbolic rhetoric it engendered, had gotten in the way of debating sensible legislation. The interviewees voiced little confidence that the legislative deadlock could be resolved, especially not when matters touched on free speech (I01) or when they affected Taiwan's influential tech industry (I11).

Despite such criticisms, several interviewees argued that Taiwanese regulators had proven adept at finding pragmatic solutions to CIF challenges. They noted, for instance, that ROC officials often relied on informal contacts or closed-door negotiations with large multi-national cooperations such as Meta or Line to establish a public-private consensus on how to tackle CIF (I05, I11, I13). Also, while there was a sense that Taiwan's regulatory approach was insufficient, the existing regulations were seen as at least a small step in a potentially fruitful direction. For instance, as one regulator argued (I13), the anti-fraud act opened the door for future fine-tuning, as it assured that platform providers had to designate a legal representative responsible for fraud-cases in Taiwan. The interviewee believed that this provision could potentially be used to hold platforms accountable for online disinformation as well.

Government communication and transparency

In addition to the broader policy landscape that enables the government to intervene in malicious information practicers, government agencies in Taiwan also implement a proactive communication approach that directly addresses rumours and false information. As one regulator outlined (I13), each ministry was responsible for addressing false information within its area of jurisdiction, and each had its own 'clarification mechanisms' (澄清的机制) to allow a rapid response to CIF, especially in crisis situations. This included press briefings and spokespersons at all levels of administration, as well as extensive training in communication methods for public-facing staff. In the ministries that dealt for instance with economic affairs, agriculture, and health and welfare, this approached had led to substantial experience over time, as these organisations were 'bombarded by disinformation on a daily bases' and had

[32] The full Digital Service Act and Digital Market Act are available as EU (2022).

consequently been forced to 'become increasingly adept at understanding how to combat false information'.

Government agencies are required to respond to problematic online discourses within 24 hours, and interviewees pointed out that this rapid response had indeed proven effective, for instance during the pandemic (I13) or environmental disasters (I06). The ministries accelerated their response rate by having templates at their disposal on how to counter false information; this included using memes and self-deprecating humour as vehicles to de-escalate polarised debates and inspire laughter rather than anger or fear.[33] In August 2025, the government vowed to further refine this approach by making its responses even more accessible to a broad public. Its so-called '222 policy' (222 法則) stipulates that departments should address misinformation in no more than 200 characters, using no more than 2 images, or with no more than 2 minutes of video footage (Yeh 2025).[34]

Finally, government communication can only be effective if the source of information is considered trustworthy, and a crucial path for assuring and safeguarding such trust was 'transparency, transparency, transparency' (I16). As one regulator put it (I13): 'If we want to combat fake news, government transparency and information provision are also extremely important. Very important.' To this end, interviewees stressed that the task of debunking fraudulent information should not be left to the government. To be trustworthy, it had to involve independent third parties in civil society, and the public itself. As one interviewee put it, a fact-checking organisation 'has to be a third party, it has to be neutral' (I06), and this was the ethos that informed much of the work in Taiwan's lively ecosystem of think tanks and civic organisations (see Appendix 2).

3.2 Socio-Technological Efforts:

When I spoke to civil society actors about that work, they described what amounted to a 'socio-technological' approach, as academics would put it, so strategies that combine technological systems and human interactions.[35] This included a range of high-tech, low-tech, and no-tech solutions. Analysts and fact-checkers in particular relied on digital technologies as they searched, catalogued, and analysed CIF. Those that were involved in outreach used a combination of digital networks and in-real-life interactions

[33] I have discussed this approach in an earlier report (Schneider 2019). The 'humour over rumour' idea was established by Audrey Tang (see e.g. Lee & Blanchard 2022) and is also discussed by Thornton (2025).
[34] This policy builds on an earlier '222 principle' that prompted government officials to debunk false information using no more than 20 characters in the title, no more than 200 characters in the main text, and no more than 2 images as illustration (see Kuo 2021, 8).
[35] The study of socio-technological interactions lies at the heart of Science and Technology Studies (STS), a field too broad to cite here in full. Useful introductions to the field come from Sismondo (2010), Felt (2017), and Kleinman & Moore (2014), though arguably one of the most accessible and thought-provoking entry-points is the seminal article by Kranzberg (1995). In the Taiwanese context, Thornton (2025) has taken an STS approach to analyse how Taiwan's whole-of-society response to disinformation makes use of socio-technical systems, specifically open-source governance and algorithmic mechanics.

to engage their audiences, especially in the crucial sector of digital media and AI literacy, where interpersonal connections were seen as key.

Tech-empowered fact-checking and analytics

Much of the civil society effort to address CIF was pragmatically built on top of existing services and infrastructures. A highly visible example is the way that organisations like Cofacts, MyGoPen, and the Taiwan FactCheck Center have integrated their crowd sourcing approach into social media platforms, most notably the Japan-based messaging app LINE. The popularity of the app in Taiwan, and its ability to foster communication within small social networks, was a major concern to these organisations, with one fact-checker seeing it as 'the frontline, on LINE' (I08). While the NGOs are also highly active on platforms by Meta, specifically Facebook and Threads, they each maintain dedicated apps that interfaced with LINE (I01, I02, I03, I08, I10): 'These are systems built to reach people who are in need of different opinions', explained one fact-checker (I03). Cofacts, for instance, has continuously refined its system to effectively index information, improve searchability, and add a chatbot for easy user interaction.[36] Similarly, MyGoPen maintains a chatbot that enables its ca. 500,000 LINE users to send in suspected false information (text, image, or video) to receive clarification on its legitimacy.

The hope is that such systems will desensitize users to the relevance of fact-checking, and that they will help users reach out even as they encounter suspicious CIF in other contexts, e.g. while using Google (I03). In fact, despite anxieties over information overload in Taiwan's media environment, some of my interviewees were carefully optimistic that the past decade of civil-society efforts had started to bear fruit: 'So now if we receive disinformation we won't just buy it. We will for example Google it, try to identify the authenticity of the information. I think many people will do that when they think the information may not be true' (I05).

Yet, to arrive at such a baseline understanding was delicate work. It meant being mindful of polarising topics, which was one reason why the fact-checkers I spoke to generally avoided attributing actors and motivations to the misinformation they debunked. Many were keenly aware that criticizing e.g. pro-Chinese messaging required a great deal of nuance, care, and tact. As one factchecker put it, 'it's not like all pro-Chinese content is Chinese propaganda – it's not like that. That's not what [we want] to tell people' (I02). This meant focusing on the kind of content that users interact with in their daily lives. 'We do have a lot of requests around spam, or around investment and policy and social welfare issues, because people really rely on those policies (...) to maintain their daily lives', the interviewee (I02) went on. By not getting sidetracked with questions about who was behind specific statements, the fact-checkers avoided the pitfalls of protracted partisan debates. 'It's so important to reach out to different kinds of people who

[36] For further discussion of Cofacts, see Thornton (2025, 19-20), who also explicitly discusses the value of normalising fact-checking efforts through such services.

don't share a similar belief with you,' they explained. Instead, they aimed to 'make baby steps, but useful steps' that empowered users and educated volunteer collaborators (ibid.):

> We teach them a framework to differentiate factual statements from personal opinions. That way, our collaborators can work out which parts are factual and also which are emotional, so those personal interpretations that go beyond the factual statements. It's very beneficial when we try to compose a counter narrative or try to compose a fact checking reply. So that is the thing that really works.

LLMs already play a role in this process. Chatbots such as the one deployed by Cofacts are built on AI, and fact-checkers now also use AI services such as Perplexity, ChatGPT, or Hive Moderation to assist them in tracing and verifying specific information, speed up the detection of fake footage, and geolocate specific images (I04). While some noted that AI detection tools were not yet sufficiently accurate (I10), many felt that they at least augmented their established analytical toolbox. 'It's like an assistant', said one fact-checker (I08), 'it's really helpful'.

Especially in their communications, fact-checkers had started to use generative AI. An example of this is how AI-generated transcripts allow fact-checking collaborators to compare contents, more easily find previous cases, and potentially reuse effective responses to misleading contents. One fact-checker envisioned a future step in which LLMs would help give writing prompts, speed up the writing process, assist with language editing, and provide directions or ideas on how to phrase an effective response to misleading information: 'So actually, what I try to do now, and in the future, is to use large language model, those services, to lower the entry barrier for fact-checking collaborators to create quality, useful fact-checking points' (I02).

Collaboration and outreach: the relevance of meaningful connection

As important as ICT-empowered analysis and outreach were, it was also clear that none of the interviewees felt it was sufficient. 'We cannot just do autopsies', explained one analyst (I12), 'we cannot just spray and pray'. They went on to stress that 'we will need to work out how to still make connections happen, how to reconnect again'.

This theme, in fact, came up time and time again: a robust effort to tackle CIF required meaningful personal interactions. Such interactions were already a feature within the fact-checking community itself, with the various organisations frequently coordinating their efforts and offering each other assistance and advice.[37] These personal networks also extended internationally, with analysts collaborating in the region and beyond to compare notes about transnational information campaigns, creating a community of support and care. 'We know that this information, it will go from language to language, from country to country', explained one fact-checker (I10). 'It may originally be English, but then it'll come to Taiwan,

[37] In fact, I found it challenging to keep my interviews anonymous, as my interlocutors often knew each other and were in frequent contact.

and someone will translate it to Chinese... so we have specialists, a network of (..) fact-checkers, and we all work together to fight this.'

But the challenge, as many described it, was to also assure personal connections across society to improve Taiwan's resilience against malicious CIF. This was particularly crucial in attempts to improve media literacy, an issue that many agreed was a priority. One regulator was explicit about this: 'I personally believe that if we really want to find an effective method to counter fake news, this is the only way to do it. Other methods may not be effective. That is, you need to improve everyone's media literacy', they argued. 'We need to continuously promote the curriculum through school education, social education, and government departments.' The fact-checkers I spoke to voiced similar sentiment, e.g.: 'We cannot debunk every single piece of information, so we think that media literacy is more important than debunking. So we teach people how to use, like, Google images, or how to distinguish a fake account' (I08).

An interpersonal touch was again crucial, in such contexts. 'It is very important that we talk to people face to face', explained one advocate (I14). Indeed, organisations like the Taiwan FactCheck Center and MyGoPen made a point of interacting with local in-real-life communities, and organisations like FactLink and Fake News Cleaners have made it their core responsibility to bring digital and AI literacy to brick-and-mortar locations around Taiwan. Fake News Cleaners, for instance, organizes extensive activities including street fairs, public lectures, volunteer training, and programs in community colleges, schools, universities, churches, and temples. The volunteer network tailors its approach to different demographics by focusing on topics the audience cares about, like health and finances for the elderly, pop culture for teenagers, or videogames for university students. These real-world contexts then allow participants to acquire practical skills, such as how to use their smartphones more effectively, how to understand social media posts in their context, how to use tools like Google Maps, and how to identify and remove malicious apps from their devices.

Accessibility stood front and centre, in this approach. As much as the media-literacy advocates I spoke to appreciated the government's efforts to provide resources and include digital literacy in school curricula,[38] they were not necessarily convinced that these official materials were sufficiently effective. 'The government, they give you these things, but they are mostly from professors – no offence – from professors, or journalists. So they're not very user friendly' (I14). Instead, a practical, playful approach that involved people from all walks of life was much better suited for reaching a broad audience, though

[38] One critic of the DPP administration was far more vocal in their criticism, especially when it came to educational programmes staying ahead of technological developments (I16): 'Now we have to do AI literacy. And do you see our government doing anything? No, nothing. Nothing! I'm sorry if I'm a bit emotional. I'm really frustrated. I'm really worried'.

it did indeed require a great deal of time and effort. 'The work we do is difficult. Very very difficult. But it works.'

4. Conclusion: Extinguishing Digital Wildfires

Taiwan is illustrative of the problems many networked societies face today, and specifically of the risks that unverified or false information create in open societies. Taiwan has been a hotbed of what I have called CIF: contested information flows. This report has outlined how concerned Taiwanese stakeholders in policymaking, government, academia, and civil society are about such information, especially when advances in machine learning are threatening to supercharge the efforts of malicious actors to use CIF to undermine social cohesion and liberal democratic values. The interviewees described how attempts to counteract negative effects of CIF were frequently hamstrung by polarisation and the inability of political elites to work across partisan lines. They also described the uphill struggle against corporate social media and the highly emotive attention economy it created. Citizens were often insufficiently trained in using digital media safely, so the impression, and information overload, widespread distrust of opposing ideological factions, and worsening socio-economic conditions made many vulnerable to exploitative online behaviours.

However, despite the very real worries about CIF, the way that Taiwan's liberal democratic society has been addressing the issue also inspires hope. 'Taiwanese people receive disinformation every day. We have many fraud cases. It's kind of become part of Taiwanese people's daily lives. To be honest, it's a kind of training', explained one interviewee (I05). Another offered a similar assessment: 'our defenses and situation awareness are getting better. So it's a little bit conflicting, in Taiwan's society right now' (I06). My impression, both from the interviews I conducted and from half a year of fieldwork, was that Taiwanese society had generally built a great deal of resilience against the risks created by CIF, despite the arguably still substantial challenges that remain. There is much to learn from Taiwan's 'whole-of-society' approach (Thornton 2025). I want to highlight five lessons:

Lesson 1: Inequality breads discontent.

At the risk of stressing the obvious: CIF, and the malicious manipulations of liberal information environments they enable, thrive on dissatisfaction. While it is important to monitor that environment, and to intervene in it where necessary, the most effective way to extinguish digital wildfires is to remove the kindling. The slow and difficult tasks of assuring fairness and equity in our societies, and alleviating social woes, must be a priority. Disinformation issues cannot be resolve in isolation: they must be flanked by serious approaches to tackling socio-economic problems and safeguarding human dignity.

Lesson 2: A robust legal framework is essential, but overreach is a risk.

Where actors deliberately spread hate or manipulate and bully others, they are causing harm, and their actions need to have legal consequence. The platforms for such actions deserve particular legal scrutiny: it

cannot be that social media, which provide crucial communication infrastructure, are run by for-profit corporations with little to no oversight and accountability. The same goes for actors in the AI sector: as such services become part of everyday processes, they need to be governed with the same concern we attribute to infrastructures such as energy, transport, or health care. Regulating the behaviours of multi-national corporations at the national level is challenging or even impossible, certainly for small states, but actors in Taiwan frequently pointed to the EU's legislative framework as a pragmatic and effective way to address digital governance. At the same time, the need for effective governance must be balanced against the risks of government overreach. Where government action erodes the legal and ethical constraints on e.g. privacy and free speech, it plays into the hands of precisely those actors who want to see such frameworks eroded. Similarly, the concern that any given ruling party might abuse the legal framework to enforce censorship and political prosecution must be counter through 1) transparency, 2) public oversight, and 3) independent non-governmental efforts to counter CIF.

Lesson 3: Support civil society.

A lively, independent network of actors in media and civil society is ideally suited to playing the roles of watchdogs, fake-news debunkers, and educators. These actors deserve appropriate public recognition and support, and, especially, funding. Donations, private grants, and foreign aid are all avenues of financial support, but they create vulnerabilities for non-profits organisations. A recent example has been the shift in political attitudes in the US, and how Trumpist politics have on the one hand gutted US foreign aid and on the other pressured private companies like Meta to discontinue their structures for funding and support. Ways forward are to shore up international financial support for aid programmes, but also to strengthen domestic funding. Much like public news media benefit from being independent but tax-funded, crucial institutions in the analytical, fact-checking, and literacy realm would benefit from similar arrangements. In short: pay your journalists, pay your fact-checkers, pay your educators.

Lesson 4: Use tech, but not at the expense of interpersonal solutions.

All aspects of dealing with CIF can benefit from advances in ICT, including AI. Transnational efforts to share best-practices and maintain support networks are a cornerstone of assuring the equitable and effective use of such tech. At the same time, tech solutions never exist in isolation, they are part of a wider social fabric. As helpful as they can be within that fabric, the most crucial way to maintain the fabric itself is interpersonal communication in-real-life. Face-to-face efforts with communities, in brick-and-mortar town halls, are just as important as in-person outreach, small-scale workshops, and playful educational initiatives. This is hard and slow work that can take years or decades to bear fruit, but it is crucial: it is an essential way to teach skills, desensitize people to digital risks, and – most importantly – establish the foundation of trust, empathy, and understanding that is necessary to undercut the pernicious effects of CIF. Ultimately, this also means reaching across partisan aisles with empathy, as hard as this may at times be. Despite the sense that we live in 'post-truth' societies, it still matters that, as one

interviewee put it, 'nobody wants to be lied to' (I01). That, as modest as it may be, can be a useful baseline for public discourse about CIF.

Lesson 5: Don't panic.

While it can seem like liberal democratic societies are under siege from hostile forces, and the risks posed by manipulation and interference are certainly real, it is also important not to give more power to malicious actors than they deserve. Finger-pointing can alienate citizens and lead them to shut themselves in, making them more rather than less vulnerable to manipulations. Hyperbole and panic further undercut our capacity to reason for ourselves and with each other, ultimately benefitting potential abusers. It would be wise to move away from the language of 'warfare' that attackers often use themselves. Talk of 'battlefields' and 'frontlines' dial up the temperature of debate rather than reducing it. A counter-intuitive task must then be to place debate on a pragmatic footing, confront anxieties with levity and humour, counter CIF by focusing on the low-hanging fruit rather than the wedge issues, and, ultimately, to stay calm and carry on.

Appendix

1. Glossary

- *AI:* machine-learning systems that aim to solve some problem, usually by emulating human experience computationally; goals of such systems can include any or a combination of the following: 1) organizing social life more fairly (ethics), 2) increasing the efficiency of social processes (governance), 3) making profits (commerce), 4) understand a specific aspect of human life (analytics), 5) learning about the human condition (philosophy), 6) facilitating human information retrieval and sharing (communication).

- *Cognitive warfare:* 'the activities conducted in synchronization with other instruments of power, to affect attitudes and behaviours by influencing, protecting, and/or disrupting individual and group cognitions to gain an advantage'. (NATO 2023)

- *Conspiracy Theory:* an often convoluted explanation for a situation characterized by uncertainty, usually invoking spiralling narratives about the secret machinations of powerful elites, imagined or real. (adapted from Moore 2016)

- *Controversial Information Flows (CIF):* the term this report uses to capture all forms of unverified information, with a particular emphasis on information that spreads rapidly through advanced information and communication networks. This includes falsehoods that people spread unwittingly (misinformation), but also the strategic use of falsehoods and half-truths (disinformation), as well as the use of facts that malicious actors take out of context (malinformation); actors can be individuals or groups with varying degrees of strategic organisation, foreign or domestic.

- *Discourse:* the collection of knowledge, true or false, held by a society, and expressed through and shaped by communication (both verbal and non-verbal).

- *Disinformation:* factually incorrect information that actors deploy strategically to achieve their objectives.

- *Foreign Information and Manipulation Interference (FIMI):* 'a pattern of behaviour that threatens or has the potential to negatively impact values, procedures and political processes. Such activity is manipulative in character, conducted in an intentional and coordinated manner. Actors of such activity can be state or non-state actors, including their proxies inside and outside of their own territory.' (EEAS 2025)

- *GenAI:* 'the class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content.' (The White House 2023)

- *Malinformation:* factually correct information that actors deploy in a manipulative way to achieve their objectives.

- *Misinformation:* information that may seem plausible but is factually incorrect.

- *Technological affordance:* the aspects of a technology's design that encourage certain uses rather than others. (adapted from Hutchby 2001)

2. Taiwan's Fact-Checker Ecosystem

Taiwan possesses a lively civil society that has produced numerous organisations involved in countering CIF and FIMI. For more information, see e.g. Wu et al. (2020, 78-82) and Dotson et al (2024, 9-13).

- **Cofacts:** Fact-checking organisation (https://en.cofacts.tw/).
- **Doublethink Lab (DTL):** Think tank studying authoritarian disinformation campaigns around the world, with a specific focus on PRC FIMI (https://doublethinklab.org/).
- **FactLink:** Fact-checking organisation focused on media and AI literacy (https://www.factlink.tw/).
- **Fake News Cleaners (假新聞清潔劑):** Volunteer network focusing on digital media literacy (https://www.fakenewscleaner.tw/).
- **Institute for National Defense and Security Research (INDSR):** Think tank under the Legislative Yuan, focused on military affairs, including cyber-warfare (https://indsr.org.tw/en/index).
- **Institute of Watch Internet Network (iWIN):** Non-profit focusing on child and youth protection online (https://i.win.org.tw/).
- **MyGoPen:** Fact-checking organisation (https://www.mygopen.com/) and member of the International Fact Checking Network (IFCN).
- **Rum Toast Rumor & Truth (蘭姆酒吐司):** Fact-checking organisation (https://rumtoast.com/).
- **Taiwan FactCheck Center (TFC):** Fact-checking organisation (https://en.tfc-taiwan.org.tw/) and member of the International Fact Checking Network (IFCN).
- **Taiwan Information Environment Research Center (IORG):** Think tank studying the Chinese-language information environment (https://iorg.tw/).

3. The Four Ds of Disinformation

Strategies to manipulate information are often broken down into the following '4D' categories, originally coined by analyst Ben Nimmo (2015). Note that these mechanisms are designed to conceptualise disinformation (intentional lies); they may not be a neat fit when examining misinformation (questionable but unintentional information uses).

Dismiss: attempt to silence opposing views and seed distrust of diverging information sources, e.g. by discrediting the source, undermining its creditability, making light of its argument, and issuing insults.

Distort: misrepresent information, twist facts, and reframe the narrative to fit a specific argument or interpretation.

Distract: changing the subject and shifting attention, e.g. through 'what-aboutism' and false comparisons.

Dismay: supress opposition through intimidation and scare tactics.

4. Democratic & ICT in Taiwan – an abridged timeline

- 1894: KMT founded
- 1912: Republic of China (ROC) founded in Nanjing
- 1927-1949: Civil War – the KMT flees to Taiwan
- 1949-1987: Martial law and dictatorship
- 1947: '2-28 Incident' (White Terror)
- 1986: DPP founded
- 1987: Martial law ends
- 1987: TSMC founded
- 1990: Wild Lily student movement
- 1996: First democratic presidential election
- 2000: First democratic power transfer
- 2004: Facebook launches (US)
- 2006: Twitter launches (US)
- 2011: LINE launches (Japan)
- 2010: ROC enacts Personal Data Protection Act (PDPA; enforced 2012)
- 2012: g0v founded and first hackath0n
- 2013: KMT appoints Jaclyn Tsai (蔡玉玲) as 'cyber-minister'
- 2014: Sunflower Movement
- 2014: vTaiwan citizen consultation platform launched
- 2015: data protection amended
- 2016 election: DPP landslide win

- 2016: DPP appoints Audrey Tang (唐鳳) as 'cyber-minister'
- 2018: local elections & referenda: Han Kuo-yu's populism begins
- 2018: Cybersecurity Management Act (CMA) passed
- 2019: data protection amended (enforced 2022)
- 2022: ChatGPT launched (US)
- 2023: DeepSeek launched (PRC)
- 2023: TAIDE initiative launches (homegrown LLM).
- 2024: Fraud Crime Hazard Prevention Act enacted
- 2025: CMA amended to prevent use of suspicious foreign apps
- 2025: 'Bluebird' recall movement

References

Ai, M. (2019). Technology empowerment and collaborative participation: The mobilization process of science and technology network community in social movement – A case study based on the "g0v" network community in Taiwan. *Taiwan Research Journal*, 3(163), 13-23. [Chinese].

Allcott, H. & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211-236.

APCSS (2010, December 16). Workshop: The strategic impact of media in comprehensive crisis management. Retrieved 6 November 2025 from https://dkiapcss.edu/the-strategic-impact-of-media-in-comprehensive-crisis-management-workshop/.

Arcos, R., Chiru, I., & Ivan, C. (Eds.). (2024). *Routledge handbook of disinformation and national security*. Routledge.

Bastos, M. T. & Mercea, D. (2019). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review*, 37(1), 38-54.

BBC (2016, January 18). Taiwan election: How a penitent pop star may have helped Tsai win. *BBC News Asia*, retrieved 6 November 2025 from https://www.bbc.com/news/world-asia-35340530#:~:text=Image%20source%2C%20Reuters,a%20variety%20show%20shown%20online.

Beauchamp-Mustafaga, N. & Drun, J. (2021). Exploring Chinese military thinking on social media manipulation against Taiwan. *China Brief*, The Jamestown Foundation, retrieved 6 November 2025 from https://jamestown.org/program/exploring-chinese-military-thinking-on-social-media-manipulation-against-taiwan/.

Becerra, M. (2022, August 24). The battle for reality: Chinese disinformation in Taiwan. *Geopolitcial Monitor Situation Report*. Retrieved 6 November 2025 from https://www.geopoliticalmonitor.com/the-battle-for-reality-chinese-disinformation-in-taiwan/.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

Berry, D. M. (2011). *The philosophy of software – Code and mediation in the digital age*. Palgrave Macmillan.

Bontridder, N. & Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy*, 3, e32.

Brady, A.-M. (2008). *Marketing dictatorship: propaganda and thought work in contemporary China*. Rowman & Littlefield.

Bui, P.-T. (2025, August 4). Disinformation, manipulation, and media: Taiwan's insights from the 2025 Taipei Forum. Taiwan FactCheck Center, retrieved 6 November 2025 from https://en.tfc-taiwan.org.tw/disinformation/.

Cervi, L., Tejedor Calvo, S., & Blesa, F. (2023). TikTok and Political Communication: The Latest Frontier of Politainment? A Case Study. *Media and Communication*, 11. 10.17645/mac.v11i2.6390.

Chang, H.-C., Haider, S., & Ferrara, E. (2021). Digital Civic Participation and Misinformation during the 2020 Taiwanese Presidential Election. *Media and Communication*, 9(1), 144–157.

Chen, S. (2024). Countering AI disinformation: Lessons from Taiwan's 2024 election defense strategies. Information Resilience & Integrity Symposium, retrieved 6 November 2025 from https://saferinternetlab.org/wp-content/uploads/2025/08/Panel-4-Summer.pdf.

Chirovici, E.-O. (2014). *Rumors that changed the world: a history of violence and discrimination*. The Rowman & Littlefield Publishing Group.

Claypool, R. (2023, September 26). Chatbots are not people: Designed-in dangers of human-like A.I. systems. *Public Citizen*, retrieved 6 November 2025 from https://www.citizen.org/article/chatbots-are-not-people-dangerous-human-like-anthropomorphic-ai-report/.

Creemers, R., Pappagianneas, S., & Knight, A. (Eds) (2025). *The Emergence of China's Smart State*. Bloomsbury, retrieved 6 November 2025 from https://www.bloomsburycollections.com/monograph?docid=b-9798881817602.

Daley, M. & Johnston, C. (2025, August 25). The rise of humanlike chatbots detracts from developing AI for the human good. *The Conversation*, retrieved 6 November 2025 from https://theconversation.com/the-rise-of-humanlike-chatbots-detracts-from-developing-ai-for-the-human-good-261787.

Diaz Ruiz, C. & Nilsson, T. (2023). Disinformation and Echo Chambers: How Disinformation Circulates on Social Media Through Identity-Driven Controversies. *Journal of Public Policy & Marketing*, *42*(1), 18–35.

DISARM Foundation (n/d). DISARM helps identify and respond to malign information influence operations, retrieved 6 November 2025 from https://www.disarm.foundation/.

Doroshenko, L. & Tu, F. (2023). Like, share, comment, and repeat: Far-right messages, emotions, and amplification in social media. *Journal of Information Technology & Politics*, *20*(3), 286–302.

Dotson, J., Wu, A., & Kielm, W. (2024). Taiwan's fight for global democracy: The role of civil society. Global Taiwan Institute, retrieved 6 November 2025 from https://globaltaiwan.org/wp-content/uploads/2024/07/OR_TWs-Fight-for-Democracy.pdf

Doublethink Lab (2021). Deafening whispers: China's information operation and Taiwan's 2020 election. Retrieved 6 November 2025 from https://drive.google.com/file/d/1FW35t93GvMJ3W6rqbPhAm6lNZ4uy66jD/view.

Doublethink Lab (2024a, January 5). *Artificial Multiverse: Foreign Information Manipulation and interference in Taiwan's 2024 National Elections* (人造多重宇宙：2024 台灣大選境外資訊操作與影響觀察報告). Retrieved 28 June 2024 from https://medium.com/doublethinklab-tw/%E4%BA%BA%E9%80%A0%E5%A4%9A%E9%87%8D%E5%AE%87%E5%AE%99-2024-%E5%8F%B0%E7%81%A3%E5%A4%A7%E9%81%B8%E5%A2%83%E5%A4%96%E8%B3%87%E8%A8%8A%E6%93%8D%E4%BD%9C%E8%88%87%E5%BD%B1%E9%9F%BF%E8%A7%80%E5%AF%9F%E5%A0%B1%E5%91%8A-493423f9bba8.

Doublethink Lab (2024b, January 19). 2024 Taiwan election: Survey report of the impact of overseas information – a preliminary analysis (2024 台灣選舉：境外資訊影響觀測報告初步分析). Retrieved 6 November 2025 from https://medium.com/doublethinklab-tw/2024-%E5%8F%B0%E7%81%A3%E9%81%B8%E8%88%89-%E5%A2%83%E5%A4%96%E8%B3%87%E8%A8%8A%E5%BD%B1%E9%9F%BF%E8%A7%80%E6%B8%AC%E5%A0%B1%E5%91%8A%E5%88%9D%E6%AD%A5%E5%88%86%E6%9E%90-fe7f819aeabd.

DSET (2024). GenAI and Democracy: AI-Driven Disinformation in Taiwan's 2024 Presidential Election and Lessons for the World. *Research Institute for Democracy, Society and Emerging Technology (DSET)*, retrieved 6 November 2025 from https://dset.tw/wp-content/uploads/2024/10/genAI-2024-Election-Report-.pdf.

EEAS (2025). *Information Integrity and Countering Foreign Information Manipulation & Interference (FIMI)*. European Union External Action Service, retrieved 14 October 2025 from https://www.eeas.europa.eu/eeas/information-integrity-and-countering-foreign-information-manipulation-interference-fimi_en.

EU (2022). *The Digital Services Act package*. Retrieved 6 November 2025 from https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package.

Fell, D. (2012). *Government and politics in Taiwan*. Routledge.

Felt, U. (Ed.) (2017). *The handbook of science and technology studies* (Fourth edition.). The MIT Press.

Feng, E. (2024, January 11). Taiwan deals with lots of misinformation, and it's harder to track down. *NPR*, retrieved 6 November 2025 from https://www.npr.org/2024/01/11/1216340756/taiwan-election-disinformation-social-media-ptt#:~:text=Fake%20stories%20abound,%22sexual%2Dassault%20island.%22.

Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2018, February). Measuring the reach of 'fake news' and online disinformation in Europe. *Reuters Institute Factsheet*. Retrieved May 1, 2024, from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/Measuring%20the%20reach%20of%20fake%20news%20and%20online%20distribution%20in%20Europe%20CORRECT%20FLAG.pdf.

François, Camille (2019). Actors, behaviors, content: A disinformation ABC – Highlighting three vectors of viral deception to guide industry & regulatory responses. Transatlantic Working Group paper, retrieved 6 November 2025 from https://www.ivir.nl/publicaties/download/ABC_Framework_2019_Sept_2019.pdf.

Gainous, J., Han, R., MacDonald, A. W., & Wagner, K. M. (2023). *Directed digital dissidence in autocracies : how China wins online*. Oxford University Press.

Grosfeld, T. (2025, September 17). 'Vriendschap' met AI: Beter een nepvriend dan een goede buur ('Friendship' with AI: Better a fake friend than a good neihbour). *De Groene Amsterdammer*, retrieved 6 November 2025 from https://www.groene.nl/artikel/beter-een-nepvriend-dan-een-goede-buur.

Halavais, A. (2009). *Search engine society*. Polity Press.

Han, R. (2018). *Contesting cyberspace in China: Online expression and authoritarian resilience*. Columbia University Press.

Han, R. (2021). Cyber nationalism and regime support under Xi Jinping: The effects of the 2018 constitutional revision. *The Journal of Contemporary China*, *30*(131), 717–733.

Ho, M. (2020). Watchdogs and partners: Taiwan's civil society organizations. In R. Youngs (Ed.), *Global Civil Society in the Shadow of Coronavirus* (pp. 11-16). Carnegie Foundation for International Peace.

Hofstede, S. E. (2022). *Claiming community: How Chinese ethno-nationalism imagines contact points for influence in Singapore and Taiwan*. ProQuest Dissertations & Theses.

Hsiao, Y. & Yu, E. C. (2020). Polarization perception and support for democracy: The case of Taiwan. *Journal of Asian and African Studies*, *55*(8), 1143-1162.

Hsu, C.-J. (2014). *The construction of national identity in Taiwan's media, 1896-2012*. Brill.

Hsu, E. (2024). 2024 Taiwan election: The increasing polarization of Taiwanese politics — Reinforcement of conspiracy narratives and cognitive biases. *Doublethink Lab*, retrieved 6 November 2025 from https://medium.com/doublethinklab/2024-taiwan-election-the-increasing-polarization-of-taiwanese-politics-reinforcement-of-2e0e503d2fe2.

Huang, C. & Kuo, T. (2022). Actual and perceived polarization on independence-unification views in Taiwan. *Asian Journal of Communication*, *32*(2), 75–92.

Huang, J.-N. (2023). China's propaganda and disinformation operations in Taiwan: A sharp power perspective. *China: An International Journal*, 21(2), 143-170.

Huang, K. (2018, September 14). Taiwanese official criticised for handling of Typhoon Jebi evacuation found dead in Osaka. *South China Morning Post*.

Hung, C.L, Fu, W.C., Liu, C.C., Tsai, H.J. (2024). *AI disinformation attacks and Taiwan's responses during the 2024 presidential election*. Thomson Foundation, retrieved 6 November 2025 from https://www.thomsonfoundation.org/media/268943/ai_disinformation_attacks_taiwan.pdf.

Hung, T.-C. & Hung, T.-W. (2020). How China's cognitive warfare works: A frontline perspective of Taiwan's anti-disinformation wars. *Journal of Global Security Studies*, 7(4), 1–18.

Hutchby, I. (2001). Technologies, texts and affordances. *Sociology*, 35(2), 441–456.

Hutchins, E. M., Cloppert, M. J., & Amin, R. M. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Lockheed Martin Corporation White Paper*, retrieved 6 November 2025 from https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf.

Jarke, J., Prietl, B., Egbert, S., Boeva, Y., Heuer, H., & Arnold, M. (2024). *Algorithmic regimes: Methods, interactions, and politics*. Amsterdam University Press.

Kania, E. B. & Costello, J. (2021). Seizing the commanding heights: the PLA Strategic Support Force in Chinese military power. *Journal of Strategic Studies*, *44*(2), 218–264.

Kao, S.S. (2021), Taiwan's Response to Disinformation: A Model for Coordination to Counter a Complicated Threat. National Bureau of Asian Research Special Report No.93, retrieved 6 November 2025 from https://www.nbr.org/publication/taiwans-response-to-disinformation-a-model-for-coordination-to-counter-a-complicated-threat/.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14-29.

Kjellen, M. & Wong, C. (2023). Governance: A 'whole-of-society' approach. In: United Nations Educational, Scientific and Cultural Organization, The United Nations World Water Development Report 2023 (pp.172-182), retrieved 6 November 2025 from https://www.un-ilibrary.org/content/books/9789210026208c021.

Kleinman, D. L. & Moore, K. (Eds) (2014). *Routledge Handbook of Science, Technology, and Society*. Routledge.

König, R. & Rasch, M. (Eds) (2014). *Society of the query reader – Reflections on web search*. Amsterdam: Institute of Network Cultures. Retrieved 6 November 2025 from https://networkcultures.org/blog/publication/society-of-the-query-reader-reflections-on-web-search/.

Kranzberg, M. (1995). Technology and History: 'Kranzberg's Laws'. *Bulletin of Science, Technology & Society*, *15*(1), 5–13.

Krumbein, F. (2023). Populist discourses in Taiwan and the case of Han Kuo-yu. *International Journal of Taiwan Studies*, 7(2), 229–263.

Kuhn, A. (2016, January 18). Backlash after singer waves Taiwanese flag rouses ruling party tensions. *NPR Politics & Policy*, retrieved 6 November 2025 from https://www.npr.org/sections/parallels/2016/01/18/463478274/backlash-after-singer-waves-taiwanese-flag-rouses-ruling-party-tensions.

Kuo, M. & Wu, A. (2022, March 27). *The battle for hearts and minds: A dive into how Taiwan's pro-China media depicts Ukraine and Russia. A Broad and Ample Road*. Retrieved May 24, 2024, from https://ampleroad.substack.com/p/the-battle-for-hearts-and-minds-a.

Last Week Tonight (2025, June 23). *AI Slop*. HBO, retrieved 14 October 2025 from https://www.youtube.com/watch?v=TWpg1RmzAbc

Lau, S. (2024, January 10). China bombards Taiwan with fake news ahead of election. *Politico*. Retrieved May 24, 2024, from https://www.politico.eu/article/china-bombards-taiwan-with-fake-news-ahead-of-

election/?fbclid=IwAR3cBIC0Ug_nWovUPqFGKVVXHSdu9_qCoZjGCnrm1wygz08rgPragcYyu8_aem_AUJKg4n7nln1ASI-8xHiJXPp3eTxS0IqZ9Jeg6B25YOUDYrKGAFORv8YbS-TpG_W1dOQXauWNFPniGtcFJ8NfxHo.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., et al. (2018). The science of fake news. *Science*, 359(6380), 1094-1096.

Lee, M.-C. (2020). Free the data from the birdcage: Opening up data and crowdsourcing activism in Taiwan. *Political and Legal Anthropology Review*, 43(2), 247-261.

Lee, M.-C. (2024). Decoding China's digital offensive: An analysis of information warfare tactics in Taiwan's 2024 presidential election. *Research Institute for Democracy, Society and Emerging Technology (DSET)*, retrieved 6 November 2025 from https://dset.tw/en/research/00100/.

Lee, Y. & Blanchard, B. (2022, September 14). 'Humour over rumour': Taiwan eyes Ukraine messaging model if China attacks. *Reuters*, retrieved 6 November 2025 from https://www.reuters.com/world/asia-pacific/humour-over-rumour-taiwan-eyes-ukraine-messaging-model-if-china-attacks-2022-09-14/

Lin, R. & Wu, F. (2019, April 27). The black hole of fake news: Taiwan's online 'opinion war' arrived. *CommonWealth Magazine*, 671. Retrieved 24 May 2024 from https://english.cw.com.tw/article/article.action?id=2375.

Lin, S. (2018, September 19). Lawmakers weigh in on fake news, diplomat's suicide. *Taipei Times*, retrieved 6 November 2025 from https://www.taipeitimes.com/News/taiwan/archives/2018/09/19/2003700718.

MacKenzie, A. & Bhatt, I. (2020). Lies, bullshit and fake news: Some epistemological concerns. *Postdigital Science and Education*, 2, 9-13.

Maddox, J.D., Gentzel, C., & Levis, A. (2021, May 10). Toward a whole-of-society framework for countering disinformation. Modern War Institute at West Point, retrieved 6 November 2025 from https://mwi.westpoint.edu/toward-a-whole-of-society-framework-for-countering-disinformation/.

MITRE (2025). ATT&CK. Retrieved 6 November 2025 from https://attack.mitre.org/.

Moore, A. (2016). Conspiracy and conspiracy theories in democratic politics. *Critical Review*, 28(1), 1-23.

NATO (2023, April 5). *Cognitive Warfare: Strengthening and Defending the Mind.* Retrieved 14 October 2025 from https://www.act.nato.int/article/cognitive-warfare-strengthening-and-defending-the-mind/

Nimmo, B. (2015, May 19). Anatomy of an Info-War: How Russia's Propaganda Machine Works, and How to Counter It. *StopFake.org*, retrieved 15 October form https://www.stopfake.org/en/anatomy-of-an-info-war-how-russia-s-propaganda-machine-works-and-how-to-counter-it/.

Nimmo, B., Shawn Eib, C., & Tamora, L. (2019). Cross-platform spam network targeted Hong Kong protests: 'Spamoflage Dragon' used hijacked and fake accounts to amplify video content. Graphika, retrieved 6 November 2025 from https://graphika.com/posts/graphika-report-spamouflage.

NSTC (2025). TAIDE: Trustworthy AI Dialogue Engine. Retrieved 6 November 2025 from https://en.taide.tw/.

Pariser, Eli (2012). *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin Press.

Pathak, S. (2023, November 16). Opinion: Racism, disinformation cast shadow on India-Taiwan cooperation. *NDTV*, retrieved 6 November 2025 from https://www.ndtv.com/opinion/racism-disinformation-cast-shadow-on-india-taiwan-cooperation-4579209.

Primig, F., Szabó, H., & Lacasa, P. (2023). *Remixing war: An analysis of the reimagination of the Russian–Ukraine war on TikTok*. Retrieved 29 June 2024 from https://doi.org/10.3389/fpos.2023.1085149.

Qi, D. & Lin, S. (2021). Dividing without conquering: Generation, class, ethnicity, and nationalism in Taiwan's 2016 presidential election. *Issues & Studies: A Social Science Quarterly on China, Taiwan, and East Asian Affairs*, 57(3), 1-32.

Qureshi, I. & Bhatt, B. (2024). Social media-induced polarisation. *Information Systems Journal*, 34, 1425-1431. https://doi-org.leidenuniv.idm.oclc.org/10.1111/isj.12525

Rainie, L., Anderson, J., & Albright, J. (2017). The future of free speech, trolls, anonymity, and fake news online. *Pew Research Center*. Retrieved 28 June 2024 from https://www.pewresearch.org/internet/2017/03/29/the-future-of-free-speech-trolls-anonymity-and-fake-news-online/.

Rauchfleisch, A., Tseng, T.-H., Kao, J.-J., & Liu, Y.-T. (2023). Taiwan's Public Discourse About Disinformation: The Role of Journalism, Academia, and Politics. *Journalism Practice*, 17(10), 2197-2217.

Rich, T., Eliassen, I., & Dahmer, A. (2018, March 6). What Happened to Taiwan's Support for Same-Sex Marriage? *Asia Dialogue*, retrieved 15 October 2025 from https://theasiadialogue.com/2019/03/06/what-happened-to-taiwans-support-for-same-sex-marriage/

Rid, T. (2020). *Active measures: the secret history of disinformation and political warfare*. Farrar, Straus and Giroux.

Rigger, S. (2014). *Why Taiwan matters: small island, global powerhouse*. Rowman & Littlefield.

Robertson, S. (2024, November 4). 'Is Sovereign AI Anti-Democratic or the Key to Securing Taiwan's Democracy?' Domino Theory, retrieved 6 November 2025 from https://dominotheory.com/is-sovereign-ai-anti-democratic-or-the-key-to-securing-taiwans-democracy/.

ROC (2024, July 31). *Fraud Crime Hazard Prevention Act*. Ministry of the Interior, retrieved 6 November 2025 from https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=D0080226.

Rühlig, T. N. (2022). *China's foreign policy contradictions: lessons from China's R2P, Hong Kong, and WTO policy*. Oxford University Press.

Sampson, T. D., Maddison, S., & Ellis, D. (Eds.). (2018). *Affect and social media: emotion, mediation, anxiety and contagion*. Rowman & Littlefield International.

Schäfer, C. & Schadauer, A. (2019). Online fake news, hateful posts against refugees, and a surge in xenophobia and hate crimes in Austria. In G. Dell'Orto (Ed.), *Refugee news, refugee politics: Journalism, public opinion and policymaking in Europe*. Routledge.

Schäfer, F. (2023). *Konnektiver Zynismus. Politik und Kultur im digitalen Zeitalter [Connective cynicism. Politics and culture in the digital age]*. Bielefeld: transcript.

Schneider, F. (2018). *China's digital nationalism*. Oxford: Oxford University Press.

Schneider, F. (2019). Digital democracy in Taiwan: The Sunflower Movement and its legacies. *Taiwan Fellowship Report*. Retrieved 6 November 2025, from https://taiwanfellowship.ncl.edu.tw/files/scholar_publish/1764-gydpxhhgkfpcxht.pdf.

Schneider, F. (2023). Reality decoupling: Rumours, disinformation, and studying the politics of truth in digital Asia. *Asiascape: Digital Asia*, 10(1-2), 181-207.

Shirky, C. (2008). *Here comes everybody – The power to organize without organizations*. Penguin Books.

Sismondo, S. (2010). *An introduction to science and technology studies* (2nd ed.). Blackwell.

Slawotsky, J. (2021). Is China guilty of committing genocide in Xinjiang? *Chinese Journal of International Law*, 20(3), 625–635.

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review*, 34(3), 150-160.

Strömbäck, J., Wikforss, Å., Glüer, K., Lindholm, T., & Oscarsson, H. (Eds.). (2022). *Knowledge resistance in high choice information environments*. Routledge. Retrieved May 24, 2024, from https://library.oapen.org/bitstream/id/b5a7538a-f752-4bef-bf35-aa97f9158df7/9781000599121.pdf.

Sullivan, J. & Smyth, J. (2016). Taiwan's 2016 Presidential and Legislative Elections. Journal of the British Association for Chinese Studies, 6, retrieved 6 November 2025 from https://bacsuk.org.uk/journal/journal-current-and-past-entries/taiwans-2016-presidential-and-legislative-elections.

Tandoc, E. C., Jr., Lim, Z. W., & Ling, R. (2018). Defining 'fake news'. *Digital Journalism*, 6(2), 137-153.

Terren, L. & Borge-Bravo, R. (2021). Echo Chambers on Social Media: A Systematic Review of the Literature. *Review of Communication Research*, 9, 99-118. Retrieved 15 October 2025 from https://www.rcommunicationr.org/index.php/rcr/article/view/94

The White House (2023, October 30). *Executive Order 14110—Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. Office of the Federal Register, National Archives and Records Administration, retrieved 14 October 2025 from https://www.govinfo.gov/app/details/DCPD-202300949.

Thomson, J. (2023, November 21). Taiwan says false information spread about Indian migrant workers online. *Taiwan News*, retrieved 6 November 2025 from https://www.taiwannews.com.tw/news/5044229.

Thornton, M. (2025). Resisting disinformation: theorising whole-of-society and sociotechnical resistance. *European Journal of International Relations*, online first. Retrieved 6 November 2025 from https://journals.sagepub.com/doi/10.1177/13540661251382639.

Tobin, D. (2021). Genocidal processes: social death in Xinjiang. *Ethnic and Racial Studies*, 45(16), 93–121.

Tseng, P.-Y. & Chen, Y.-J. (2020), *An analysis on the impact of false information on Taiwanese voters*. Taiwan Democracy Lab & Doublethink Lab, retrieved 6 November 2025 from https://medium.com/doublethinklab/an-analysis-on-the-impact-of-false-information-on-taiwanese-voters-c061500a898c.

Ventre, D. (Ed.) (2014). *Chinese cybersecurity and defense*. Wiley.

Waisbord, S. & Amado, A. (2017). Populist Communication by Digital Means: Presidential Twitter in Latin America. *Information, Communication and Society* 20(9): 1330-46.

Waisbord, S. (2018). Truth is what happens to news: On journalism, fake news, and post-truth. *Journalism Studies*, 19(13), 1866-1878.

Walorska, A. M. (2020). *Deepfakes and Disinformation*. Friedrich Naumann Foundation for Freedom, retrieved 6 November 2025 from https://shop.freiheit.org/download/P2@897/269071/FNF_DEEPFAKES_Broschuere_EN_web.pdf.

Wang, A. H.-E. (2019). The myth of polarization among Taiwanese Voters: The missing middle. *Journal of East Asian Studies*, *19*(3), 275–287.

Wang, T.-L. (2020). Does fake news matter to election outcomes? The case study of Taiwan's 2018 local elections. *Asian Journal for Public Opinion Research*, *8*(2), 67–104. Retrieved 15 October 2025 from https://doi.org/10.15206/ajpor.2020.8.2.67

Wang, H.E. (2024). AI-generated disinformation in the 2024 Taiwan presidential election. Friedrich Naumann Foundation, retrieved 6 November 2025 from https://shop.freiheit.org/#!/Publikation/1825.

Wang, X., Tsai, T.Y., Lin, L. Et al. (2025). Spotting the fakes: A deep dive into GAN-generated face detection. *ACM Transactions on Multimedia Computing, Communication, and Applications*, 21(7), 1-24. Retrieved 6 November 2025 from https://dl.acm.org/doi/full/10.1145/3742786.

Wasserstrom, J. (2025). *The Milk Tea Alliance: Inside Asia's struggle against autocracy and Beijing*. Columbia Global Reports.

Wigell, M., Mikkola, H., & Juntunen, T. (2021), *Study: Best practices in the whole-of-society approach in countering hybrid threats*. Policy Department for External Relations, Directorate General for External Policies of the Union, retrieved 6 November 2025 from https://www.europarl.europa.eu/thinktank/en/document/EXPO_STU(2021)653632

Wu, C.E., Chu, Y.H, & Taiwan Foundation for Democracy (2020). Social Media, Disinformation, and Democracy in Taiwan. In Asia Democracy Research Network (Eds). *Social Media, Disinformation, and Democracy in Asia: Country Cases* (pp.67–95). Retrieved 6 November 2025 from http://www.adrnresearch.org/publications/list.php?at=view&idx=118&ckattempt=3.

Xu J. & Yang L. (2025). Celebrity public relations in China: Power, politics and pop propaganda. *The China Quarterly*, 262, 499-514.

Xu, J. & Qu, L. (2025). 'Telling China's Stories Well' through wanghong: Rethinking China's soft power and public diplomacy in the influencer era. In Xu, J., G. Donnar, & D. Garg (Eds). *Asian Celebrity Cultures in the Digital Age*, Hong Kong University Press (pp.54-69).

Xu, J. & Schneider, F. (2025). Influencers as emerging actors in global digital propaganda. *European Journal of Cultural Studies*, 0(0). Retrieved 6 November 2025 from https://journals.sagepub.com/doi/full/10.1177/13675494251351221.

Yang, T.-T., Hsung, R.-M., Chen, S.-H., Du, Y.-R., Lin, Y.-J., Yen, N.-S., Wu, C.-T., & Liu, H.-L. (2021). The mechanisms of trust formation under different conditions of political identity: An experiment among Taiwanese voters. *Current Sociology*, *69*(6), 879–899.

Yeh, S.-P. (2025, August 27). The Democratic Progressive Party (DPP) is strengthening its public opinion defense, with Hsu Kuo-yung demanding the 222 rule (民進黨強化輿情攻防，徐國勇要求 222 法則). *CNA*, retrieved 6 November 2025 from https://www.cna.com.tw/news/aipl/202508270112.aspx.

Yu, C.H. (2023, August 8). *US Skepticism Narratives and Where They Come From*. Taiwan Information Environment Center (IORG), retrieved 28 June 2024 from https://iorg.tw/_en/a/us-skepticism-238.

Yu, H. & Creemers, R. (Eds) (2025). *Automating Governance in China? Data-Driven Systems in the Scoring Society*. Leiden University Press.

Zhang, X. (2011). *The transformation of political communication in China : from propaganda to hegemony*. World Scientific.