Ministerial Executive Summary

This policy brief focuses on the experiences of young people exposed to, engaging with, and fighting information disorder via social media during Covid-19. Globally, young people are the world's largest and fastest growing age group, with 40% of the world's population currently younger than 24 years old.¹ Where young people were once considered (and still often believe themselves to be) more resistant to Covid-19 infection than older groups, the fastest growing global infection rates today are occurring among those in their twenties.² A similar paradox emerges with information disorder: although young people frequently self-report high levels of conviction that they can recognise misinformation online³ and do seem to recognise text-based manipulation (e.g. clickbait headlines) faster than older users, this appears to lessen when visual material is involved.⁴ Indeed, a recent study in the U.S. found young people more likely than older users to believe misinformation related to Covid-19.⁵

⁴ See Nieman Lab (2018) Younger generations are actually better at telling news from opinion than those over age 50 - Nieman Journalism Lab. (n.d.).
A main concern of this brief is the fact that young people constitute the largest percentage of believers in Covid-19 conspiracy theories, especially in wealthy countries with high levels of education\(^6\) Social science and industry research reveals that as a consumer market, young people in these countries tend to favor messengers and messages perceived as ‘authentic’ (alternately described as personally trustworthy, entertaining, or relatable) over messages delivered by institutions that are driven solely by argument and presentation of facts.\(^7\) Misinformation formats like rumour, stigma and conspiracy likewise draw power from emotionally driven narratives that appeal to authentically felt personal experience.\(^8\)

Another concern is that the substantial amount of time young people have amassed on platforms puts them at higher than average risk for ‘social pooling’ in algorithmically created communities.\(^9\) Left undisturbed, these can stagnate, with young people who once thought themselves to be gathered according to consumer taste suddenly charged with navigating disturbing levels of ideological polarisation and/or online grooming efforts by organised political extremists.\(^10\)

This brief addresses how the public health sector along with a coalition of civil servants, technology companies, CSOs (civil society organisations), and researchers should understand

---


\(^8\) These modalities have been assessed as the top misinformation threats globally. Islam, M. S., et al. (2020). COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. The American Journal of Tropical Medicine and Hygiene, 103(4), 1621–1629. [https://doi.org/10.4269/ajtmh.20-0812](https://doi.org/10.4269/ajtmh.20-0812)

\(^9\) Throughout this brief, the term “pooling” is used to refer to algorithmic sorting of users into platform-based groups, as a way to bypass more common but largely inaccurate terms like “echo chamber” and “filter bubble.” For more see the “Ideological Communications Circuit” framework in this document. See also Bruns, A. (2019). Filter bubble. Internet Policy Review, 8(4). [https://policyreview.info/concepts/filter-bubble](https://policyreview.info/concepts/filter-bubble)

and respond to young people as they encounter, evaluate, and circulate information in social media contexts. Compiled by a media researcher at Macquarie University with twenty years experience tracking young people’s internet use, this brief focuses on young people’s unique needs in the areas of education, protection, and care. In correspondence with the WHO’s already suggested framework for dealing with the infodemic, this brief covers the following terrain:

- It analyses the challenges of speaking about young people as a demographic to be publicly served, when age disaggregated data collection is scarce outside of industry, and conditions for protecting and regulating the behaviour of young people differ across institutions, nations, legal jurisdictions and digital platforms.

- It summarises the major challenges currently faced by young people during Covid-19, with a focus on their self-perceptions of health exceptionalism, their economic precarity, their mental health concerns, their low levels of institutional trust, their taste for emotionally authentic messengers over argument-driven messaging, and their at-risk status for extremist targeting.

- It describes how young people are currently navigating the infodemic, analysing their perceptions and realities with regard to identifying and resisting health misinformation, with a particular focus on rumour, stigma and conspiracy theories.

- It considers how new patterns of social media use among young people pose challenges for those fighting the infodemic. These include platform choices that emphasise co-creating video, gaming, and live streaming; a preference for contextual ambiguous visual communication practices like selfie, GIF and meme circulation; social engagement styles that are largely determined by algorithmic protocols; and high awareness of the demand to present and protect the ‘self as brand’ online by simultaneously proclaiming an aversion to, yet engaging with internet ‘drama.;

This brief also assesses the effectiveness for young people of strategies currently used to battle the infodemic, providing best practice examples of campaigns directed toward and led by young people. The findings are as follows:

---

● **Debunking** is most effective when used among trusted communicators (to avoid shaming dynamics); least effective when in online environments devoted to entertainment, where it can be perceived as combative.

● **Pre-bunking** is most effective when results are shared as part of an online socialisation process; least effective when effective pre-bunking games languish unpublicised online.

● **Media literacy** is most effective when it is updated to reflect platform, algorithmic, and datification dynamics; least effective when it treats online communication as if it were an extension of broadcast media.

● **Prohibitive efforts** are most effective when they occur via platform regulation (e.g. flagging content, removing users for violations, etc.); least effective when occurring via legal regulation (because of varying age limits for criminal prosecution of young people.)

● **Info-activism** is most effective when it occurs as a spontaneous on the ground response to an event (e.g. K-pop activists organising against alt-right misinformation campaigns on Twitter); least effective when it is perceived as opportunism (often referenced as ‘you are using us for clout.’)

● **Influencing** is most effective when messengers are perceived as authentic, funny, speaking from personal experience and/or ‘one of us’; least effective when messages appear official, mechanical, and/or delivered solely for compensation.

● **Nudging** is most effective in environments where users have been primed to expect brief interruptions (e.g. gaming platforms) and use certain tools (e.g. online stickers); least effective when it is perceived as too frequent, or too overbearing.

The recommendations in this brief are grounded in four frameworks. Each provides researchers, educators, policy makers, and practitioners interlocking perspectives on how young people receive information online; how they find themselves within separate ‘pools on the same platforms; how their social regulation processes mimic social media logics; and how young people deploy concepts like authenticity as code terms to describe their perceptions of environmental risk.
• **The Influence Ecosphere** framework helps explain how and why information with a high emotional charge seems to dominate social media flows. It can be especially useful for understanding dynamics like the spread of viral rumours online.

• **The Ideological Communication Circuit** illustrates how social media market segmentation technologies like personalisation, customisation and ratings wind up pooling users into communities of ideological self-reinforcement. This framework can be useful for understanding how particular conspiracy theories map to particular demographics online.

• **The Social Media Shame Cycle** can be useful for understanding how personal and social stigma is communicated in online contexts, and when opportunities exist for policy makers, platform owners, activists, educators and health practitioners to engage in various forms of intervention, such as dialogue, confrontation, education, and care.

• **The Social Media Risk Matrix** helps explain how young people have learned to incorporate the dynamics described above into self-perceptions of their emotional, economic, and physical risk, online and off. This framework can aid anyone seeking to better understand young people’s tendency to narrate their experiences of trust and distrust through the language of authenticity, referencing both personal embodiment (e.g. authentically feeling a certain way about something or someone), and social mediation (e.g. deliberating the authenticity of others; knowing that our authenticity is likewise being assessed.)

This brief offers actionable insights for addressing the needs of young people in three critical areas during the infodemic: protection, care, and inclusion. These have been allocated according to four stakeholder groups: policy makers, platform owners, members of civil society (usually in the fields of health, education, and/or activism), and researchers (usually in policy, health, media, and/or education.) This brief includes advice on how to:

• Collect and document young people’s experiences with misinformation while protecting their right to inclusion, consent, privacy, and data protection.

• Stay abreast of new technological and social challenges connected to young people by forging alliances between policy makers and platforms.
• Boost the impact of relevant, local, authoritative information by understanding which sorts of messengers young people trust, where and when messages should appear, and which delivery styles young people trust most.

• Foreground young people in the fight against misinformation by tapping into acknowledging their demographic expertise, their tactical mastery of preferred communication styles, and their demonstrated effectiveness in the realm of information activism.

• Assess early warning signs of young people at risk within Social Media Shame Cycles, rather than waiting until they prove harmful to themselves or others.

The policy recommendations in this document are informed by a corpus of research conducted by organisations focused on protecting and empowering young people globally; organisations measuring cultural attitudes during Covid-19; organisations devoted to assessing attitudes towards the press, the internet, and social media; organisations who track social media use among young people for industry purposes; organisations developed to promote health, media, digital and data literacy; as and organisations explicitly committed to fighting misinformation, such as the Shorenstein Center and First Draft. The frameworks discussed in this brief have been reviewed by professionals in the fields of risk communication and media studies, as well as experts in the health field, including the CDC and WHO. Feedback from these reviews has been incorporated into each framework.

In their WHO policy brief addressing media manipulation during Covid-19\(^\text{12}\), the Shorenstein Center introduces a framework for understanding the chronology of misinformation flows called the Media Manipulation Lifecycle.\(^\text{13}\) We endorse the use of this model, as well as the valuable

---


advice offered by First Draft directed toward platform owners, researchers teachers and information activists seeking to combat misinformation. Like the Shorenstein Center, this brief calls for a “whole of society” networked response to issues surrounding the infodemic, where stakeholders in the private sector, public sector, and civil society are able to work toward a common goal – to provide accurate health information. To this we add the following charge: delivering accurate health information and caring for society’s most politically vulnerable members are tasks that ought to go hand in hand. What follows are suggestions for coordinating such an effort with and on behalf of young people across the world.

15 See for example, Ultimate beginner’s guide on how to monitor content on TikTok—First Draft. (n.d.). https://firstdraftnews.org/long-form-article/ultimate-beginners-guide-on-how-to-monitor-content-on-tiktok/
16 We also echo Shorenstein Center’s long-range recommendations to understand how political cleavages, wedge issues, and socioeconomic issues affect the spread of misinformation among people; monitor the community’s perception of trust in media and authorities to understand how best to address them, and continuously call for transparency by technology companies.
This policy brief focuses on the experiences of young people exposed to information disorder during Covid-19.\textsuperscript{17} The WHO characterises the current state of global health communications as an \textit{infodemic}, with too much health information to take in, much of it inaccurate and contradictory.\textsuperscript{18} Overloaded at the cognitive level, and tasked with navigating environments frequently charged with emotional conflict, young people find it harder to assess risks and make rational choices regarding their health practices. Where clarity, education and care should be, the main variants of information disorder--rumour, conspiracies and stigma--flourish.

A recent UNICEF U-Report found that four out of five young people have been exposed to Covid-19-related misinformation,\textsuperscript{19} which at this writing has led to 800 deaths and 5800 hospitalisations globally.\textsuperscript{20} Although misinformation is frequently associated solely with social media use, it need not be. Just as we can post a photo of the ‘real world’ on social media, so too can information move from online contexts to offline ones. For young people, this can happen through conversations with relatives, visits home from school, or requests to translate

\textsuperscript{17} A recent literature review categorised ‘information disorder’ according to six genres and eight agents, each of which affects young people. Genres include: disinformation (deliberately spread); misinformation (unintentionally spread); malinformation (possibly true, but damaging, and generally hyper-partisan); online propaganda (purely promotional); fake news (false and generated to promote profit); clickbait (true but sensationalised); rumours (non-confirmed information); and conspiracy theories (false stories positioned against common knowledge.) Agents of information include politicians; trolls (unpaid individuals, as well as groups working in ‘troll farms’); fake news sites; conspiracy theorists; mainstream media; highly partisan news outlets; foreign governments; and bots (generally working to increase followers or likes and/or attack dissident views, but can also be deployed to alter search engine results, etc) Tucker, J., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. SSRN Electronic Journal, \url{https://doi.org/10.2139/ssrn.3144139}

\textsuperscript{18} World Health Organisation(2020) An ad hoc WHO technical consultation managing the COVID-19 infodemic: Call for action 15 September, \url{https://www.who.int/publications-detail-redirect/9789240010314}

\textsuperscript{19} Get this citation from Lis (in her Indonesia proposal)

\textsuperscript{20} Ibid.
for a community.\textsuperscript{21} Especially when engaging with young people who live among migrant communities or in environments like refugee camps, these distinctions matter.\textsuperscript{22}

**Young People: Defining a Demographic**

For the purposes of this brief, young people will be defined as those between the ages of 16-24. When evaluating research on young people, it is important to understand that there are at present no global standards for disaggregating data by age.\textsuperscript{23} This has led to multiple (and sometimes contradictory) understandings of who young people are, and how to study them.\textsuperscript{24} Figure 3 below compares policy and industry approaches to categorising young people, which tends to name ‘generations’ according to year of birth. For industry researchers, ‘young people’ covers Generation Z, and sometimes, Generation Alpha.\textsuperscript{25}

The behaviour of young people is frequently regulated at national and statutory levels, where guiding terms are: age of protection; statutory age; age of majority; and purchasing age. When discussing how young people produce, consume and circulate disinformation and misinformation, these distinctions matter. Consider, for instance, the rise of teenagers spreading disinformation while working for so-called ‘troll farms.’\textsuperscript{26} While there are a number of laws


\textsuperscript{24} For example, the WHO currently defines young people as those between the ages 10-24, with two overlapping cohorts: adolescents (10-19 years) and youth (15-24). By comparison, the Johns Hopkins Global Covid-19 Attitudes Report only segments its responses by those aged older and under 31 years. See KAP COVID. Johns Hopkins Center for Communication Programs. Retrieved October 31, 2020, from https://ccp.jhu.edu/kap-covid/


\textsuperscript{26} For an example of young people spreading disinformation, see Turning Point Action enlists teenagers in “troll farm” spreading coordinated pro-Trump messaging—The Washington Post. https://www.washingtonpost.com/politics/turning-point-teens-disinformation-trump/2020/09/15/c84091ae-f20a-11ea-b796-2dd09962649c_story.html
around the world designed to punish such activities, \(^{27}\) to be subject to such laws, a young person needs to be over the age of legal protection. Depending on the jurisdiction in which a young person lives (which may be different to where they circulate disinformation or misinformation) this can be as young as ten years, or as old as fifteen.

Equally important in the study and fight against misinformation is statutory age: the point at which a young person can consent without a parent to medical procedures (including vaccine inoculation) and/or activities like participating in research studies. Statutory age ranges by activity and jurisdiction, and is normally between 14-18 years of age. A final categorisation affecting young people is purchasing age, which can range from 16-21 years of age, depending upon jurisdiction and product.

Appendix 4 illustrates how these regulations overlap with platform approaches to young users. On most social media platforms, a user can join at thirteen years old. On social media platforms featuring sexually explicit content, a user must be eighteen. However, young people are under the age of eighteen are frequently subject to additional content restrictions, \(^{28}\) and on platforms where young users dominate, some activities that might be harmful to young people are banned for all. \(^{29}\)

Of course, young people identify as much more than their age. Identity is an intersectional phenomenon, and especially when it comes to gender, race, religion, or outsider status, young people are as vulnerable as anyone else to stigma. \(^{30}\) An increasingly important way young people distinguish themselves is through taste. This is done by signalling a preference for particular consumer goods and brands or for entertainment practices and formats that signal


\(^{28}\) TikTok’s ban on weight loss ads directed to users under 18. Get citation

\(^{29}\) Instagram’s ban on posing with weapons. Get citation

adherence to (or at times, divergence from) group norms. Social media platforms encourage and track such “taste performances”, organising users online according to their preferences, which are later marketed as evidence of consumer sentiment.

**Young People: Experiences and Attitudes During Covid-19**

This section describes the features that appear to typify young people’s experiences of Covid-19, globally. These include a belief in personal health exceptionalism; high levels of economic precarity, high levels of emotional distress; high levels of vulnerability for domestic abuse; and high levels of vulnerability for targeting by extremist organizations.

**Health exceptionalism.** Although infection rates continue to rapidly rise among young people (especially among those in their twenties), many maintain they aren’t at high risk for Covid-19. This belief in “health exceptionalism” appears to have originated from young people’s early experiences in the pandemic, when they were told they were not at high risk for contracting the virus and were sent to work and to school.

**High economic and educational precarity.** Young people around the world report high levels of concern over economic and education-related issues, with greater concern across migrant

---


34 According to Johns Hopkins Medical Center, 44% of Americans and 41% of Indonesian respondents under 31 believe they are at lower risk than others for infection. See John Hopkins Medical Center (2020) KAP COVID Study. July https://ccp.jhu.edu/kap-covid/

35 According to a recent UNICEF poll, one-third of youth in Latin America and the Caribbean believe they are not at risk for the disease. One third of youth in Latin America and the Caribbean believe they are not at risk from disease, new UNICEF poll shows.”


37 For more on this process, see McStay, A. (2018). *Emotional AI: The Rise of Empathic Media*. SAGE.


39 According to Johns Hopkins Medical Center, 44% of Americans and 41% of Indonesian respondents under 31 believe they are at lower risk than others for infection. See John Hopkins Medical Center (2020) KAP COVID Study. July https://ccp.jhu.edu/kap-covid/

40 According to a recent UNICEF poll, one-third of youth in Latin America and the Caribbean believe they are not at risk for the disease. One third of youth in Latin America and the Caribbean believe they are not at risk from disease, new UNICEF poll shows.”


43 For more on this process, see McStay, A. (2018). *Emotional AI: The Rise of Empathic Media*. SAGE.


47 For more on this process, see McStay, A. (2018). *Emotional AI: The Rise of Empathic Media*. SAGE.
populations. In wealthier countries, young people face a paradox: forced to work in insecure ‘essential worker’ roles and take classes on campus, they are frequently shamed by mainstream media as ‘partying super spreaders’ while at universities, public parks and beaches or gathering for activist causes like Black Lives Matter.

High levels of concern for mental health, and vulnerability for domestic abuse and violence. Young people are reporting higher than average levels of concern for their mental health during the pandemic. In a recent study of young Australians, many reported that they don’t seek mental health assistance because they felt it was not as important as everything else going on. Another important issue for young people globally is linked to marked rises in at-home harassment and physical abuse experienced during the pandemic.

High targeting for political extremist grooming. Covid-19 lockdowns have ushered in a wave of online extremist activities across the political spectrum including efforts by movements like

The OECD reports that prior to the pandemic, young people were 2.5 times more likely to be unemployed than people aged 25-64, one in eight young people lived in poverty. During the pandemic, those who can find employment are often doing temporary or part-time work, and can easily slip below the poverty line within 3 months, should their income suddenly stop or decline, affecting their access to housing, ability to pay back school loans, etc.

39 Blamed for ignoring the coronavirus warnings, young people say it’s not all “quarantine parties.” (2020, March 23).
the Alt-Right, who already have a long and documented history of grooming and recruitment efforts waged over young people's gaming platforms.\textsuperscript{44}

**Young People: Ongoing Perceptions & Preferences**

This section describes attitudes that have characterised young people prior to the pandemic, and continue to shape their attitudes and perceptions of the current moment. These include: low levels of institutional trust; a preference for messengers over messages; a tendency to resonate w more strongly with feelings more than facts.

**Low levels of trust.** As a cohort, young people report substantially lower levels of trust than the general population in both government institutions and social media platforms. Less than half of young people (45\%) across the OECD countries express trust in government;\textsuperscript{45} further decreased among marginalised communities.\textsuperscript{46} In a recent Axios survey, 83\% of young people identified social media as their primary source of news, with 93\% reporting they found news over social media to be untrustworthy.\textsuperscript{47}

**Tendency to trust messengers first, and messaging second.** Industry studies show young people are increasingly turning to influencers, celebrities, spokespeople and brands as sources

https://doi.org/10.1177/2056305118768302

\textsuperscript{45} OECD (2020) Youth and COVID-19: Response, recovery and resilience.  


\textsuperscript{47} Kight, S. W. (2020.). *Gen Z is eroding the power of misinformation*. Axios. from  
https://www.axios.com/gen-z-is-eroding-the-power-of-misinformation-5940e3cd-e3d0-44a1-b66c-93be45fe1d2c.html According to a recent Reuters report, young people do see traditional media as more credible, yet avoid stories there, finding them too downbeat, narrow, repetitive, and driven by nationalist biases for their tastes. See *How Young People Consume News and the Implications for Mainstream Media*. (n.d.). Reuters Institute for the Study of Journalism.  
https://reutersinstitute.politics.ox.ac.uk/our-research/how-young-people-consume-news-and-implications-mainstream-media
of all sorts of information, including health information. While the propensity to gather ‘trusted influencers’ is most frequently remarked upon in discussions of young women, extremist groups like the Alt Right have long deployed influencer tactics in their communications with young men. A recent study by Axios showed that young people were far more adept at identifying sponsored and influencer-driven material than their older counterparts. Argues Jonathon Morgan, CEO of AI software company Yonder “Most older generations — even millennials — don’t always understand online influencer culture,” and frequently mistake influencer communications as disinformation or misinformation.

High preference for narratives focused on authenticity of feeling. Over and again, industry research shows that young consumers prefer messages delivered through feeling-driven narratives built on claims to experience and authenticity, rather than purely evidence-driven arguments built on presentations of facts. One upshot of this appears to be a predisposition to conspiracy theories: monological belief systems in which any new information that threatens existing beliefs is either neglected, or inserted in existing beliefs. Across all social media platforms, Covid-19 conspiracy theories appear woven into projects with larger political narratives, such as the rise of conspiratorial group QAnon in the United States and Europe. In

---


50 See Marwick, A., & Lewis, R. (2018). Media Manipulation and Disinformation Online (p. 106). Data & Society. [https://doi.org/10.2196/16540](https://doi.org/10.2196/16540)


a recent study conducted by King’s College London, young people displayed the highest numbers of belief for every major Covid-19 conspiracy, regardless of education level. In Australia, researchers found the majority of Covid-19 denialists were not just young, but white and male.

**New challenges: young people’s social media use patterns**

Social media use has risen among all age groups during Covid-19, but industry research shows that young people use social media differently than their older counterparts. This section describes some of the distinct features of young people’s social media use.

**Platform Preferences.** Although young people continue to crave news and education content, they prefer platforms that highlight collaborative entertainment activities, such as Tik Tok, which allows for video creation and sharing, and Twitch, which frequently features users who live-stream themselves playing video games. Among social media users as a whole, the top social media platforms are Facebook, YouTube and WhatsApp. By comparison, Gen Z consumers worldwide prefer Snapchat, TikTok, Discord and Instagram. Even when it comes to news gathering, young people favor entertainment platforms over more traditional venues like Facebook and Twitter.

---


58 More Young Teens Use TikTok Than Facebook | Morning Consult. (n.d.). [https://morningconsult.com/form/more-young-teens-use-tiktok-than-facebook/](https://morningconsult.com/form/more-young-teens-use-tiktok-than-facebook/)

59 Online video watching jumped 50% in this cohort during the pandemic. See


When it comes to misinformation, these preferences pose two major challenges. At the technological level, the content of these platforms is often co-created, and/or embedded at screen level, which often makes it difficult to source the origins of circulating material.\textsuperscript{64} At the regulatory level, the “just for fun” tone of these platforms lends itself to lax policies around misinformation. As reporter Daniel Kelley notes, although gaming platforms appear to be growing in popularity as venues for political activities, no gaming platform has issued formal guidelines regarding the spread of health misinformation on their platforms.\textsuperscript{65}

**Communication Practices.** A much-remarked upon feature of young people’s social media use is their preference for communication through images. Emojis have taken on new meanings during the pandemic\textsuperscript{66}, and selfies continue to remain a staple of personal expression and civic discussion\textsuperscript{67}, along with other photo material like reaction GIFs and memes\textsuperscript{68}. Because visual communication tends to signal differently to different communities of users,\textsuperscript{69} its meaning is always in some way local,\textsuperscript{70} yet the challenges posed by visual misinformation are global in

\textsuperscript{64} There are a number of conversations circulating about solving this problem via digital watermarking and other technological strategies. See Gugelmann, D., Sommer, D., Lenders, V., Happe, M., & Vanbever, L. (2018). Screen watermarking for data theft investigation and attribution. 2018 10th International Conference on Cyber Conflict (CyCon), 391–408. https://doi.org/10.23919/CYCON.2018.8405027


\textsuperscript{68} For a discussion of selfies as civic witnessing and activist display, see Kuntsman, A. (2017). Selfie Citizenship. Springer. For an example of government-mandated selfies during Covid-19, see Ieva was on the phone to the ABC when the police showed up. (2020, April 24).

https://doi.org/10.1080/10584609.2019.1674979

\textsuperscript{70} For a provocative case study, see Images and misinformation in political groups: Evidence from WhatsApp in India | HKS Misinformation Review. (7 July 2020.).
scope, argues First Draft researcher Hannah Guy, pointing out that “Graphs and charts—with their uniform, scientific format...can give a false sense of authority.”71 A recent study of visual misinformation circulating during Covid-19 confirms this. In addition to using visuals to provide false evidence for claims (“helping to establish objects as specific as a diseased chicken or as diffuse as a global conspiracy”) visual branding that appeared to represent reputable institutions was frequently used to provide “authoritative agency” for claims. 72

Of all forms of visual communication, memes pose special challenges to those fighting misinformation. The individual and institutional73 circulation of memes74 has been particularly active during this time: memes have been served a range of purposes during the pandemic. While memes like “flatten the curve” deliver necessary health messaging in entertaining formats, other sorts of Covid-19 related memes have been created for emotional coping, 75 public shaming (discussed later in this brief), and exacerbating political polarization.76

Limor Shiffman defines memes as visual artifacts that “(a) share common characteristics of content, form, and/or stance; (b) are created with awareness of each other; and (c) are circulated, imitated, and transformed via the Internet by multiple users.”77 The figure below represents how memes are frequently dominated by ‘in group’ practices, and images are

frequently altered to suit group needs, making it difficult to assess meaning from an outsider’s perspective. An image created by one group as a misinformation parody can easily circulate among others as accurate information.

Politically speaking, memes present two challenges. Because they tend to be created and circulate without it is relatively easy to amass groups of memes in what Joan Donovan dubs “evidence collages” that seem to represent a groundswell of opinion, but actually originate from a few individuals. The second challenge is summed up by Donovan, who writes, “Meme wars are tilted in favor of those willing to use negative stereotypes and cruelty to tap into broader audiences. Memes that invoke racist, sexist, LGBTQ, and religious intolerance resonate with a broad swath of internet users, particularly white men.”

---


Protocols. Even more than the general population, the online lives of young people are heavily influenced by mechanisms like rankings, feeds, and trend—the result of algorithms that have been tracking them for almost a decade now. Although young people demonstrate higher than average understanding of algorithmic dynamics, it is unclear how much of this knowledge translates to autonomy, power or control. This matters, because platform dynamics pose particular challenges for those attempting to dislodge misinformation that ‘pools’ in communities of ideological self-reinforcement, such as those involving stigma and/or conspiracy theories.

Procedures. As a cohort, young people display a conflicting awareness about their lives online. On the one hand, they understand that the realities of surveillance capitalism require them to curate their identities online in a way that puts their best public face forward. On the other hand, young people rely on social media for activities like learning, friendship and building political solidarity—all of which require a degree of personal vulnerability. To manage, they have developed psychosocial procedures for assessing personal risk online that often turns on a loathing for and fascination with online ‘drama.’ Regan and Sweet identified three related functions of online drama among teen girls: it theatricalises social/relational aggression; it allows self-regulation via social surveillance; and it works as entertainment. To reach young people,
it is critical to understand the significance of each of these functions within their communities.

Frameworks for Understanding

The recommendations in this brief are grounded in four frameworks. Each provides researchers, educators, policy makers, and practitioners interlocking perspectives on how young people receive information online; how they find themselves within separate ‘pools’ on the same platforms; how their social regulation processes mimic social media logics; and how young people deploy concepts like authenticity as code terms to describe their perceptions of environmental risk. These frameworks incorporate and extend existing work in the social and behavioural sciences in the areas of threat perception and navigation; socio-technical influences on human behaviour; psycho-social processes of decision-making, and responses to emotional stress. 89

The Influence Ecosphere

The influence ecosphere framework helps explain how and why information with a high emotional charge seems to dominate social media flows. It can be especially useful for understanding dynamics like the spread of viral rumours online, such as the recent emergence of the #FilmYourHospital conspiracy from a single tweet. 90 Illustrated in the image below, this framework combines recent research in the fields of emotional communication, persuasive computing, and internet virality. 91


It is common to hear even the most seasoned social media users describe their online social interactions as if they were organic processes. In truth, all exchanges over platforms are the result of highly controlled operations, designed to amplify ‘emotions on the move,’ with the objective of keeping users online as long as possible. The Influence Ecosphere framework maps these operations by way of the following elements: social media actors; automated events; psychological states; interface triggering; social and platform drivers (including algorithmic tracking and platform control); social discourses about social media at large; and specific social media practices.

Figure: Influence Ecosphere

Social media actors include any ‘players’ engaged in exchanges online, be they individuals, institutions, humans, or bots. Mediated events refer to moments (personal, social, political, etc.) in the lives of users, ideally ones that can be automated. The most common of these are the birthday announcements; the most frequent are notification alerts that let us know others have
reacted to our posts, or updated their own. Events online and off trigger psychological states in users ranging from positive, to negative, to liminal. In a recent analysis made by Brandwatch of Covid-19 posts around the world, the dominant psychological state appeared to be disgust, and the second most was fear.  

In his review of the role of ‘persuasive technology’ in the history of platform design, Christian Stocker likens contemporary interface triggers to psychologist BF Skinner’s boxes. “Instead of lights, speakers, levers, food pellets and electric shocks, the user interface designers of today employ multimedia content, like, share and reaction buttons, likes, reactions and comments as stimuli, actions and rewards. Although a substantial amount of research has been done linking what Megan Boler terms the ‘affective feedback loop’ to the release of ‘happiness chemicals’ like dopamine levels in the brain, Boler points out that emotions like outrage also figure in reward systems. For social media platforms, she points out, what makes money isn’t what keeps users happy, but what keeps “emotions on the move,” compelling users to spend more time online.

Offline, emotions moved through populations by way of social drivers like gossip and social contagion. Online, these dynamics combine with those of technological drivers, such as the Social Media Shame Cycle, discussed later in this brief. Psychologists have observed that in general, people need group identity-based motivations to share content, noting that when

---


content is emotionally fueled, it captures the attention of others. Platforms follow these exchanges by way of algorithmic tracking, noting properties like speed, frequency, user location, and so forth. To amplify these emotions on the move, platforms provide technological affordances, such as “like” buttons, upvote mechanisms, monetisation opportunities, and so forth. To keep attention trained on some areas, others may need dampening: on platforms, this is done via technological constraints that ‘drip’ information, reduce the frequency of users on one’s feed, and so forth. From the user’s point of view these processes are felt before they are understood or even visible. When our emotional experiences with these processes are positive, we express that through discourses about social media as a whole, using terms like promoting, donating and even protesting. When our emotional experiences are negative, terms like drama and trolling and bullying appear. Social media practices are mapped to these discourses, which is why they run the cultural gamut, from practices generally accepted (for instance, using emojis) to things heavily contested (overly edited images, something that feels like spam, etc.)

**The Ideological Communication Circuit**

The Ideological Communication Circuit framework can be useful for understanding how conspiracy theories map and other hard to dislodge beliefs begin to map to particular demographics online. By invitation and default, user experiences online are increasingly shaped by processes like customisation, personalisation, and ratings. Pitched as time-saving ways to “optimise experience,” an alarming byproduct of these market segmentation practices has been the pooling of users into communities that put them at risk for ideological self-reinforcement. This framework combines Young and Bleakly’s research on “ideological health spirals,” Axel Bruns’s thoughts on ‘filter bubbles,’ and Klein et. al.’s observations regarding conspiracy narratives online.

---

97 Brady, W. J., Crockett, M., & Bavel, J. J. V. (2019). The MAD Model of Moral Contagion: The role of motivation, attention and design in the spread of moralized content online. PsyArXiv. [https://doi.org/10.31234/osf.io/pz9q6](https://doi.org/10.31234/osf.io/pz9q6)

98 Ibid


The Ideological Communication Circuit presents social media information as a flow, moving through three sorts of user pools, online. Psycho-social pooling is characterised by thoughts about personal identity and taste preferences. Statements along these lines include: “These are things I like (or don’t like)”; “This is what I find normal (or weird)”; “This is what I find funny (or cruel)”; “These are (or are not) my values.” Socio-political pooling is characterised by thoughts about social identity. Statements along these lines include: “I feel included here (and sometimes, an outsider elsewhere)”; “I act in a way that is consistent with my group (and sometimes, distrust group members when they act otherwise); “I can distinguish my group from others” (and “I know which are the most disliked ‘out-groups’ by those in my pool.”)

Techno-cultural pooling is characterised by thoughts about social media preferences and informational interests, such as: “This is the platform/channel/news source everyone uses”; “This topic is something I should (or don’t need to) know about, and “In my group, nobody follows (or supports, or believes) this messaging (or this messenger.”

Figure: Ideological Communication Circuit Framework
This framework relies on the metaphor of pools, rather than the common "echo chamber" or "filter bubble," reflecting Axel Bruns’s observation that contrary to common wisdom, research shows that most users have actually “positively increased the diversity of their information diet,” which has in turn “prevented them from becoming locked into ideological monocultures.” For Bruns, the echo chambers/filter bubbles discussion has morphed into a moral panic that obscures a question that urgently needs addressing; namely, “Why and how have different groups in society come to develop such highly divergent personal readings of the same information?”

Polarisation occurs in two forms. Of these, attitudinal polarisation (differing stands on issues) is far easier to dislodge than affective polarisation (liking or trusting an opposing party.) In a moment where political decisions are increasingly marketed as taste choices, affective polarisation seems the order of the day on platforms like Tik Tok, where traditional political communities like Trump Tik Tok sit side by side with publics generated by algorithms tracking the sort of content users seem to like online. When a young person says, “You don’t find Alt-Tik Tok, Alt-Tik Tok finds you,” this is the dynamic they are referencing. As the photo evidence below suggests, Alt and straight on Tik Tok don’t signal sexual identifications, but ‘alternative’ and mainstream consumption tastes. The image below includes a joke that “Soon political parties won’t exist in America, and you will just have to identify as either Straight or Alt Tik-Tok.” While many members of communities like Alt Tik Tok do appear to lean toward progressive politics, transferring boilerplate ideas about ideologically linked communities to platforms like Tik Tok fails to explain why many Alt Tik Tokers are also part of Brand Tik Tok, where users devote their time to personifying brands.

---

102 Ibid
To begin to answer these questions, the Ideological Communication Circuit conceptualises publics as pools made up of information that can stagnate if interventions are not made to not periodically freshen its flows. These interventions might be at the user level (e.g. making deliberate and new curatorial choices in order to vary information flow\textsuperscript{105}); at the social level (sharing and discussing new information with one’s group); and/or at the platform level (elevating or dampening the flow of certain types of information to certain parties by increasing or limiting access in some way.)

At each pool depicted, this framework displays three challenges to keeping information flowing, and polarisation from sedimenting. The first of these occurs at the psycho-social level, when feeling wrong about one’s understanding of information can sometimes translate to feeling wronged as an individual. When this occurs, Young points out, people resort to ego reinforcement behaviours.\textsuperscript{106} One of these is denialism. Another is the Dunning Kruger effect (DKE): a refusal to relinquish belief in one’s knowledge, expertise or ability, in spite of mounting evidence to the contrary. The DKE may provide a plausible explanation for why young people continue to insist they are less prone to misinformation than older generations.

\textsuperscript{105} Naderer, B., Heiss, R., & Matthes, J. (2020). The skilled and the interested: How personal curation skills increase or decrease exposure to political information on social media. \textit{Journal of Information Technology & Politics}, 1–9. https://doi.org/10.1080/19331681.2020.1742843

A socio-political inversion of the DKE can be found in the so-called “Imposter Syndrome,” in which someone who can offer new knowledge or expertise feels insufficiently prepared to do so, or else feels they lack ‘real’ authority within a particular context. Imposter syndrome can arise as a result of the second challenge listed in the framework: social sorting. The final challenge listed on the framework, informational fragmentation reflects the technological concerns voiced by those who raise the “filter bubble” argument. For full polarisation to occur, however, this technological realm must combine with psycho-social and socio-political stagnation.

The Social Media Shame Cycle

This framework has been designed to illustrate a pattern of social engagement ‘drama’ that increasingly characterises all discussions with young people about information trustworthiness, whether these discussions happen online, or off it. Combining Crystal Abidin’s work on “influencer drama cycles”\textsuperscript{107} and behavioural science research on the operation of shame and blame for social cohesion, the Social Media Shame Cycle can also be a useful way to understand how the communication of stigma works online.

The Cycle involves four concurrent stages: shame, blame, mediation and remediation. In public consciousness, the cycles are most recognisable when they appear as actions, or call to actions such as public ‘call outs’ of individuals and organisations perceived to have transgressed the common good; large-scale “piling on” to protest a wrongdoing or a wrongdoer; demands for public apologies staged over social media (with accompanying assessments of their authentic sincerity); and continued fascination with the dynamics of so-called ‘cancel culture.’

\textsuperscript{107} Get citation from Crystal Abidin (new work 2020)
Shame: The feeling of shame is best described as a self-conscious awareness that one is somehow out of sync with the norms of a particular group or environment. Shame may or may not incorporate feelings of guilt, which center on action (“I feel guilty for doing that”), whereas shame centers on identity (“I feel ashamed of myself.”) Two recent examples of institutional attempts to induce shame during Covid-19 are the Indonesian government’s requirement that those breaking quarantine publicly dig graves (and have their photos spread across social media), and the publication of the names (and photos) of two young Australians declared “Enemies of the State” for violating travel bans. As mentioned earlier, a great deal of Covid-19 related shame relating to young people concerns their financial, educational and emotional precarity, as well as the media’s characterisation of them as self-interested super spreaders. For young people especially, shame is commonly voiced as a spectrum of feeling, with awkwardness and embarrassment on the milder end, and humiliation at the most extreme point.

Blame: In his historical overview of shame, Stearns argues that while guilt and shame have long been used to govern societies over the course of history, the two strategies tend to deliver different psycho-social outcomes. When someone feels guilt, there generally follows a desire for forgiveness. By comparison, people who feel shame first freeze, and then attempt to shift their state by fleeing (self-isolation, avoidance), fighting (getting angry or aggressive), or both of these. In the Blame part of the cycle, an individual attempts to lessen the emotional pressure of shame by condemning (consciously or not) a person, group, object and/or situation. Condemnation can include oneself. Entry into the blame segment of the cycle is most readily signalled through language that turns the speaker from subject to object, e.g. from “I was embarrassed” to “they embarrassed me.” The milder end of the blame spectrum can be seen in words like “annoys” or “frustrates”; the more extreme end includes words like “infuriates” and “enrages.”

During Covid-19, “Life has been difficult, individuals and families have lost their jobs or lost income, people’s lives have been constrained into small bubbles. All of this has reinforced anxiety, anger and a desire to find someone to blame. To a substantial degree, these dynamics mimic those at regional, national and international levels. Many statements that Covid-19 was developed as a bioweapon, for example, have come from official sources. Among extremists, early denials of the pandemic appear to have morphed into a "particular focus on racial groups and promotion of antisemitic and xenophobic positions," with Covid-19 providing the “perfect opportunity for extremists to groom people, through the use of conspiracy theories but also through a presence on social media.”

Mediation: in this part of the cycle, blame moves from language to action, normally with efforts to publicise, retaliate, or otherwise punish a blamed party. ‘Blaming and shaming’ practices range from mild (resentment, gossip) to extreme (scapegoating, stigmatising.) As debates on “mask shaming” in the United States make plain, efforts to blame and shame have seen mixed

---

111 https://www.governmentevents.co.uk/covid-19-conspiracy-theories-and-extremist-grooming/  
113 https://gnet-research.org/2020/05/18/call-to-arms-social-media-and-far-right-narratives-of-covid-19/  
114 https://www.governmentevents.co.uk/covid-19-conspiracy-theories-and-extremist-grooming/
results during Covid-19, particularly when such efforts move online\textsuperscript{115}. Consider the triptych featured below. The left-hand side features what might be called a pro-mask meme, designed to shame anti-maskers. The middle image features an anti-mask meme, designed to shame pro-maskers. The right-hand side features an anti-mask meme (depicting pro-maskers as sheep) that has been marked with pro-mask comment, “This anti-mask meme is actually very accurate insofar as that lion is going to kill everything else in the picture.”

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{mask_shaming.png}
\caption{Covid-19 and Masking Shame}
\end{figure}

**Remediation:** this part of the cycle account is best understood in terms of reactive responses to blaming and shaming. These can range from actions signalling low interest in conflict (apology, restitution, reconciliation) to high interest (censorship; defensiveness; counter-attack). Between these poles are non-actions, such as ignoring blame, or evading it. With the proximity of vaccines to fight Covid-19 on the horizon, it is worth considering in advance how current efforts to shame any and all “anti-vax” responses may well backfire into well-orchestrated counter-attacks.

Intervention. Within and between each stage of the Social Shame Cycle, opportunities exist for policy makers, platform owners, activists, educators and health practitioners to engage in various forms of intervention. Understood here as action intended to prevent or alter the course of events, interaction includes methods such as dialogue, confrontation, education, and care. In general, these strategies tend to be most effective during early stages of emotional arousal: awkwardness is easier to recover from than humiliation; irritation is more easily assuaged than rage; low-level resentment more easily redirected than entrenched cultural stigmatising. It should also go without saying that intervention strategies should always be assessed for contextual appropriateness. Just as confrontation strategies of police differ from those of health practitioners, so too should dialogue strategies for teachers and students differ from those used by close friends of similar ages and established bonds of trust.

Social Media Risk Matrix

This framework provides a way to understand how young people have learned to incorporate the dynamics described above into self-perceptions of risk in online environments. Combining insights from the fields of public health, management science, and youth culture research, it characterises young people’s experiences online by way of the following elements:

- **Risk**, situated between poles of **esteem, trust, shame and harm**
- **Esteem**, amplified via online narratives of **empowerment**
- **Trust**, amplified via online dynamics of **promotion**
- **Shame**, amplified via online narratives of **redemption**
- **Harm**, amplified via online dynamics of **controversy**
- **Authenticity**, expressed as personal feeling and social discourse
Risk. “Popular and professional ideas about risk often differ,” observes epidemiologist James Trostle.\textsuperscript{116} Health communications expert Gaya Gamhewage adds, “For experts, the risk is great when the hazard is great.” By contrast, the public’s perception of risk is great “when their sense of emotional engagement is great - fear, anger, outrage.” \textsuperscript{117} This relationship between actual and perceived risk is worth exploring in greater detail with regard to young people.

In their efforts to designate ‘at risk’ populations for research purposes, The Belmont report\textsuperscript{118} identifies four categories of risk: physical (possibility of bodily harm); psychological (adversely affecting the perception of self, emotional suffering); social (possibility of social stigmatisation) and economic (bearing financial costs.) In research settings, young people under the age of 18 years old are at particular risk of social stigmatisation, emotional suffering, and financial costs due to their lack of experience and maturity. The Belmont report recommends that researchers consider these factors when selecting research participants and designing interventions.

eighteen are considered *de facto* at risk for exploitation by adults. Young people over eighteen may be likewise deemed vulnerable if they have difficulty granting informed consent due to cognitive disabilities, or as a result of living in an environment where physical or emotional coercion exists, as in most domestic violence settings. All people are considered at risk if they live in ‘difficult environments’ where daily life is unpredictable and permeated with fear, such as conflict zones, politically repressive environments, authoritarian regimes, and/or areas where ‘terrorist’ groups predominate.\(^{119}\)

Regardless of their perceptions of health exceptionalism, young people have already found themselves on the frontlines of risk for speaking out about health misinformation. in locations like the Philippines\(^{120}\), Vietnam\(^{121}\) and Nepal,\(^{122}\) they have met government censorship, pressure and pushback. In the U.S., American teenager Ethan Lindenberger (who frequently shares that he turned to Reddit for answers when he first began questioning his parent’s resistance to vaccines) has been repeatedly targeted by extremist factions for his pro-vaccine activist work.\(^{123}\)

Whatever their political stance, all young people are at higher risk than the general population for general harassment and bullying online, with girls particularly vulnerable.\(^{124}\)

**Esteem.** Trostle notes that abstract arguments about the benefits of changing a behaviour will always be less effective with patients than laying out “the real risk a disease would pose to their plans and dreams.”\(^{125}\) In social media spaces, positive reinforcement for one’s ‘plans and dreams’ by way of the Influence Ecosphere where every user must undertake a mental calculus to determine whether risks (of potential exposure, censure, ridicule, harm) are worth the rewards

---


\(^{120}\) Duterte’s troll armies drown out Covid-19 dissent in the Philippines—Coda Story. [https://www.codastory.com/disinformation/information-war/philippines-troll-armies/](https://www.codastory.com/disinformation/information-war/philippines-troll-armies/)


(of appeal, connections, friendships, inclusion, monetary rewards, etc.) In persuasive computing circles, this calculus has been codified as the The Fogg Behavioural Model (FBM) consisting of three tiers: personal motivations; capacities for action (shaped by factors like time, money, physical effort, mental effort, and routine), and prompts to trigger engagement.\footnote{Agha, S., Tollefson, D., Paul, S., Green, D., & Babigumira, J. B. (2019). Use of the Fogg Behavior Model to Assess the Impact of a Social Marketing Campaign on Condom Use in Pakistan. \textit{Journal of Health Communication}, 24(3), 284–292. \url{https://doi.org/10.1080/10810730.2019.1597952}}

**Trust.** Trust can be understood as “the willingness to increase one’s vulnerability to another person whose behavior is not under one’s control,” writes management scholar Dale Zand.\footnote{Zand, D. (n.d.). \textit{Trust and Managerial Problem Solving on JSTOR}. Retrieved November 20, 2020, from \url{https://www.jstor.org/stable/2393957?origin=crossref}} For young people, willingness tends to be assessed using the language of authenticity. A common way young people signal an erosion of trust is by moving from discussions that frame authenticity as an experience felt in the body (e.g. “I really like you”), to deliberations that frame it as a performance that can and should be collectively evaluated (e.g. “they don’t seem sincere.”) While engaged in the Influence Ecosphere, these shifting understandings of authenticity are often mapped onto discussions of what ‘really’ goes on in social media environments, which can sometimes be a young person’s way of indicating their personal engagement with the Social Media Shame Cycle.

**Shame.** Although shame was discussed at length earlier, it is worth noting which individuals and groups are most at risk for shame responses. From a psychological perspective, some people’s personalities are just more shame prone (whereas others are more guilt prone.)\footnote{Stearns, P. N. (2017). \textit{Shame: A Brief History}. University of Illinois Press. \url{https://doi.org/10.5406/j.ctt1vjqrq8}} From a developmental perspective, individuals or groups who have current or prior experiences with trauma and/or stigma tend to be more prone to shame. From a sociological perspective, socially dominant groups are at far lower risk for shame than subdominant groups.

**Harm.** When algorithmically enhanced controversies burn too hot online, actual harm can emerge offline. When successful interventions are made to mitigate these, trust levels might be re-evaluated as transformed. From a health communications perspective, these interventions map to general principles of risk management: raising awareness; encouraging protective...
behaviour; reducing anxiety and outrage; improving relationships through dialogue, cooperation and network-building; and involving people in decision making processes.¹²⁹

Fighting the Infodemic: Strategies to Date

This section describes the most current strategies for fighting information disorder, evaluating their strengths and weaknesses vis a vis young people. Throughout, examples of best practice are provided of efforts targeted at young people, and led by them. Social responses include debunking; pre-bunking; information activism; and influencing. Institutional responses include education; legislation; regulation; and nudging.

Debunking. The classic approach to correcting misinformation has been to concentrate on providing verifiable facts, and countering objections with evidence-driven arguments. This tactic is commonly known as debunking.¹³⁰ For some, timing is a critical element in the debunking process: Claire Wardle of First Draft suggests misinformation that is corrected too early may receive “unintended oxygen,” allowing it to spread further.¹³¹ However, insists Timothy Caulfield, “the evidence is far from definitive, and what evidence is available suggests it doesn’t often happen.”¹³² At least as important is how to debunk: UNICEF advises addressing the underlying cause of misinformation; communicating in ways that focus on solutions; emphasizing self efficacy, hope and agency; and promoting trust through dialogue and questions.”¹³³

¹³⁰ Caulfield, Timothy. (2020). CHAPTER B-4: Does Debunking Work? Correcting COVID-19 Misinformation on Social Media. In Vulnerable: The Law, Policy and Ethics of COVID-19. CanLII. https://commentary.canlii.org/w/canlii/2020CanLIDocs1866#fragment/zoupio-_Tocpdf_bk_22/BQCwhgziBcwMYqK4DsDWslQewE4BUBTADwBdoAvbRABwEtsBaAfXx2zhoBMAzZql1TMATEICUAAGmTZShCAEVfXAE9oAcjIihMLgQKlqjVp16QAZTtkAQAoBKAUQAyDgDBBAHIPhB+NJgfnCk7KIIQA
¹³² Caulfield, Timothy. (2020). CHAPTER B-4: Does Debunking Work? Correcting COVID-19 Misinformation on Social Media. In Vulnerable: The Law, Policy and Ethics of COVID-19. CanLII. https://commentary.canlii.org/w/canlii/2020CanLIDocs1866#fragment/zoupio-_Tocpdf_bk_22/BQCwhgziBcwMYqK4DsDWslQewE4BUBTADwBdoAvbRABwEtsBaAfXx2zhoBMAzZql1TMATEICUAAGmTZShCAEVfXAE9oAcjIihMLgQKlqjVp16QAZTtkAQAoBKAUQAyDgDBBAHIPhB+NJgfnCk7KIIQA
When directed at young people, debunking strategies seem to work most effectively when they emphasise shared humanity (e.g. an official message that features a representative speaking face-to-camera); allow for anonymous use (e.g. a chatbot that can be consulted privately); and/or involve young people as active agents in fact-checking, reporting and general community efforts. Debunking strategies appear to work less effectively when they are deployed in online environments where users are focused on entertainment, where attempts to correct can be perceived as off topic, combative and/or shaming.

**Pre-bunking.** To better prepare the public to deal with the emotional side of misinformation, communication specialists advocate pre-bunking.\(^{134}\) Also known as social inoculation, this strategy teaches users to identify potentially dangerous arguments and deconstruct them before they gain wider public traction.\(^{135}\) Pre-bunking efforts often take the form of online games and quizzes with names like Spot the Troll\(^{136}\) and Bad News,\(^{137}\) and have been proven to reduce susceptibility to misinformation across cultures.\(^{138}\) There are also a number of young people involved in ersatz pre-bunking operations of their own. Twitch stream Casey Neiham routinely covers techniques of Covid-19 ‘scammers.’ International students have initiated the “I am not a Virus” selfie campaign to make visible ethnic and national stigmatisation during Covid-19. Meme parodies of “questionable health guru” advice demonstrate the aestheticization of medical misinformation on platforms like Instagram.

When directed at young people, pre-bunking efforts seem most effective when they become part of online socialisation processes (e.g. ability to share your score on a game or quiz.) They are perhaps least effective when presented as parodies, which run the risk of misinterpretation,

---


\(^{136}\) Spot the Troll is at http://spotthetroll.org

\(^{137}\) Bad News is at http://getbadnews.com

as happened with a Facebook meme mocking the “peppers protect against the pandemic” rumour, which later circulated on WhatsApp as true information.\(^{139}\)

**Information activism.** Information activism can be used to deliver a unified health message to groups that may be ideologically opposed in other ways. This is the case in southern Thailand, where youth activist group THE MOTIVE have joined forces with Malay Muslim activists the PATHANI to create short videos that are focused on Covid-19 misinformation and hate speech directed toward Muslims. Activism can be used to complement or supplement misinformation monitoring efforts going on at regional or national levels. Indonesian information activists MAFINDO maintain a rich hoax busting database. Activist identities can develop as momentary responses to the pandemic. More Viral than the Virus is an activist movement to fight medical misinformation begun by medical students. Sometimes, opportunities for activism present in guises that borrow from fan culture. An example of this sort can be seen in the recent politically inflected activities of K-pop fans.\(^{140}\)

In general, information activism seems to be most effective with young people when it is perceived as having a grass roots origin on the ground, or else arising as spontaneous response to an event (e.g. K-pop activists organising against alt-right misinformation campaigns on Twitter.) It appears least effective when it is perceived as opportunism (often referenced as ‘clout chasing.’)\(^{141}\)

**Influencing.** In the context of fighting misinformation, influence refers to the power of health messengers to educate, activate, and where necessary, soothe an anxious population. UNICEF emphasises the importance of providing clear communication through trusted and familiar channels, using simple and easily digestible language\(^{142}\) -- a point to underscore, given the

---


inability of many health organisations to meet basic literacy guidelines. Standard guidelines for messaging include: emphasizing benefits to the recipient; focusing on protecting others as well as oneself; aligning with group values; appealing to scientific norms; and/or highlighting social group approval. Given young people’s continued belief in their health exceptionalism, it is worth considering messaging approaches that break through what has been called the “optimism bias,” provided these can be delivered without arousing fear or shame. Among young people, one messaging approach that appears to resonate well is what might be called “you are here with me” message. These generally focus on the messenger’s placement in a real-time location (e.g. a hospital ward); allude to relatable struggles (e.g. the livestream posted after a long day’s work); and/or signal a groundswell of larger public opinion (e.g. the selfie with the messenger holding a sign of some sort.)

As the rise of ‘influencer culture’ demonstrates, young people prefer messages delivered by messengers who emphasise personal presence, in addition to institutional authority. Partnerships between health organisations and online influencers are now common, and rightfully so. As UNICEF points out, “Online influencers often have economic resources and wield socio-cultural influence, dedicate substantial time to online communication, and have the ability and knowledge to influence algorithms. Whether they are working with an institution or own their own, the most effective influencers among young people are those perceived as funny, or speaking from authentic personal experience and/or seem to be ‘one of us.’” As illustrated by the Thai “Spokes-Sheba” below, influencers need not even be human. The least

---


effective influencers are those who seem to be appearing out of institutional duty, or solely for compensation, as when an organisation hires a disinterested celebrity spokesperson to voice their messages.

**Education** One of the most common responses to the rise of misinformation has been a call for better health and media literacy, with a focus on the critical skills needed to de-bunk and pre-bunk low quality information encountered online. For young people, education efforts are most effective when existing media literacy curricula are updated to reflect platform, algorithmic, and datification dynamics. They are least effective when they treat online communication as if it were an extension of broadcast media.

While nobody would dispute its continued value, traditional media literacy fails to engage two truths every young person actively using social media instinctively knows. First, social media always amplifies some voices over others, and quality of content rarely seems to figure into the equation. Second, while it seems possible to “game” the platform (harnessing the power of speed and popularity for one’s promotion), it is equally possible to find oneself “gamed” by the platform—that is, abused online in ways that are worsened by algorithmic amplification.

Enter digital and data literacy. Digital literacy emphasizes the ‘doing’ side of networked computing that encourages user behaviour in some directions and discourages it in others. Data literacy allows users to understand how localized communications feed into global efforts to track our behaviours online (and increasingly offline), and package these as quantifiable trends and target markets, which are then sold to third parties like advertisers and pollsters.

In addition to digital and data literacy, education efforts may well require what might be called “influence literacy.” When it comes to media, young people have chiefly been conceptualized as audiences (for entertainment, for advertising, for education, etc.) and at times, as victims (of mediated “effects,” of social panics, etc.) On social media platforms, however, a user’s greatest value is as a producer (of information, of responses, of data, etc.) For many young people who grow up socialized through social media, perceptions of success on platforms—and by extension, life itself—maps to production-oriented metrics of popularity (likes, followers, ratings), and public perceptions of personal reputation. Policymakers, educators, and health practitioners all require a better understanding of how young people can and do internalize conflicting geographical and platform-specific norms as hard-to-articulate feelings of personal
risk, reward, shame, and self-esteem. Nobody wants to encourage a young person to fight against misinformation, only to learn they’ve risked their mental or emotional health in the process.

**Prohibitive efforts** (Legislative & Platform based). As noted earlier, many countries now have laws designed to punish users over the age of eighteen who create disinformation, and sometimes, circulate misinformation. However, far more frequent in the lives of young people are prohibitive efforts at the platform level. These can be generally thought of in three categories. Platforms can ban users, as Twitch did when it banned a game streamer making Covid-19 denialist broadcasts, who is also the son of the current President of Brazil. Platforms can decide to ban content (e.g. YouTube’s recent decision to ban health misinformation content relative to Covid-19.) They can also choose to ban a particular technology associated with misinformation (e.g. Tik Tok’s decision to ban technologies for so-called “deep faking.”)

**Nudging** The last decade has marked the rise of ‘nudging’ practice—automated environmental cues designed to trigger social behaviour in particular directions. Although it began online, (the FOGG model is frequently cited by nudge enthusiasts) nudging is now an oft-cited staple of offline policy offerings, particularly in the field of public health. On platforms frequented by young people, misinformation-oriented nudges tend to take the form of automated fact-check pop up warnings and windows. General Covid-19 related nudges show more variety: platforms have released emojis, stickers, and even in-game ‘players’ to deliver health-related messaging, usually sponsored by organisations like the WHO. Among young people, nudging seems most effective in environments where users have been primed to expect brief interruptions (e.g. gaming platforms) or use certain tools (e.g. online stickers), and least effective when it is perceived as too frequent, or overbearing.

**Recommendations for Action**

This section offers actionable insights for addressing the needs of young people in three critical areas during the infodemic: protection, care, and inclusion. These have been allocated

---


150 Ibid
according to four stakeholder groups: policy makers, platform owners, members of civil society (usually in the fields of health, education, and/or activism), and researchers (usually in policy, health, media, and/or education.)

**Policy makers.** To better protect young people from health misinformation as well as possible exploitation by extremist organisations, policy makers are urged to fortify existing education efforts in the area of media and digital literacy by immediately adding material relating to influence literacy. If they have not already done so, policy makers should begin to partner with platform owners popular with young people to trouble-shoot forthcoming areas of concern, such as the online gaming industry’s current lack of Covid-19 misinformation guidelines. To demonstrate care for their most vulnerable constituents, policy makers must ensure their messages ‘land’ well with young people by using channels and messengers they favor—ideally using youth ambassadors. True care requires both short-term and long term efforts, which is why policy makers who have not yet done so are advised to initiate discussions within their communities about how, when, and in what ways interventions might need to occur in efforts at extremist grooming observed online that threaten local stability. Finally, policy makers demonstrate a commitment to inclusion when they make efforts to grant young people a ‘seat at the table’ in decisions that affect their lives, and by adopting standards of inclusive recovery for all.¹⁵¹

**Platform owners.** To better protect young people from health-related misinformation and politically polarising phenomena like viral conspiracy theories, platform owners must commit to developing public trend-tracking mechanisms similar to Google’s Trend Tracker, per recommendations of First Draft.¹⁵² Platforms that have not yet done so should immediately develop guidelines to inform users about and prohibit the spread of Covid-19 related misinformation. To care for their most vulnerable users and combat the rise of dangerous polarising practices, platform owners are urged to develop more robust protocols to prohibit hate speech and marginalising language, especially on gaming and streaming platforms. Finally,


platform owners can demonstrate a commitment to civic inclusion by partnering with organisations expressly concerned with issues like mental health, education, and youth activism--all of whom have been on the front lines of intervening with young people--to develop better protocols for responding to moments of crisis.

**Civil society members.** Whether their focus is misinformation or elsewhere, all civil society members working with and for young people are advised to adopt practices outlined in UNICEF’s Practical Tips for Engaging with Adolescents During Covid-19. Additionally, organisations that have not already done so are urged to develop protocols about how, when, and in what ways to intervene in observed efforts at extremist grooming of young people. Finally, it is imperative that the next rollout of efforts like the WHO’s “Infodemic Manager” programs have versions that train young people--ideally recruited from existing youth activist and advocacy organisations--to work with their adult counterparts around the world.

**Researchers.** To best represent young people’s experiences negotiating misinformation of all kinds, it is imperative that researchers spend more time studying the platforms they frequent, and the communication practices they prefer. Transnational research is especially needed on the following social media practices favoured by young people: deployment of visual markers of authenticity (e.g. selfies, live streams); the global rise of influencer culture (e.g. the collapse of beauty and health advice; the rise of denialism as free speech); the rise of conspiracy theory uptake among young people (ideally focused on how health-related claims ‘fold’ into larger conspiracy narratives; especially within gaming culture); the global spread of meme culture (including its local variations and its growing political economies.) Just as important, researchers must respect and protect young people (where feasible) from the harm that might come to should their images, words or videos be treated as publicly viewable, and therefore “fair game” for study, publication, and/or circulation beyond original intention. For researchers focused on misinformation, this commitment to protection should include a plan for when, how, and with whom to intervene when observing harassment, bullying, large scale weaponised ‘pile ons’ or other sorts of aggression towards young people online. These plans are best developed in collaboration with young researchers who are already active online in areas like platform and

---

communication studies, and could use their voices amplified.

**Guiding Questions.** In the interest of protecting our most vulnerable members from the infodemic, the entirety of society would do well to ask ourselves: Are we meeting young people where they are, by following their preferred platforms and behaviours, or are we just speaking to them in places and ways where we feel most comfortable? Are we taking care to balance our responsibility to protect young people with a concern for their individual needs and voices? Those committed to principles of care should ask: How are we ensuring that we do not endanger young people further in our efforts to study them, or enact policies in their name? When we encounter admissions or information that worry us how, when, and where should we intervene? Finally, those committed to inclusion might ask: In our policies and communications, are we talking at young people, or talking with them? What steps are we taking right now to ensure that young people lead the infodemic charge tomorrow?

**Recommendations for Communication**

This section offers actionable insights for health communications regarding Covid-19 related information. These have been categorised by way of messages, messengers and audiences, with questions to guide throughout.

**Messages.** When crafting messages directed at young people, it is important to avoid shame cycle and stigma dynamics. One way to do this is to avoid messages that state or imply a blamed “you.” To foster a sense of unity, use “we” or “us.” To foster a sense of authenticity, consider the first person “I” or “me.” When choosing where a message will appear, consider that every platform features some technologies (e.g. filters, hashtags, split screens, etc.) that it is known for. To give the impression you aren’t “copy-pasting” everywhere, try to incorporate one feature ‘native’ to that platform in your message. Also bear in mind that most people in the world will be viewing any message sent using an Android phone. When crafting your message, commit to clarity, limited focus, and a visual style that places the heaviest emphasis on the most important point. Although the message should pierce the optimism bias, the most persuasive tone will be one that balances fear with hope. When deciding when (or if) a message should be released, it may be wise to consider the Media Manipulation Lifecycle.
Messengers. One common way to signal authenticity to young people is to personalise official messaging through spokespeople, using formats like, “I am X from organisation Z.” Another way is to partner with outside messengers, ideally from competing factions, to deliver a unified message. Messengers tend to deliver messages in three sorts of venues: on official channels solely focused on organisational communication (e.g. the official WHO account on Twitter); over ‘home’ channels on ‘home’ platforms (e.g. an Instagram influencer who embeds WHO health messaging into overall lifestyle communications); or through online/offline channels (e.g. a youth activist who delivers WHO messaging to residents of a refugee camp.) Among young people, the most trusted messengers are those who are perceived as “authentic.” There are a range of ways this can be achieved: via testimony, through humour, using ‘you are here’ strategies like first person address, etc. It is worth noting that the more visibility a message or messenger receives, the more liable it becomes for hijacking, weaponsation, and other activities described in the Social Media Shame Cycle. When deciding when to engage messengers to directly rebut misinformation, it can be useful to consult the Media Manipulation LifeCycle. To assist influencers with Covid-19 oriented messaging, be sure to consult Best Practices Guidelines developed by Pathos Labs. Should constant intervention be required, consider partnering with a platform to issue periodic nudges toward desired information spreading practices.

Audiences. Young people aren’t monoliths. Rather than hoping to reach everyone young, drill down more deeply into specific age brackets or identifications like gender, race, language, etc. To reach ‘follower’ communities, consider partnering with popular influencers on each platform. To reach algorithmically generated communities, consider developing ‘tracking accounts’ to trace taste identifications in music, fashion, and humour. When deciding when to engage young audiences, consider using tools like Google trends.

Questions worth asking. Have we crafted a message that breaks through the optimism bias? Have we avoided shame cycle dynamics? Would someone viewing this message understand why it appears at this time, on this platform? Are we signalling that our messages are genuine, and our messengers can be trusted? Are our messengers appearing in the right places and at the right times to get our message across? Do we understand the main identifications of the

---

154 See https://www.pathoslabs.org/research
cohort we are trying to reach? What platform tracking technologies are we using to ensure this?

**APPENDIX**

**1. Methodology**

The recommendations contained in this policy brief are built upon the research of Theresa M. Senft, PhD, Senior Lecturer in Media, Music, Communication & Cultural Studies at Macquarie University in Sydney, Australia. Dr. Senft takes an interdisciplinary approach to researching the impact of technology on society, situating her work at the nexus of gender, race, visual culture and Critical Internet Studies.

Observations on young people have been derived using methods of digital investigative ethnography. In the detection phase, behaviors are observed across different websites and social media platforms. Evidence of patterns of behavior are documented through privately held screenshots, and where necessary, screen recordings for deeper analysis. Reflections are then crowd sourced from organisations like the Selfie Researchers Network in order to gain global perspectives on practices from areas where English is not the dominant language. Where observed behaviors raise issues connected to safety, trust, risk, or harm (e.g., social media weaponization, harassment, bullying, stigmatizing, sexual or extremist grooming), crowd sourced guidance is obtained from representatives of the Association of Internet Researchers Ethics Committee.

In the analysis phase, patterns of behavior are considered by way of the Influence Ecosphere framework, the Ideological Communication Circuit framework, and the Social Media Shame Cycle. Patterned after models of persuasive computing and “emotional AI” these three frameworks describe how digital platforms work to transform personal and social emotions into self-reinforcing informational flows that can be observed (although not completely, due to the proprietary nature of platform technologies.) When it appears necessary for the public good that

---

155 See http://selfieresearchers.org
156 See http://aoir.org
these flows be altered, (e.g., when behaving as an educator in an online classroom where a student appears at risk) the Social Media Risk Management Matrix is consulted.

These frameworks were developed to give common frames for journalists, researchers, technologists, and members of civil society to understand the impacts of persuasive computing environments on youth identity, community formation, and ideological pooling. The policy recommendations in this document have been informed by a literature review of research conducted by organisations explicitly focused on the following areas:

- Protecting and empowering young people
  (e.g. UNICEF; Save the Children)
- Measuring young people’s attitudes regarding Covid-19
  (e.g. King’s College; Youth Insight)
- Reporting young people’s patterns of social media use
  (e.g. Global Web Index; App Annie)
- Evaluating young people’s attitudes toward social media
  (eg. Pew Center; Common Sense; Axios)
- Fighting misinformation
  (e.g. The Shorenstein Center; Data & Society; First Draft; Reboot Pathos Labs)
- Evaluating the role of social and behavioural science during Covid-19
  (e.g. Social Science in Humanitarian Action; Campaign for Social Science)

The frameworks discussed in this brief have been reviewed by professionals in the fields of risk communication and media studies, as well as experts in the health field, including the CDC and WHO. Feedback from these reviews has been incorporated into each framework.
2. The Disordered Ecosphere

3. Agents of Dis and Misinformation

---

**THE DISORDERED INFORMATION ECOSPHERE**

- **CONSPIRACY THEORIES**
  - false stories repeated over time with known contrast to received wisdom. Includes reference to fact that others are trying to suppress the truth

- **RUMOURS**
  - spreading non-confirmed info

- **CLICKBAIT**
  - non-false information presented to maximize clicks for profit

- **FAKE NEWS**
  - false information produced to maximize clicks for profit

- **MISINFORMATION**
  - unknowingly spreading false info

- **MALINFORMATION**
  - spreading info intended to denigrate the other party

- **HYPER PARTISAN INFO**
  - spreading info intended to promote one party/candidate

- **ONLINE PROPAGANDA**
  - spreading info intended to promote one party/candidate

---

**DISINFORMATION AGENTS**

- **BOTS**
  - inflate followers & likes
  - influence political discourse
  - attack dissidents
  - manipulate public opinion
  - manipulate news search rankings

- **FOREIGN GOVERNMENTS**
  - 'information warfare'

- **HIGHLY PARTISAN OUTLETS**
  - RIGHT WING EXAMPLES
    - Daily Caller
    - Breitbart News
    - InfoWars
  - LEFT WING EXAMPLES
    - Occupy Democrats
    - Addictive Info
    - Daily Newshub

- **POLITICIANS**

- **TROLLS**
  - unpaid individuals
  - paid groups ('troll farms')

- **FAKE NEWS WEBSITES**
  - motivated by profit
  - motivated by politics

- **CONSPIRACY THEORISTS**
  - can feed mainstream media

---

3. Young People: Policy & Industry Definitions

**Policy Definitions**

**Young People: 10-24 Years Old**

There is no universally agreed upon definition, but policy researchers tend to combine overlapping categories of adolescence and youth, as below:

- **Adolescents**: For statistical purposes, WHO defines adolescents as 10-19 years.
- **Youth**: For statistical purposes, the United Nations defines youth as 15-24 years.

**Industry Definitions**

**Young People: Gen Z, Gen Alpha**

Researchers in digital industry and marketing studies classify young people by ‘generation,’ as per the chart below:

- **Gen X**: 1961 – 1981
- **Gen Y**: 1981 – 1997
- **Gen Z**: 1997 – 2012
- **Gen Alpha**: Born 2012

- **Age Groups**:
  - Age 54-74
  - Age 40-55
  - Age 25-39
  - Age 16-24
  - Age 16 below

4. Young People: Platforms & Laws

**Platforms & Law**

**Platform Regulation**

<table>
<thead>
<tr>
<th>Platform Entry</th>
<th>10-15</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Non-kids&quot; Sites</td>
<td>13</td>
</tr>
<tr>
<td>Platform Entry</td>
<td>14-16</td>
</tr>
<tr>
<td>Sexually Explicit Sites</td>
<td>18</td>
</tr>
<tr>
<td>Platform Screening</td>
<td>17-18</td>
</tr>
<tr>
<td>Advertising for Age-delimited Products</td>
<td>18-21</td>
</tr>
</tbody>
</table>

**Protected Age**

Age one can be treated in the criminal justice system. Varies by country.
- In Australia: 10
- In France: 13
- In Norway: 15

**Statutory Age**

Age of consent for sexual activity, medical procedures and informed consent. Varies by regions within a country.
- In Australia: 14-16 depending on state, or of sufficient maturity that they are able to comprehend the procedure and give informed consent.

**Age of Majority**

Age someone passes from ‘minor’ status into adulthood. Generally associated with voting age. Varies from country to country, although usually 18.

**Purchasing Age**

Age someone can legally purchase a substance, usually alcohol or cigarettes.
- Varies from country to country; usually 18-21, although sometimes younger.
## 5. Young People: Attitudes

### Young People: Key Distinctions
Focus: Attitudes re. COVID-19 & Misinformation

<table>
<thead>
<tr>
<th></th>
<th>What we know</th>
<th>Significant because</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 SIZE</td>
<td>Large demographic: 16-24 year olds make up 30% of the global population; and 40% are under 24</td>
<td>Infection rates highest among 20-29 year olds. High possibilities of transmission due to high economic vulnerability</td>
</tr>
<tr>
<td>02 MINDSET</td>
<td>Lower concern for physical health relative to general population; higher concern for mental and emotional health.</td>
<td>Less apt to search for accurate health information when overwhelmed with dire economic, education and/or domestic violence concerns</td>
</tr>
<tr>
<td>03 TRUST</td>
<td>Low levels of trust in public institutions, private corporations, social media platforms, and news information found online.</td>
<td>Blanket distrust of authority, combined belief they can spot social media misinformation, leaves them vulnerable to manipulation</td>
</tr>
<tr>
<td>04 VALUES</td>
<td>High interest in narratives that emphasise emotional authenticity of felt experience; lower interest in arguments driven by evidence.</td>
<td>Debunking strategies depend on openness to argument. Rumour, stigma and conspiracy theory turns on feeling-centered narratives.</td>
</tr>
<tr>
<td>05 RISK</td>
<td>This demographic is at high risk for profiling and targeting by political extremist groups.</td>
<td>Extremist groups have histories with gaming and other platforms online, and have already woven health misinformation into narratives there.</td>
</tr>
</tbody>
</table>
## Young People: Social Media Practices

### Young People: Key Distinctions

**Focus:** Engagement Patterns with Social Media

<table>
<thead>
<tr>
<th>What we know</th>
<th>Significant because</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>01 PLATFORMS</strong></td>
<td>Prefer platforms that prioritise collaborative entertainment over deliberative discourse (e.g., TikTok, gaming platforms, etc.).</td>
</tr>
<tr>
<td><strong>02 PRACTICES</strong></td>
<td>Heavy reliance on images over textual communication. Accept alterations and/or appropriations, as long as intended meaning is understood by the ‘right’ parties.</td>
</tr>
<tr>
<td><strong>03 PROTOCOLS</strong></td>
<td>Higher than average awareness of cross-platform algorithmic processes like feeds, trends, monetisation flows, etc.</td>
</tr>
<tr>
<td><strong>04 PROCEDURES</strong></td>
<td>Greater reliance on social media for employment and education leads to highly curated procedures for presenting ‘best self online, along with lower tolerance for conflict outside of anonymous contexts.</td>
</tr>
<tr>
<td><strong>05 PREFERENCES</strong></td>
<td>Preferences for online engagements are aided using risk management procedures that combine awareness of platform protocols with sensitivity to social stigma through unspoken “drama.”</td>
</tr>
</tbody>
</table>

Need to source misinformation when platforms are visually embedded (e.g., watching someone live-streaming online game play via Twitch, or made using split screens (common on TikTok).

Memes especially are subject to greater ambiguity as they go viral. Many pieces of COVID-19 misinformation began as humour parodies.

Although young people frequently report ‘training’ algorithms to return information they desire and require, with more time online, grows greater propensity ‘filter bubble’ risk.

Fast-driven discussions online can be easily derailed by accusations of creating ‘drama’ (a charge that tends to carry class, race and gender connotations), meant to induce shame and fear of social stigma in the accused.

Young people may be quite open to serving as information advocates and activists in some online environments, uninterested in doing so in others; and at high risk for targeted harassment and abuse in others.
7. Strategies for Fighting Misinformation

MISINFORMATION: TACTICAL RESPONSES TO DATE

DEBUNKING
- Defined: Spotting and correcting misinformation, directing where verifiable sources of information.
- Example: Encouraging family members to use WhatsApp chatbot to verify COVID-19 rumors (http://story姹n.org)
- Good strategy for: contexts where users have an established emotional investment.
- Drawbacks: public debunking efforts perceived as unwarranted, incentivizing or intensifying rather than reining in, difficulties subjecting original misinformation.

PRE-BUNKING
- Defined: Teaching users to spot disinformation techniques, such as disinformation, create or distribute false accounts, clickbait, etc.
- Example: “Playing game” “Spot the FAKE” at at standoff.
- Good strategy for: maximising power of disinformation before it spreads, as disinformation campaigns are often spread faster.
- Drawbacks: unless efforts wind upinding or gaining viral, information warfare is limited.

INFO ACTIVISM
- Defined: mobilizing social goodwill, party members, fan communities to fight (targeted) attacks made by individuals and/or groups on social media.
- Example: “Whip users” “ghost purchasing” notes inleshooting Trump Rally, Twitter & Flip Long posting成coal surveillance app, among others.
- Good strategy for: clamping down or downing accounts that spread false information.
- Drawbacks: Activist groups are highly resistant to being used for social disinformation purposes.

SOCIAL RESPONSES

INFLUENCING
- Defined: influencing individuals or groups to leverage their popularity on behalf of authenticity.
- Examples: “Let’s change the narrative” “I live in a very COVID safe area” “We need to take this seriously” “I heard this from a news reporter.”
- Good strategy for: increasing social goodwill within a person’s sphere of influence.
- Drawbacks: Not all change is positive; it all change is not necessarily a positive change in the absence of evidence.

INSTITUTIONAL RESPONSES

EDUCATION
- Defined: usually as health or media literacy, sometimes as information literacy.
- Example: “Breaking down walls” “Breaking down barriers” to understanding the impact of misinformation.
- Good strategy for: building awareness and understanding of the impact of information on health and society.
- Drawbacks: many efforts focused on understanding the impact of misinformation.

LEGISLATION
- Defined: laws and regulations put in place to regulate the spread of false information.
- Example: “Legislation” “Regulations” “Laws” put in place to stop the spread of false information.
- Good strategy for: regulating misinformation in a way that is practically enforceable.
- Drawbacks: legislation can often be seen as a way of limiting freedom of expression.

REGULATION
- Defined: “Regulation” “Regulations” “Rules” put in place to stop the spread of false information.
- Example: “Regulation” “Rules” “Guidelines” put in place to stop the spread of false information.
- Good strategy for: regulating misinformation in a way that is practically enforceable.
- Drawbacks: legislation can often be seen as a way of limiting freedom of expression.

NUDGING
- Defined: implementing small, incremental changes in an environment that make it easier to adopt behaviors.
- Example: “Let’s change the narrative” “I live in a very COVID safe area” “We need to take this seriously” “I heard this from a news reporter.”
- Good strategy for: increasing social goodwill within a person’s sphere of influence.
- Drawbacks: Not all change is positive; it all change is not necessarily a positive change in the absence of evidence.

PRESIDENTIAL APPROACH
- Defined: “President” “Approach” “Strategies” put in place to stop the spread of false information.
- Example: “President” “Approach” “Strategies” put in place to stop the spread of false information.
- Good strategy for: regulating misinformation in a way that is practically enforceable.
- Drawbacks: legislation can often be seen as a way of limiting freedom of expression.

Process Draft
9. Examples of Best Practice

**DEBUNKING**

**Best Practices Targeting Young People**

Organisations like Poynter have learned that by centering teens in leadership roles, they are more likely to be perceived by young speakers as talking with, rather than at them.

---

**DEBUNKING**

**Best Practices Led by Young People**

California youth group Youth Will livestreams COVID-19 Town Hall sessions, where young people debunk pandemic fictions, and discuss the disease’s impact on their lives. See youthwill.org
In BAD NEWS, game-players assume the role of fake news-monger. Translated into ten European languages beyond English, the game is free online. It is designed for ages 15-35, with a "junior" version for ages 8-11.

Developed by the Social Decision Making Lab at Cambridge University, BAD NEWS is the first game to feature in peer-review research on pre-bunking. See www.getbadnews.com

In GO VIRAL, also developed at Cambridge, game-players take the role of an online user trying to spot misinformation. The five-minute game has three levels: fearmongering, using an expert to bolster false information, and the art of creating a conspiracy-theory.

Developed at Clemson University, SPOT THE TROLL delivers images featuring social media content, asking which are from legitimate sources, and which are from trolls. spotthetroll.org

On the game-streaming site Twitch, broadcaster Kitboga delivers regular commentary on COVID-19 related scams as part of his ongoing scam coverage, pointing out common tactics deployed. See https://www.twitch.tv/kitboga

The #IamNotAVirus selfie campaign was an informal movement spearheaded by international students who found themselves repeatedly targeted by COVID-19 related racial discrimination.

This parody features typical imagery featured in posts made by "natural health gurus" on Instagram, known to spread misinformation on the platform. It shows how false information can gain traction through attractive packaging.

While parodies can be powerful instruments, they can also backfire: a recent parody image circulated on Facebook to mock the "peppers protect against COVID" hoax was widely circulated on WhatsApp as true information.
INFO-ACTIVISM

Best Practices Led by Young People

In Southern Thailand, youth activists THE MOTIVE joined forces with Malay Muslim activists THE PATANI to create short videos focused on COVID-19 misinformation and hate speech directed at Muslims.

The effort mobilized various Islamic religious groups, including separatist insurgents to embrace public health recommendations and more peaceful community relations.


In Indonesia, young activists at MAFINDO are spearheading COVID-19 hoax-busting efforts in the region, assembling a robust database of common hoaxes. MAFINDO is currently partnered with the WHO.

See www.mafindo.or.id

Around the world, medical student activists have joined MORE VIRAL THAN THE VIRUS, an activist campaign aimed at spotlighting COVID-19 myths

MoreViralThanTheVirus
In the UK, online influencers are paid to spread messages regarding best health practices. Filipina health expert Bianca Gonzales (more than 7 million YouTube followers) and Bangladeshi doctor Jahangir Kabir (1 million Facebook followers) are among those listed as part of the effort.

The most effective online influencers aren’t always human. In Thailand, the Health Ministry uses a shiba inu nicknamed Zonghchai (“chief shiba.”) On Facebook and printed posters, Zonghchai shares messages like, “Stand three dog-lengths apart when indoors.”

On TikTok, the viral “hand washing dance challenge,” based on a popular pop song called Ghen, was the result of a collaboration by singer Khac Hun and Vietnam’s National Institute of Occupational Safety and Health.

People communicating through at-work selfies are often perceived to be more authentic than those posing in “ideal self” shots. When someone holds a sign in a selfie, they are signaling that they feel themselves to be speaking as part of a larger movement.

Communications that reference fan communities can tap into the fun of belonging to a group. The example here features a model wearing a mask printed with a Tweet made about the K-pop group BTS, whose fans are known as the BTS Army.

Messages focused on the authenticity of the broadcaster’s location can be extremely powerful way to fight location-oriented rumours. Here, Nurses of TikTok take on misinformation while standing on wards, helping debunk claims about empty hospitals.
**Banning users.** Gaming streamer Jair Bolsonaro was recently banned from Twitch for spreading COVID-19 misinformation, including denialism. This story made larger headlines because he is also the son of the current President of Brazil.

**Banning content.** In October 2020, YouTube announced it was expanding its policy surrounding misinformation by removing any content that includes claims about COVID-19 vaccinations that contradicts information from health authorities.

**Banning technologies.** In October 2020, TikTok announced it was banning deep fake technologies from its platform. Deep faking is artificial intelligence-powered manipulation of audio and video designed to mislead people about something someone may have done or said.

---

**YouTube’s fact check window** is an example of nudging by way of **technological control.** The platform now automates fact checks on videos corresponding to particular ‘hot’ terms in searches. Rather than relying on users to fact check themselves, pop-up windows assume a need for a assistance from the moment a users clicks a link to watch a video.

**Snapchat stickers** are an example of nudging by way of **technological affordance.** Stickers are a widely used feature on Snapchat, adding health advice allows users to share messages in a format they enjoy using.

**Minecraft’s COVID Creepers** are a good example of nudging by way of **platform-orchestrated interaction.** Players of the popular game now periodically encounter “creepers” (in-game villains) giving COVID-19 related information. Since it’s not uncommon for characters to interrupt users during gameplay, the nudging isn’t seen as particularly intrusive.
### 10. Recommendations for Action

#### Action Recommendations

<table>
<thead>
<tr>
<th><strong>PROTECTION</strong></th>
<th><strong>CARE</strong></th>
<th><strong>INCLUSION</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>POLICY MAKERS</strong></td>
<td><strong>PLATFORM OWNERS</strong></td>
<td><strong>CIVIL SOCIETY MEMBERS</strong></td>
</tr>
<tr>
<td>- Repeat testing health literacy education efforts to achieve platform culture and social media influence literacy.</td>
<td>- Partner with platform owners for upcoming areas of concern (e.g., misinformation policies on gaming platforms).</td>
<td>- Include young people at the policy level when making decisions directed toward them.</td>
</tr>
<tr>
<td>- Partner with platform owners for upcoming areas of concern (e.g., misinformation policies on gaming platforms).</td>
<td>- Develop protocols for monitoring extremist growing efforts online that leverage local statistics.</td>
<td>- When developing policies, commit to standards for mental health recovery of all.</td>
</tr>
</tbody>
</table>

**RESEARCHERS**

- Develop and make public trend tracking features similar to YouTube's trend watch, per First Draft's recommendations.
- Develop and circulate clear hate misinformation policies for gaming environments.
- Adopt communication strategies that comply with Unicef's Ethical Engagement with Young People Guidelines.
- Develop organizational protocols for reporting and tracking patterns of observed extremist growing.

**QUESTIONS TO GUIDE**

- Are we meeting young people where they are: following their preferred platforms and behaviors?
- Are we balancing consent for protecting young people as a vulnerable group with consent for their individual needs and voices?
- Who are the most vulnerable among us, and how are we ensuring we don’t endanger them further?
- When we encounter admissions or information that worries us, how, when, and where should we intervene?

### 11. Recommendations for Communication

#### Communication Recommendations

<table>
<thead>
<tr>
<th><strong>MESSAGES</strong></th>
<th><strong>MESSENEGERS</strong></th>
<th><strong>AUDIENCES</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WHO</strong></td>
<td><strong>WHERE</strong></td>
<td><strong>HOW</strong></td>
</tr>
<tr>
<td>- To help with shame stigma dynamics, avoid messages that state or imply a “blamed” you.</td>
<td>- Software level: make use of at least one tech feature “authentic” on the platform on which message appears.</td>
<td>- Social level: create language that is clear, precise, and relevant to the audience.</td>
</tr>
<tr>
<td>- To foster a sense of utility, use “you” or “we.”</td>
<td>- Visual level: state the message using graphics, symbols, and visuals to make it more accessible to the audience.</td>
<td>- Emotional level: “fear” (limited) to overcome optimism bias.</td>
</tr>
<tr>
<td>- To signal authenticity, humanize messengers from organizations (e.g., Tim S. from NTA)</td>
<td>- Establish official channels on platforms</td>
<td>- Messengers signal authenticity through formats like testimony, humor, and first-person address.</td>
</tr>
<tr>
<td>- To build trust, partner with credible messengers; avoid messages with competing groups.</td>
<td>- Partner with influencers for “embedded” messaging offline</td>
<td>- All messages are at risk for Nailing and vaporization should messenger be caught in “shame cycle fallout.”</td>
</tr>
<tr>
<td>- To foster a sense of authenticity, use “you” or “we.”</td>
<td>- Partner with platforms for embedding messaging offline</td>
<td>- Messengers signal authenticity through formats like testimony, humor, and first-person address.</td>
</tr>
<tr>
<td>- To foster a sense of utility, use “you” or “we.”</td>
<td>- Establish official channels on platforms</td>
<td>- Messengers signal authenticity through formats like testimony, humor, and first-person address.</td>
</tr>
<tr>
<td>- To foster a sense of utility, use “you” or “we.”</td>
<td>- Partner with platforms for embedding messaging offline</td>
<td>- All messages are at risk for Nailing and vaporization should messenger be caught in “shame cycle fallout.”</td>
</tr>
</tbody>
</table>
12. Reconfiguring Health Information Literacy

Health Information Literacy, current state

Expanded Health Info Literacy

Health Information Literacy, revised