CCOD

Council for Countering Online Disinformation (CCOD)

Advancing trustworthy information in the digital age | https://ccod.co.uk/

Using the Adaptive Cycle Framework in Policy Planning Approaches

Disinformation and the Case for Preventative Policy Design



CCOD

Council for Countering Online Disinformation (CCOD)

Advancing trustworthy information in the digital age | https://ccod.co.uk/

Using the Adaptive Cycle Framework in Policy Planning Approaches

Disinformation and the Case for Preventative Policy Design

By Daniel Lewis

with Giovanna da Custódia, Holly Barnett, James Rice, and Sakshee Singh

Edited by Sakshee Singh

Submitted on 31st July 2026

A CCOD Policy Fellow paper. This does not reflect the views of the CCOD; each author's view is their own.

Summary

Disinformation policy tends to respond to each crisis and then wait for the next. Drawing on resilience theory, this report applies the adaptive cycle framework – exploitation, conservation, release, reorganisation – to online safety regulation, financial markets, public health, and conflict, emphasising the importance of pre- emptive, strategic, policymaking. Durability, not disruption, is the test.

1. Context and Use-cases

As Esme Fowler-Mason has outlined in her 2023 critique of the OSA, "[t]he lack of engagement by Parliament with academics in the field of algorithmic accountability could lead to a legislative instrument unable to make social media safe."[1] There were several oversights in the rollout of the OSA, and this shortsightedness has allowed the continued proliferation of disinformation - amplified by platform algorithms and their intrinsic 'attention economy' design.[2]

The information environment the OSA is faced with constitutes a CAS, where (different) platforms, users, investors, governments, catalysts, and trends, together form the "emergent properties,"[3] meaning no single entity controls or designs the information environment that disinformation flourishes in. With the application of the ACF, which assesses the CAS heuristically,[4] this encourages a long-term policy approach which seeks to transform the environment from a reoccurring cycle of policy-manoeuvre-return (to disinformation), into an escape from this cycle.

Ofcom does have plans to facilitate algorithm regulation, which are mandated for 2027,[5] but this will likely be informational with enforcement unlikely to alter platforms underlying profit models.[6]



Image Source: https://www.instituteforgovernment.org.uk/explainer/ofcom (2026)

KEY DEFINITIONS

Disinformation

False or misleading information created and spread deliberately to deceive, manipulate, or serve a political, financial or social aim.

Information Environment

The wider ecosystem in which information is produced, ranked, distributed and consumed: platforms, recommender algorithms, actors, and norms.

The Framework Terms:

Complex Adaptive System (CAS)

A system of many interdependent agents whose interactions produce system-level "emergent properties" that no single actor designs or fully controls. (The ACF will presuppose that a disinformation environment behaves as a CAS.)

Adaptive Cycle Framework (ACF)

A heuristic from resilience theory describing how complex systems move through four recurring phases; used here to locate where a disinformation environment sits and where intervention is possible.

Resilience

A system's capacity to absorb disturbance and persist. Worth flagging the double edge: resilience is value-neutral, so an entrenched disinformation regime is highly resilient. The aim of policy is to reduce the resilience of the disinformation environment while building resilience into the counter-measures.

Escape

A reorganisation that exits the cycle: the structural conditions enabling the disinformation environment are removed, so it cannot reconstitute.

[1] Fowler-Mason, E., 2023. The Online Safety Bill needs more algorithmic accountability to make social media safe. [Online] Available at: https://blogs.lse.ac.uk/medialse/2023/02/08/the-online-safety-bill-needs-more-algorithmic-accountability-to-make-social-media-safe/

[2] Fondation Hirondelle, 2025. Strengthening Information Integrity: Towards a Preventive and Inclusive Approach to Disinformation. [Online] Available at: https://www.hirondelle.org/wp-content/uploads/2025/07/Disinformation-policy-paper-2025.pdf

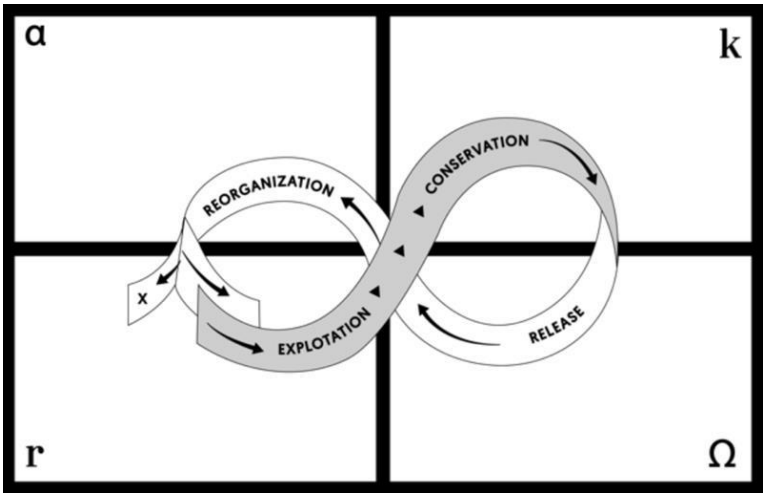
[3] Miller, J. & Page, S., 2007. Complex Adaptive Systems: An Introduction to Computational Models of Social Life. Princeton: Princeton University Press.

[4] Gunderson, L. H. & Holling, C. S., 2002. Panarchy: Understanding Transformations in Human and Natural Systems. Washington: Island Press.

[5] techUK, 2025. Online Safety Act: Categorisation Register Pushed Back to July 2026 in New Timeline. [Online] Available at: https://www.techuk.org/resource/online-safety-act-categorisation-register-pushed-back-to-july-2026-in-new-timeline.html

[6] Starr, K., 2025. Systemic Blindness: How EU and UK Law Misidentify Harm and Protect the Platforms That Enable It.

Drawn from resilience theory and increasingly applied to social and political systems, the ACF would frame disinformation issues as dynamic circumstances that move through four recurring phases (exploitation, conservation, release, reorganisation) that can be located, anticipated, steered, and nullified. Its value for policy planning is that it makes the reorganisation phase visible as the decisive moment. Good policy, on this reading, is preventative by design.



KEY DEFINITIONS (cont.)

Exploitation (r)
Rapid emergence and growth, as a disinformation capacity seizes available resources (an information vacuum, attention, engagement, tools).

Conservation (K)
Entrenchment: the configuration becomes rigid, connected and self-reinforcing. The stable, hard-to-reform state where most systems sit.

Release (Ω)
Disruption: accumulated structures collapse or are broken open by exposure, scandal, coordinated pressure, or a policy shock.

Reorganisation (α)
The decisive phase: the system rebuilds, and either escapes (x) the cycle or reseeds a new exploitation phase.

In sum, the OSA was a short-sighted reorganisation phase, which facilitated an easy return to the cycle into the exploitation phase, faced with an impending conservation phase of entrenchment. This can be pinned down to three policy issue points:

- A failure to penalise the financial models behind algorithmic amplification
- An exemption of "legal but harmful" adult disinformation from platform responsibility [7]
- A lack of enforceable transparency for algorithmic code and corporate governance [8]

---

[7] techUK, 2022. Removing confusing 'legal but harmful' definition for adults is the right way to ensure the Online Safety Bill can achieve its objectives. [Online] Available at: https://www.techuk.org/resource/removing-confusing-legal-but-harmful-definition-for-adults-is-the-right-way-to-ensure-the-online-safety-bill-can-achieve-its-objectives-to-make-the-uk-the-safest-place-to-be-online-for-children.html

[8] Molly Rose Foundation, 2026. Written evidence submitted by the Molly Rose Foundation: Response to the Science, Innovation and Technology Committee's call for evidence on social media restrictions. SMR0002. [Online] Available at: https://committees.parliament.uk/writtenevidence/162918/default/

2. Applying the ACF to Disinformation Case Studies

2.1. Incitement of Civil Unrest via Algorithmic Feeds: Belfast 2026

By Daniel Lewis

Northern Ireland offers the clearest recent case and mapped onto the adaptive cycle it shows a system the Online Safety Act has disrupted but not escaped.

Exploitation **(r).** A knife attack in Belfast opens an information vacuum that bad-faith actors fill within hours with false claims about an assailant's identity and the location of migrant accommodation. Researchers note that falsehoods rush in precisely while authorities are still under pressure to confirm details, and engagement optimised feeds surface the most provocative material fastest, converting local panic into swift physical mobilisation.[9]

Conservation **(K).** This is no longer episodic. The same pattern ran from the 2024 Southport spillover to the June 2025 Ballymena riots and again in Belfast in 2026, with social media shifting from primarily an amplification platform to an organising tool used to coordinate protests and even identify properties for attack. The configuration is entrenched and self-reinforcing: platforms profit from the engagement local crises generate, and the PSNI Chief Constable has named the malign influence of bad actors spreading anti-immigration disinformation a standing feature.

Release **(Ω).** Disruption arrives reactively. Ofcom issued open letters reminding platforms of their duties, and the technology minister asked the regulator to raise compliance urgently with X; days earlier it has finalised a crisis-response protocol requiring response teams and police hotlines.

Reorganisation **(α).** Relapsing. The escape is being missed. The protocol covers only illegal content and content harmful to children, and expressly not disinformation campaigns; it must still clear Parliament, and rests on platforms themselves judging when a crisis applies rather than on immediate enforcement against the amplifying algorithm. The structural driver survives intact, which is why the cycle has returned three summers running, whilst MPs warn the Act cannot tackle such disinformation, leaving the UK exposed to a repeat.[10] Where senior law officers, the Attorney General among them, are now abandoning X rather than the regime curing the environment is a symptom of relapse, not escape.

RECOMMENDATION(S):

An escape-oriented reorganisation would shift from reactive takedown to pre-emptive, content-neutral duties on algorithmic amplification during designated information incidents, while simultaneously making use of the momentum surrounding the incident and thus utilising the release phase.

[9]Right Response Team, 2026. The Belfast Riots: Disinformation, Mobilisation and the Far Right. [Online] Available at: https://hopenothate.org.uk/2026/06/15/the-belfast-riots-disinformation-mobilisation-and-the-far-right/

[10]Milmo, D., 2025. UK MPs warn of repeat of 2024 riots unless online misinformation is tackled. [Online] Available at: https://www.theguardian.com/technology/2025/oct/17/uks-riots-of-2024-will-repeat-unless-misinformation-is-tackled-mps-warn

Such a reorganisation must be binding rather than advisory, enforced against the recommender systems that monetise panic, and closing the disinformation gap the protocol leaves open. It should be paired with mandated crisis-time prominence for authoritative local sources, such as those recommended by Fondation Hirondelle, so the vacuum is filled by verified detail before falsehood seizes it. Durability must be the test.

2.2. Economic Engineering of Systematically Disruptive Financial Disinformation: A Global Case Study

By James Rice

Financial disinformation — false claims engineered to move markets, trigger deposit flight, or monetise panic — now circulates within a global information environment that behaves as a complex adaptive system (CAS). Platforms, algorithmic feeds, retail investors, state-linked actors, and regulators interact to produce emergent properties no single actor designs or fully controls. Mapped onto the adaptive cycle, the system shows repeated disruption without escape.

Exploitation **(r).** Fabricated financial content seizes available resources — attention, credulity, and the speed advantage falsehood holds over verification. A single hacked Associated Press tweet in 2013 erased an estimated $130 billion in equity value within minutes, and a fabricated 2023 image of a Pentagon explosion moved the S&P 500 in seconds.[11] As the IMF's financial stability surveillance makes clear, markets remain exposed to shocks that propagate faster than authorities can respond — the information vacuum is filled before circuit-breakers engage.[12]

Conservation **(K).** The dominant digital power configuration has become increasingly entrenched. Engagement-optimised recommender systems monetise volatility; synthetic media tools have industrialised production; and the EU's monitoring of foreign information manipulation and interference (FIMI) finds manipulation concentrated overwhelmingly on a handful of platforms.[13] World Bank work on financial-sector stability underscores the stakes: well-functioning banking systems rest on public confidence, a trust that digital misinformation erodes faster than supervisors can rebuild it.[14] The system is rigid, connected, and self-reinforcing — a stable, hard-to-reform state.

Release **(Ω).** The 2023 collapse of Silicon Valley Bank was the paradigmatic shock: social-media amplified panic drove $40 billion in withdrawals — 23% of deposits — in a matter of hours, and the Federal Reserve conceded

[11] Deepfakes, financial stability, and EU regulation: navigating the risks of synthetic media, 2026. Capital Markets Law Journal. [Online] Available at: https://academic.oup.com/cmlj/article/21/2/kmag008/8644171

[12] IMF, 2025. Global Financial Stability Report, October 2025: Shifting Ground beneath the Calm. Washington, DC: International Monetary Fund. [Online] Available at: https://www.imf.org/en/publications/gfsr/issues/2025/10/14/global-financial-stability-report-october-2025

[13] EEAS, 2026. Information Integrity and Countering Foreign Information Manipulation & Interference (FIMI). [Online] Available at: https://www.eeas.europa.eu/eeas/information-integrity-and-counteringforeign-information-manipulation-interference-fimi_en

[14] World Bank, n.d. Financial Stability (Banking Regulation, Supervision and Crisis Management). [Online] Available at: https://www.worldbank.org/en/topic/financialstability

that social media had fundamentally changed the speed of bank runs.[15] Accumulated regulatory assumptions collapsed with the bank.

***Reorganisation* (α).** The most prominent risk is lapsing into public disorder. The EU's Digital Services Act imposes systemic-risk duties on very large platforms,[16] and the UN Global Principles for Information Integrity (2024) name the monetisation of falsehood directly.[17] Yet market-abuse instruments such as MAR and MiFID II remain ill-suited to anonymised, high-velocity synthetic manipulation, *(see n 11)* and the advertising-driven business model — the structural driver — survives intact. The system is reseeding a new exploitation phase, now AI-accelerated.

RECOMMENDATION(S):

An escape-oriented reorganisation would:

1. Impose content-neutral duties on algorithmic amplification during designated financial information incidents, targeting the release phase before panic converts to withdrawal;
2. Operationalise the UN's "healthy incentives" principle by penalising the monetisation of market-moving falsehood, dismantling the resilience of the disinformation regime rather than merely describing it;
3. Mandate crisis-time prominence for authoritative sources — central banks, deposit insurers — so the vacuum that opens each phase is filled by verified detail before falsehood seizes it; and
4. Build cross-border, FIMI-style threat-sharing between financial regulators, since CAS spanning jurisdictions cannot be escaped by national policy alone.

Durability, not disruption, must be the test.

2.3. Synthetic Political Personas and the OSA's Blindspot: Danny Bones (2026)

By Sakshee Singh

Danny Bones offers a test of the ACF's anticipatory function, applied here to a system still mid-cycle rather than one already resolved, in keeping with the framework's prior use in planning for harms not yet fully realised.

***Exploitation* (r).** Danny Bones is an artificial-intelligence-generated persona produced by an anonymous collective operating under the name the Node Project, whose output addresses immigration, crime, inequality, and national identity.[18] Advance UK, a registered political party, commissioned the collective to produce its

[15]The Conversation, 2023. Digital bank runs: social media played a role in recent financial failures but could also help investors avoid panic. [Online] Available at: https://theconversation.com/digital-bank-runs-social-media-played-a-role-in-recent-financial-failures-but-could-also-help-investors-avoid-panic-204947

[16]European Commission, 2026. Tackling Disinformation (EC Library Guides). [Online] Available at: https://ec-europa-eu.libguides.com/disinformation-and-misinformation/websites

[17]United Nations, 2024. United Nations Global Principles for Information Integrity. New York: United Nations. [Online] Available at: https://www.un.org/en/information-integrity/global-principles

[18]The Bureau of Investigative Journalism, 2026. Danny Bones: the AI rapper funded by a far-right party. [Online] Available at: https://www.thebureauinvestigates.com/stories/2026-03-12/danny-bones-meet-the-ai-rapper-funded-by-a-far-right-party

principal campaign video for the Gorton and Denton by-election, incorporating Danny Bones material directly into electoral messaging.[19] The exploited vacuum is regulatory rather than informational: no existing classification, user-generated content, journalism, or paid political advertising, adequately accounts for a synthetic persona that is simultaneously commercially monetised and functioning as commissioned campaign material.

Conservation **(K).** The configuration entrenches through a self-reinforcing financial structure. Spotify confirmed that Danny Bones' output was eligible to generate royalty revenue on the platform. Following investigative reporting on the persona, cryptocurrency tokens referencing Danny Bones appeared on the Solana blockchain within a day, and the Node Project solicited donations to an associated wallet.[20] A researcher at Leiden University's Institute of Political Science observed that, should other political actors regard the tactic as effective, its use for extremist political messaging may diffuse more widely. This suggests the entrenchment is not confined to a single persona, but to a replicable production model.

Release **(Ω).** Disruption arrived through investigative journalism rather than statutory enforcement. Reporting by the Bureau of Investigative Journalism prompted TikTok to remove the Node Project's account and Instagram to take down a portion of its content, while YouTube applied AI-generated-content labels and Spotify stated the material had undergone human review under its existing content standards. The UK Electoral Commission, the statutory body responsible for regulating political finance and campaign spending, stated only that it was assessing the findings in line with its remit.

Reorganisation **(α).** No binding reorganisation has yet occurred, leaving the system exposed to reseeding rather than escape. The All-Party Parliamentary Group on Political and Media Literacy took the findings to Parliament, with its co-chair characterising the case as revealing a regulatory gap permitting synthetic personas to circumvent democratic transparency norms for commercial gain. Whatever the by-election's limited electoral impact, this does not resolve the underlying structural gap: AI-generated personas funded by registered political actors remain outside both the OSA's content categories and existing electoral-transparency requirements.

RECOMMENDATION(S):

A pre-emptive, escape-oriented reorganisation would introduce statutory disclosure obligations for AI-generated personas commissioned or financed by registered political actors, classifying such material as regulated campaign expenditure irrespective of format. This should be paired with platform-level labelling duties triggered by the underlying funding relationship rather than content classification alone, closing the gap between user-generated content and paid political communication before the tactic diffuses to other actors.

---

[19] Novara Media, 2026. Meet the AI Rapper Working for the Far-Right. [Online] Available at: https://novaramedia.com/2026/03/12/meet-the-ai-rapper-working-for-the-far-right/

[20] The Bureau of Investigative Journalism, 2026. The hate economy in action: how an AI rapper tried to cash in on our coverage. [Online] Available at: https://www.thebureauinvestigates.com/stories/2026-03-26/the-hate-economy-in-action-how-an-ai-rapper-tried-to-cash-in-on-our-coverage

2.4. The Strengthened 2022 Code of Practice on Disinformation and Covid-19 Health Disinformation

By Giovanna da Custodia

The COVID-19 pandemic generated not only a global public health emergency but also what the World Health Organization described as an "infodemic": an overabundance of information, including false and misleading claims, that made it increasingly difficult for citizens to identify trustworthy health advice. Online platforms became central channels through which misinformation concerning COVID-19 vaccines, treatments, public health measures and the origins of the virus spread at unprecedented speed. Recognising the scale of this challenge, the European Commission introduced the Strengthened Code of Practice on Disinformation in June 2022, expanding the commitments contained in the original 2018 Code through stronger measures on transparency, fact-checking, researcher access, demonetisation and platform accountability. The Code represented the European Union's principal governance response to online disinformation during the latter stages of the COVID-19 pandemic.

Viewed through the Adaptive Cycle Framework, this case demonstrates both the achievements and limitations of the Code. While it strengthened coordination between platforms and public authorities and improved accountability mechanisms, its voluntary and predominantly reactive approach limited its capacity to fundamentally reshape online Covid dis(mis)information.

Exploitation **(r): Rapid Growth in an Information Vacuum.** The COVID-19 pandemic created an unprecedented information vacuum as scientific knowledge, public health guidance and government advice evolved rapidly. False claims concerning the origins of the virus, vaccine safety and unproven treatments spread quickly across major platforms, amplified by recommendation systems that rewarded highly engaging, emotionally charged content. The speed with which misleading information spread quickly outpaced the ability of public health authorities to verify and correct it, prompting the European Commission and the EEAS to characterise the phenomenon as an "infodemic" requiring coordinated action between governments, platforms and public health authorities.[21]

Conservation **(K): Entrenchment into a Self-Reinforcing Ecosystem.** What began as isolated false claims evolved into a stable, self-reinforcing disinformation ecosystem. COVID-19 misinformation became embedded within online communities bound by shared distrust of governments, public health institutions and mainstream media. Indeed, empirical research on COVID-era social networks has found that such communities functioned as genuine echo chambers, with information and engagement concentrated almost entirely within like-minded clusters and rarely crossing into or out of them, while anti-vaccine networks specifically have been found to be denser and more tightly organised than their pro-vaccine counterparts.[22] Rather than merely disseminating individual claims, platforms increasingly facilitated the formation of durable networks that reinforced existing beliefs through repeated exposure and community identity. The European Commission concluded that the

[21] European Commission and High Representative, 2020. Tackling COVID-19 disinformation – Getting the facts right. JOIN(2020) 8 final, 10 June 2020.

[22] Retweet-BERT polarization study, JMIRx Med, 2021; network-density study on anti-vaccine vs. pro-vaccine community structure, cited in Chen et al., "Misinformation and Polarization around COVID-19 vaccines in France, Germany, and Italy," arXiv, 2024.

original 2018 Code of Practice lacked sufficiently robust commitments, transparency and accountability to address this increasingly systemic challenge, prompting a comprehensive review of the framework.[23]

***Release* (Ω): Disruption through the Strengthened Code of Practice.** The European Commission sought to disrupt this entrenched ecosystem through the 2022 Strengthened Code of Practice on Disinformation. Developed jointly with major platforms, advertisers, fact-checkers, researchers and civil society organisations, the strengthened framework introduced 44 commitments and 128 specific measures, including reducing the monetisation of disinformation, strengthening fact-checking partnerships, improving researcher access to platform data, increasing transparency in political advertising, establishing a Transparency Centre, and introducing a permanent monitoring framework.[24] Together, these measures represented the EU's principal governance response to COVID-19 health disinformation, seeking to reduce the incentives, visibility and reach of harmful content through coordinated multi-stakeholder action.

***Reorganisation* (α): Adaptation Rather than Disappearance.** The disruption proved incomplete. As major platforms intensified moderation and reduced the visibility of COVID-19 misinformation, many communities reorganised rather than disappeared. Anti-vaccine and conspiracy networks migrated toward alternative and encrypted platforms, most notably Telegram, which remained outside the Code's signatory framework;[25] documented "lifeboat" strategies among leading anti-vaccine communicators specifically anticipated this shift, planning migration to less-moderated platforms in advance of anticipated account removals.[26] At the same time, influential actors reframed explicit anti-vaccine messaging around broader narratives of "medical freedom," natural health and institutional distrust, a pattern well documented among wellness influencers, whose anti-establishment messaging has been found to invoke themes of parental rights and distrust of pharmaceutical companies at substantially higher rates than comparable non-influencer accounts.[27] To reduce the likelihood of detection, communities also adopted coded language, intentional misspellings, emojis and euphemisms - for example, renaming groups to innocuous-sounding names or substituting emoji for flagged terms - allowing misleading content to continue circulating while evading automated moderation.[28] Rather than eliminating health disinformation, the intervention altered the environment in which it operated, enabling misinformation networks to preserve audiences and reseed similar narratives through new platforms, actors and communication strategies.

RECOMMENDATION(S):

If content-moderation-centred interventions share a structural tendency to displace disinformation rather than eliminate it, the practical implication for policy design is that reorganisation should be anticipated at the drafting stage rather than addressed only once displacement has already occurred. The Code's own history illustrates the cost of the reactive alternative: Telegram remained outside the signatory framework throughout the period under study, and it was not until the Code's integration into the Digital Services Act's co-regulatory

[23]European Commission, 2021. Guidance on Strengthening the Code of Practice on Disinformation. COM(2021) 262 final, May 2021; European Court of Auditors, 2021. Special Report: Disinformation affecting the EU.

[24]European Commission, 2022. 2022 Strengthened Code of Practice on Disinformation.

[25]European Commission, list of signatories to the 2022 Strengthened Code of Practice (Telegram absent); TechCrunch, 2022. EU unveils tougher industry Code to combat disinformation. June 2022.

[26]Center for Countering Digital Hate, 2020. The Anti-Vaxx Playbook. December 2020.

[27]O'Brien, G., Ganjigunta, R. & Dhillon, P.S., 2024. Wellness Influencer Responses to COVID-19 Vaccines on Social Media: A Longitudinal Observational Study. Journal of Medical Internet Research 26: e56651.

[28]PolitiFact, 2021. People are using coded language to avoid social media moderation. Is it working? November 2021; "Algospeak," Wikipedia.

framework in February 2025 that a mechanism existed to extend obligations beyond voluntary signatories.[29] A more preventative design would treat likely destination platforms as within scope from the outset, using the DSA's systemic risk assessment powers proactively against platforms hosting displaced communities rather than waiting for migration to be documented after the fact. A second implication concerns what moderation systems are built to detect. The persistence of "medical freedom" framing and wellness-oriented messaging succeeded precisely because it operated below the threshold of explicit anti-vaccine claims that automated systems and fact-checkers are calibrated to catch *(see n 27)*; a Code oriented toward reorganisation would therefore supplement claim-based detection with structural indicators tracking shifts in community rhetoric and framing over time, since a decline in flagged content is fully compatible with an undiminished audience for the same underlying narrative. A third implication follows from the documented use of coded language, renamed groups and emoji substitution to evade detection *(see n 28)*. Because this pattern recurs across misinformation and extremism contexts alike *(see n 26)*, it should be treated as a predictable feature of the adaptive cycle rather than a series of unforeseen loopholes to be patched retrospectively - favouring monitoring capacity built around pattern recognition (sudden co-occurrence of emoji with health topics, cluster renaming) over static keyword lists that this literature shows communities can route around within weeks of their introduction. None of this implies that the Strengthened Code was poorly conceived relative to comparable interventions elsewhere; it implies that its central limitation was shared with content-moderation-first governance generally, and that a Code designed with the adaptive cycle in mind - rather than with the exploitation or conservation phase alone in view - would build its reorganisation-stage response into its initial architecture rather than its subsequent revision.

2.5. Health Disinformation Targeting Older Adults: Covid Era Digital Vulnerabilities

By Holly Barnett

Older adults experienced one of the most sustained and harmful disinformation cycles during the COVID-19 pandemic. This case study shows how UK-based older adults were structurally exposed to health misinformation through increased cognitive load, social isolation, and platform-level algorithmic design. Mapped onto the adaptive cycle, the system shows a disinformation environment that leaves older adults vulnerable during new health scares, such as the 2026 meningitis cases.

***Exploitation* (r).** The early pandemic opened an information vacuum in which fear, uncertainty, and contradictory international guidance created ideal conditions for falsehood to spread. As people turned to online sources for rapid answers before official updates were available, conflicting international approaches to lockdowns, masks, and treatments deepened confusion and drove search-based exposure to curated feeds. Due to social isolation, Facebook community groups and WhatsApp family chats acted as high-trust environments where social proof amplified rumours. Meanwhile, widely shared "miracle cure" claims, including those

[29] European Commission and European Board for Digital Services, 2025. Endorsement of the integration of the 2022 Code of Practice on Disinformation into the DSA framework as a Code of Conduct. 13 February 2025.

pushed by influential figures, played on older adults' heightened health anxiety. Structural vulnerabilities such as higher perceived risk, lower digital skills and confidence, and strong trust in familiar contacts made disinformation exposure likely. Engagement-optimised feeds then pushed provocative, emotional content, making disinformation feel urgent and personal.

Conservation **(K).** By mid-2020, COVID misinformation hardened into a set of self-reinforcing narratives. Anti-vaccine claims painted the vaccines as "rushed," "experimental," or "dangerous," while anti-lockdown messages framed restrictions as threats to personal freedom or wellbeing. At the same time, distrust toward government, the NHS, and public health bodies deepened, and conspiracy theories about 5G towers, bioweapons, or microchips gained traction through repetition. As online communication became central to maintaining social connection during isolation, older adults experienced heightened exposure to repetitive disinformation, and psychological mechanisms helped sustain these narratives: the illusory truth effect made repeated claims feel accurate; social proof increased perceived legitimacy when peers reshared content; and the "better safe than sorry" principle outlines how older adults can accept health misinformation to avoid perceived risks.[30]

Release **(Ω).** Disruption arrived reactively through platform interventions and public health communication efforts. WhatsApp introduced strict forwarding limits[31] so "highly forwarded" messages could be sent to only one chat at a time, and social media companies backed targeted campaigns to counter vaccine misinformation among ethnic minority communities.[32] Twitter suspended accounts under its COVID-19 Misleading Information Policy,[33] and fact-checking organisations rapidly expanded their capacity to debunk content.[34]

Reorganisation **(α).** Algorithmic amplification of health content, even when it's inaccurate, still isn't penalised, and older adults remain under-researched in digital literacy and misinformation-resilience policy. Recent events, like the meningitis cases in early 2026, showed how quickly vaccine-related falsehoods can spread again. The main drivers, engagement-optimised feeds and closed-network forwarding, also remain in place. Without structural changes, the system keeps defaulting back into exploitation whenever a new health scare emerges.

RECOMMENDATION(S):

1. Impose limits on algorithmic amplification during public-health incidents, slowing the spread of inaccurate health content before repetition produces the illusory truth effect.
2. Require transparency for closed-network forwarding chains during health emergencies so public-health teams can understand how rumours travel.
3. Penalise monetisation of health-related falsehoods, reducing incentives for creators who profit from misleading content.

[30]Zhou, J., Xiang, H. & Xie, B., 2022. Better safe than sorry: a study on older adults' credibility judgments and spreading of health misinformation. Universal Access in the Information Society. https://doi.org/10.1007/s10209-022-00899-3

[31]MIT Technology Review, 2020. WhatsApp is limiting message forwarding to combat coronavirus misinformation. https://www.technologyreview.com/2020/04/07/998517/whatsapp-limits-message-forwarding-combat-coronavirus-misinformation/

[32]UK Government, 2020. Government targets false vaccine information on social media. https://www.gov.uk/government/news/government-targets-false-vaccine-information-on-social-media

[33]Evidence Hub. Accounts Suspended for Violations of Twitter COVID-19 Misleading Information Policy. https://evidencehub.net/chart/accounts-suspended-for-violations-of-twitter-covid-19-misleading-information-policy-295.0

[34]Full Fact. Covid 19 fact checks. https://fullfact.org/covid-19/

4. Fund targeted digital-literacy programmes for older adults, focusing on confidence with online health information, recognising manipulative content, and navigating high-trust social spaces like online community groups.
5. Support community-based resources (GPs, pharmacists, local health charities) to deliver rapid, trusted corrections during outbreaks, as these are the sources older adults actually use.

2.6. AI-Generated Disinformation in the India-Pakistan Conflict, May 2025

By Sakshee Singh

The four-day military confrontation between India and Pakistan in May 2025 demonstrates the ACF operating at a different scale: not a single incident, but a sustained, cross-border information contest, occurring in the near-complete absence of any regulatory framework governing synthetic conflict media specifically.

***Exploitation* (r).** Following the April 22 terror attack in Pahalgam and India's launch of Operation Sindoor on May 7, an information vacuum opened alongside the military one. Deepfakes proliferated within hours: a fabricated video showed a Pakistani military spokesperson admitting to lost jets, and another depicted Pakistan's Prime Minister conceding defeat. A widely shared AI-generated image of a supposedly bombed stadium accumulated millions of views without correction or platform labelling.[35] BOOM, an India fact-checking outlet, found that over two-thirds of its checks that month concerned the operation, while India's Chief of Defence Staff acknowledged that a substantial share of operational time was spent debunking fabricated claims.[36]

***Conservation* (K).** This is not a novel configuration but a recurring one, with commentators tracing similar dynamics back to the 2019 Balakot standoff. The entrenchment is structural: platforms including X, WhatsApp, and television news channels function as an integrated amplification circuit in which state-linked actors, verified accounts, and commercial broadcasters all contribute to the same pattern, reinforced by delayed content moderation.[37]

***Release* (Ω).** No meaningful release phase occurred. What existed on both sides was general fact-checking capacity, but its application followed the same partisan pattern as the disinformation itself, each side's mechanism verified only claims damaging to the other, rather than applying a consistent standard.[38] The absence, then, is not of institutional infrastructure but of any rule specifically governing synthetic and AI-generated conflict content, at either the state or platform level, the pattern of exploitation was met with a mirrored pattern of correction, not a disruption of it.

---

[35] IFJ, 2026. #IFJBlog: Operation Sindoor and the Two Wars – Real and Imagined. [Online] Available at: https://www.ifj.org/media-centre/news/detail/category/articles-blogs/article/ifjblog-operation-sindoor-and-the-two-wars-real-and-imagined

[36] The News Minute, 2025. Deepfakes, voice clones and AI images amplified disinformation on India-Pak conflict. [Online] Available at: https://www.thenewsminute.com/news/deepfakes-voice-clones-and-ai-images-amplified-disinformation-on-india-pak-conflict

[37] Bulletin of the Atomic Scientists, 2026. Disinformation and deepfakes: Improving crisis communications in India and Pakistan. [Online] Available at: https://thebulletin.org/2026/01/disinformation-and-deepfakes-improving-crisis-communications-in-india-and-pakistan/

[38] Center for the Study of Organized Hate, 2025. Inside the Misinformation and Disinformation War Between India and Pakistan. [Online] Available at: https://www.csohate.org/2025/05/16/india-pakistan-digital-war/

Reorganisation **(α).** Consequently, no reorganisation has taken place. The structural conditions behind the surge, an unregulated synthetic-media environment and partisan verification, remain intact, meaning a comparable surge should be expected in any future escalation, not treated as exceptional.

RECOMMENDATION(S):

Given the lack of trust between the two states, bilateral fact-checking bodies are unlikely. More realistic is pressure on platforms operating in both countries to apply uniform labelling of synthetic conflict media, regardless of which side a claim favours, paired with use of existing military hotlines for information-specific de-escalation, rather than new institutions requiring cooperation neither government has shown.

---

3. CONCLUSION: Key Takeaways

By Daniel Lewis

It is not proposed that the Adaptive Cycle Framework is a drafting instrument. It produces no clauses, no thresholds, no enforcement architecture, and it cannot adjudicate between competing rights or price the cost of an intervention. This report does not suggest policy papers should be written in the vocabulary of exploitation, conservation, release and reorganisation. Its value sits upstream of that, at the planning stage, where the operative questions are focused on *where is the system now*, *when will it be movable*, and *what would have to change for it not to return*. Used in that position, give takeaways follow from the cases above.

1. **The framework diagnoses position; it does not supply the remedy.**

Ofcom's crisis-response protocol (1.1) is a completely drafted instrument that lands in the wrong place. It covers illegal content and content harmful to children, expressly excludes disinformation campaigns, and leaves the trigger to platforms' own judgement. Very little in the drafting is defective on its own terms. What the ACF supplies is the prior question the drafting was never asked to answer: which phase is this aimed at, and does it touch the driver that regenerates the cycle? The answer: a reorganisation-stage response to a system whose exploitation driver survives intact – is available before a single clause is written and is the reason the cycle has returned three summers running.

2. **Release is the window, and it is short.**

The release phase is where entrenchment is weakest and political attention highest, and it is consistently where the cases show the opportunity being lost rather than taken. The collapse of Silicon Valley Bank (1.2) produced an explicit official concession that the speed of bank runs had changed, yet MAR and MiFID II remain calibrated to a slower, attributable form of manipulation. The Belfast recommendation makes the same point, positively, that the value of acting during disruption is that the momentum surrounding the incident is itself a policy resource. This is the planning implication that matters most. Because release arrives faster than legislation moves, the instrument has to exist, at least in draft, *before* the shock that makes it passable. Identifying the likely shock is a hypothesis, not a forecast – but it is a hypothesis that can be made in advance and revised, which reactive policymaking cannot.

3. **Reorganisation should be designed for, not discovered.**

The Strengthened Code (1.3) is the clearest demonstration. Migration to Telegram, the reframing of explicit anti-vaccine claims as "medical freedom", and the adoption of coded language were not unforeseeable; migration to less-moderated platforms were documented as a deliberate strategy in advance of the removals that prompted it. Yet, not mechanism existed to extend obligations beyond voluntary signatories until the Code's integration into the DSA framework in February 2025. The gap between a successful disruption and a structural response to its consequences is where the system reseeds. Planning-stage use of the framework converts this into a design question asked at the outset: where does this go when it works?

4. **Escape is defined by the driver, not by the content.**

The same variable survives every case in this report. Amplification remains unpenalized for the financial models behind it in Northern Ireland (1.1); the advertising-driven business model survives the SVB shock intact (1.2); the funding and royalty structure sustaining a synthetic persona sits outside both content and electoral-transparency categories (1.3); algorithmic amplification of inaccurate health content is still not penalised (1.4); and platforms, broadcasters and state-linked accounts operate as an integrated amplification circuit across the India-Pakistan escalation (1.6). That a single untouched structural condition recurs across civil unrest, financial stability, electoral integrity, public health and interstate conflict is this report's strongest finding, and it is visible only because one framework was applied consistently to five unrelated domains. The corollary is the on flagged in the definitions: resilience is value-neutral. The objective is to reduce the resilience of the disinformation environment while building resilience into the counter-measures – and a policy that leaves the driver untouched has done neither.

5. **The framework's planning value is that it commits to expectations.**

Two of the cases here are unresolved, and the ACF still yields determinate statements about them. Danny Bones (1.2) has produced no binding reorganisation, which supports the expectation that a replicable production model diffuses rather than a single persona persisting. The India-Pakistan escalation (1.6) had no meaningful release phase at all, which supports treating a comparable surge in any future confrontation as the baseline rather than the exception. This is the anticipatory function doing real work, where it converts a "this may occur" claim about which specific conditions must change for it not to, and it supplies a negative test for proposals – an intervention that leaves the phase diagnosis unchanged should be expected to fail, and can be identified as such before it is enacted.

Limits. The ACF is a heuristic, not a model. Phase boundaries are clean analytically and blurred empirically; the SVB case shows that a release may belong to the host system rather than to the disinformation environment riding on it, and the two do not always move together. The framework cannot say how long a conservation phase will hold, nor how large an intervention must be to prevent reseeding it. It is also retrofittable – almost any sequence of events can be narrated into four phases after the fact – which is why its discipline depends on being applied while outcomes are still open, as in 1.2 and 1.6 above.

None of this displaces the central claim. Used at the planning stage, the ACF forces three questions that current disinformation policy has largely not asked:

- Where is this system now?
- When will it next be movable?
- What has to be removed for it not to reconstitute?

Policy capable of answering those before drafting starts *begins* preventatively.

Policy paper by



Daniel Lewis

Daniel Lewis is Policy Fellow at the Council for Countering Online Disinformation (CCOD), working on online safety and digital disinformation. He holds a BA in Politics and International Relations from Royal Holloway, University of London, where he wrote his dissertation applying the capabilities approach to online safety, and is studying for an MSc in Development Studies at the London School of Economics. He is a Research Assistant at the London Social Media Observatory and serves as Amnesty International UK's Country Coordinator for Thailand and Laos.

Contributed by



Giovanna da Custodia

Giovanna da Custódia is a Policy Fellow at the Council for Countering Online Disinformation (CCOD), working on online safety and platform governance policy. She also works as a Research Analyst at Chambers and Partners, where she focuses on the TMT markets across the EMEA region. Giovanna holds a Distinction in an MSc in Public Policy from the London School of Economics and a BA in Politics and Government from Sciences Po Paris. Her research has informed publications for the European Parliament, UNESCO, and the Mozilla Foundation, and she has contributed to work on AI governance, platform accountability, and digital safety.



Holly Barnett

Holly Barnett is a PhD candidate at Lancaster University, researching how psychological interventions support misinformation detection across age groups. Her work focuses particularly on older adults, examining how cognitive and confidence-related factors influence their ability to identify false news content. She aims to develop strategies that help people detect misinformation more easily and engage with emerging digital technologies with greater confidence.



James Rice

James Rice is a political data scientist and PhD candidate in the Sustainable Transitions Programme at the University of Essex, with an interdisciplinary academic background spanning degrees at the London School of Economics and the University of Edinburgh. Alongside his doctoral work, James teaches at the UCL Department of Political Science and has engaged with research and specialist training across institutions including Oxford's DPIR and Economics Departments, the Mercatus Center, the LSE Hayek Programme, and the Essex Summer School in Data Science, reflecting a broader interest in bringing ideas from economics, political theory, and AI to bear on some of the most difficult problems in contemporary politics.



Sakshee Singh

Sakshee Singh is a development professional specialising in knowledge systems and stakeholder engagement. She holds an MSc in Development Studies from the London School of Economics and Political Science and has experience across international organisations, policy research, and community-focused initiatives in South Asia, with a particular interest in digital resilience and the intersection of technology and society.

About the CCOD

The Council for Countering Online Disinformation (CCOD) is a member organisation and advocacy group designed to fight back against the growing challenge posed by online disinformation.

With the world entering a new technological era, influencing opinion and human decision-making through false and misleading content has never been so easy. Ofcom now estimates that more than four in ten have been exposed to online mis/disinformation, with seven in ten of those exposed having seen it online. Meanwhile, according to the Reuters Institute for the Study of Journalism, approximately six in ten are worried globally about what is real and what is fake when it comes to news online. The matter has been made worse by the deregulatory stance taken by the Trump administration, which wishes to see regulatory requirements in both the UK and the EU lifted from major U.S. tech companies.

The CCOD is the only actor of its kind aimed at defending the information integrity field and championing all its various stakeholders to policymakers and the UK government. Founded by the minds behind Polis Analysis, a non-profit organisation, the CCOD draws upon the considerable advocacy and research experience of the Working Team and the technical and political insight of an expert Advisory Board, whose backgrounds range from politics to academia, journalism, and regulatory bodies. Working together, these two entities enable a coordinated, informed, and impactful approach to tackling online disinformation, which delivers on our core mission.