Disinformation futures

How we will build mental resilience against disinformation

A experiential futures case study

Introduction and context Concern over disinformation and limitations of regulation

- Increased concern over disinformation given its material impact on the ability of societies and governments to effectively co-create and implement collective responses to global issues
- The increased adoption of social media platforms and technology to communicate and share information has enabled disinformation to be spread more easily, widely and quickly (Southwell et al., 2018)
- Regulation of disinformation generation and sharing is difficult companies like Meta have near-monopolies over how media is shared and can significant leverage against government interference (Pickard et al., 2020); regulations are also complicated by discussions over individual freedom of speech and expression (Pielemeier, 2020) and debating whether regulating content is aligned a democratic society (Tan & Sijie, 2020)
- With disinformation supply difficult to regulate, some researchers have turned their attention to understanding how individuals and communities can become more resilient to disinformation

Building mental resilience A focus on information consumers vs content generators

- Interventions to build individual/societal resilience to disinformation ranges from general media literacy in the classroom to more direct interventions that aim to "inoculate" against disinformation (Roozenbeek et al., 2022)
- "Inoculation" methods are more direct and specific in helping people identify and understand the intent behind disinformation; for example, asking individuals to create and then deconstruct disinformation on highly political topics (e.g. immigration)
- Like content regulation, these more direct methods can be perceived to be highly personal, intervening with individual thinking, expression and psychology and make them difficult to implement
- Therefore, testing public perception and receptivity to methods of building resilience against disinformation can help policymakers and practitioners understand their likelihood of adoption.

Origins of disinformation From paper to digital

- False information is broadly categorised as misinformation a factually incorrect idea, notion or theory that spreads without any malice or negative intent behind its creation (Fallis, 2014). An example of this is the pop culture emergence of the "Mandela effect", where individuals insist that popular logos, icons and names were different when they were young (Figure 1)
- In contrast, disinformation is misinformation that was created and spread with the intent to deceive and create a desired reaction in society even if that desired reaction is chaos and conflict (Fallis, 2014),
- Disinformation is often attributed to political and military origins: from the Cold War KGB propaganda department (Taylor, 2016) to CIA interventions in South American elections (Ferreria, 2008) however, this was done mostly via print or radio
- Digital society has disinformation to be created more quickly, by more people and to spread at faster rates
- There is also increased efficacy of disinformation through the creation of bots, false profile building and rapid message testing

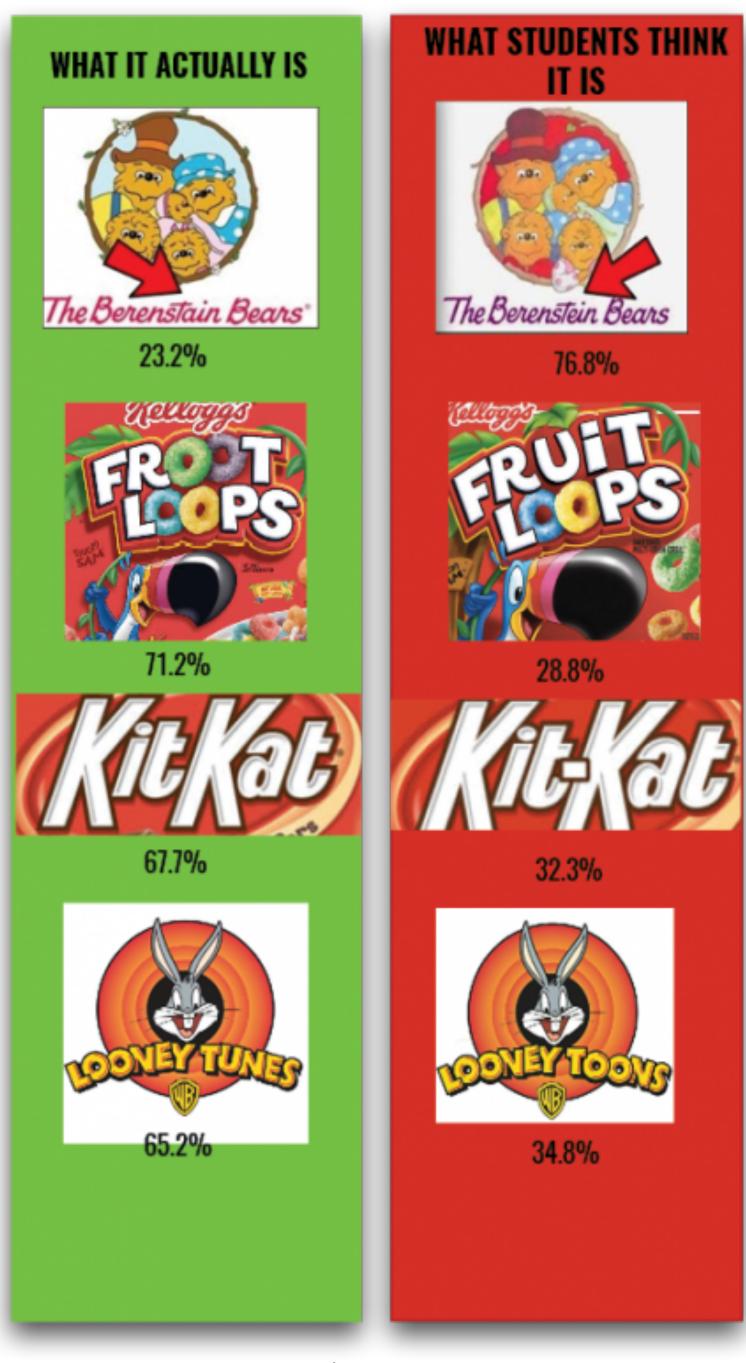


Figure 1

Impacts of disinformation Impacts on the individual

- Most individuals cannot tell the difference between facts and disinformation when they are presented in similar ways, which makes the increased volume, efficacy and spread of disinformation a significant challenge
- Even more challenging, disinformation can modify human neurology and behaviour. For example, disinformation can manipulate memory recall, changing an individual's memories of past events to create a sound and logical path to justify behaviours and belief in current disinformation (Greenspan & Loftus, 2021).
- This manipulation of memory can create a particularly challenging causal loop (Figure 2) for the pursuit of preserving of true and factual information and combating disinformation. As disinformation can manipulate memories, it can also shift the worldview of individuals, making them less resilient and more susceptible to similar disinformation.
- Conversely, these individuals build more resilience and resistance to any facts that contradict the narrative of disinformation, further supporting this reinforcing loop.

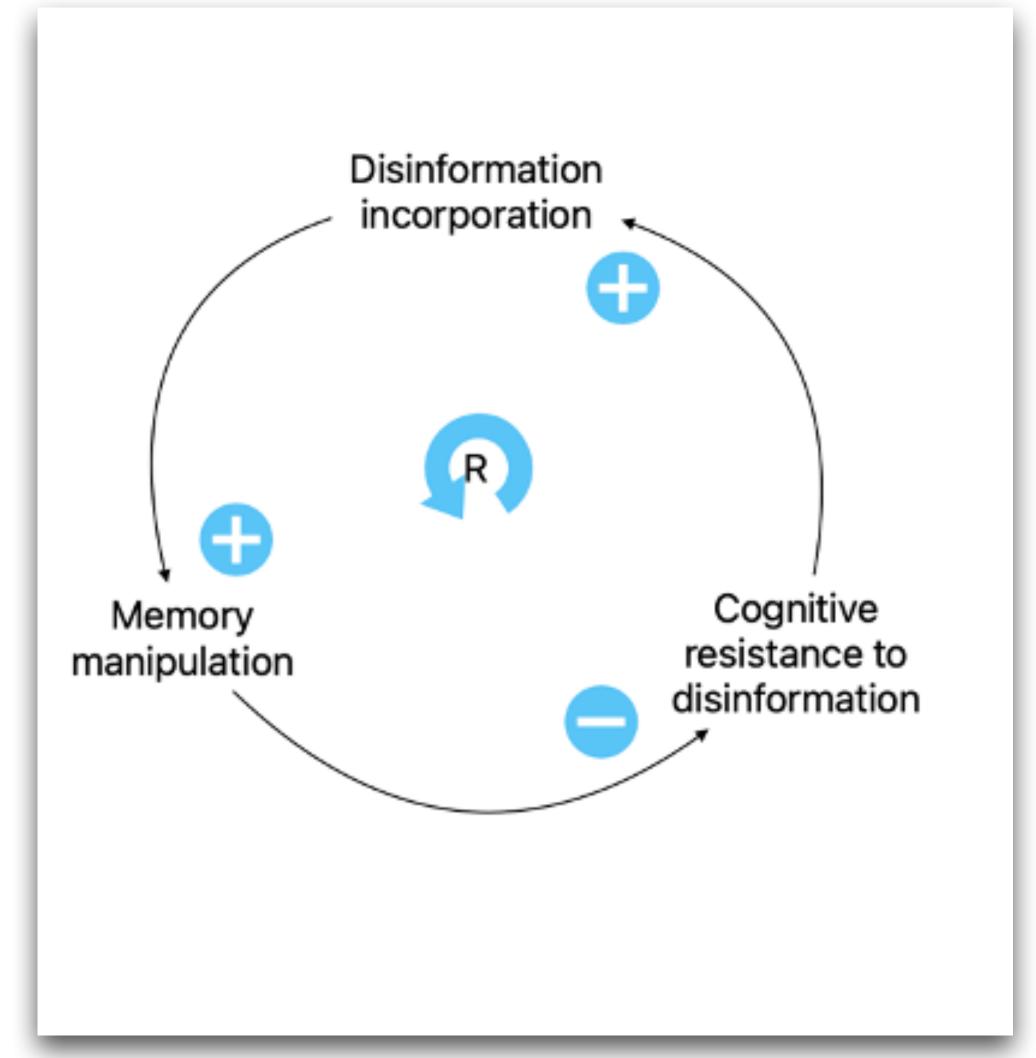


Figure 2

Causal loop diagram demonstrating how consumption of disinformation reinforces openness and reduce friction to incorporating further disinformation.

Impacts of disinformation Impacts on communities

- In communities, the impacts of disinformation can be exacerbated through mutually reinforcing causal loops (Figure 3)
- The reinforcing loop of memory manipulation on the uptake of disinformation (R1) is further reinforced by cultural and language divergence in society (R3)
- Memory manipulation from disinformation can further drive divergence in world reviews and drive isolation (Liv & Greenbaum, 2020), further reducing diversity of thought and perspectives (R3).
- Isolated communities may form more homogenous culture and language and communications styles that may be specific to them (R3)
- With less diversity of thought and cognitive friction as a result, these communities become more susceptible to disinformation that nudges them to become more extreme in their worldviews (R3)

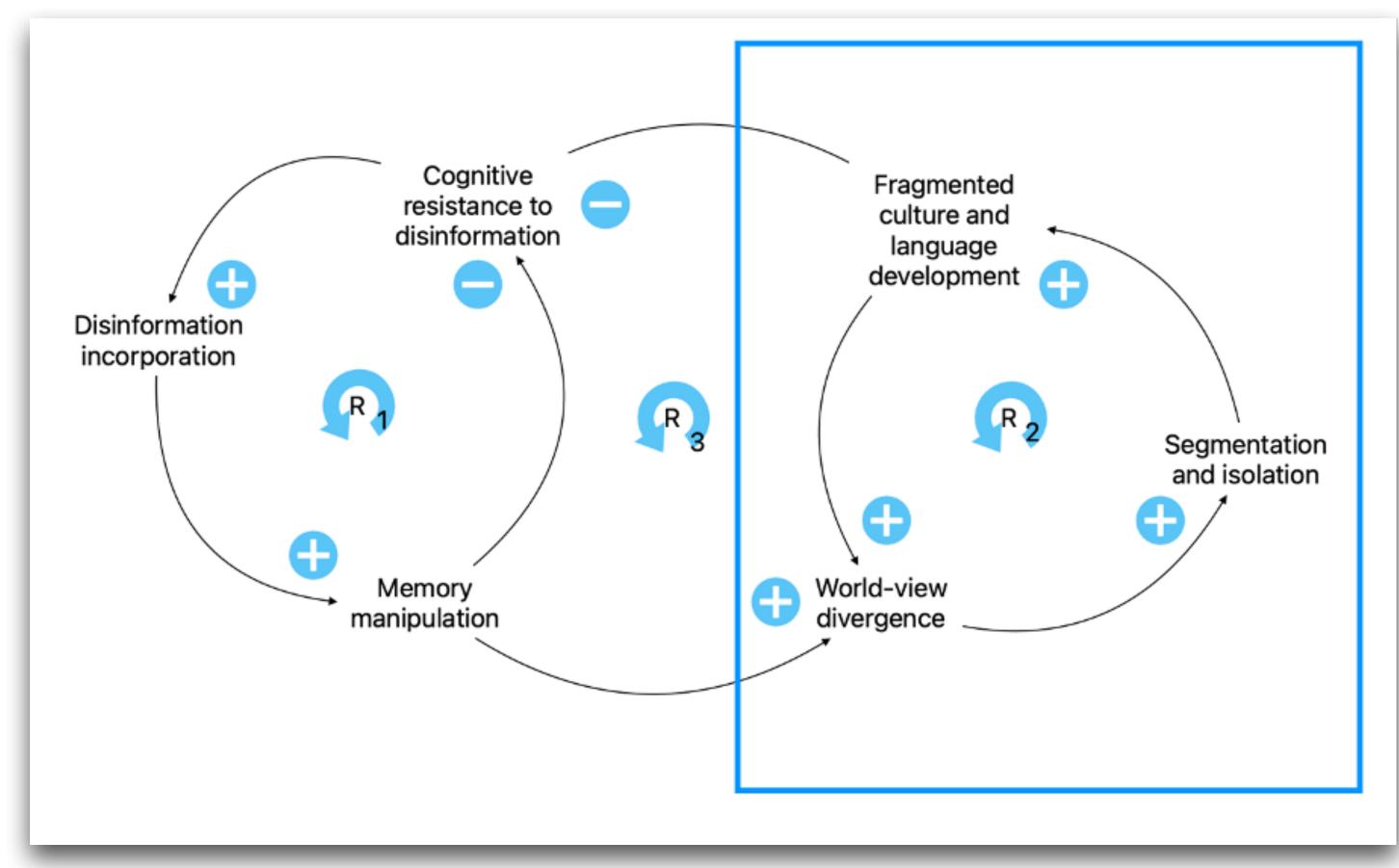


Figure 3

Casual loop diagram demonstrating the community level impacts and effects of disinformation on cultural divergence and reinforcement.

Interventions

Generation and consumption

- Disinformation interventions focus on two key areas of the causal loop (Figure 5): preventing the fragmentation of isolation of society by limiting the generation and dissemination of disinformation (F1), and increasing cognitive resistance to the incorporation of disinformation (F2)
- In the first focus area (F1), there are a wide range of solutions from direct limits on what can be expressed on the internet in the EU (Wood & Ravel, 2017) to letting large social media players self regulate in the US (Lorenz, 2022)
- In the second focus area (F2), interventions are often referred to as "inoculation" or "prebunking", referring to their preventative intent; however, these are less tested and studied with little insight into their public receptivity

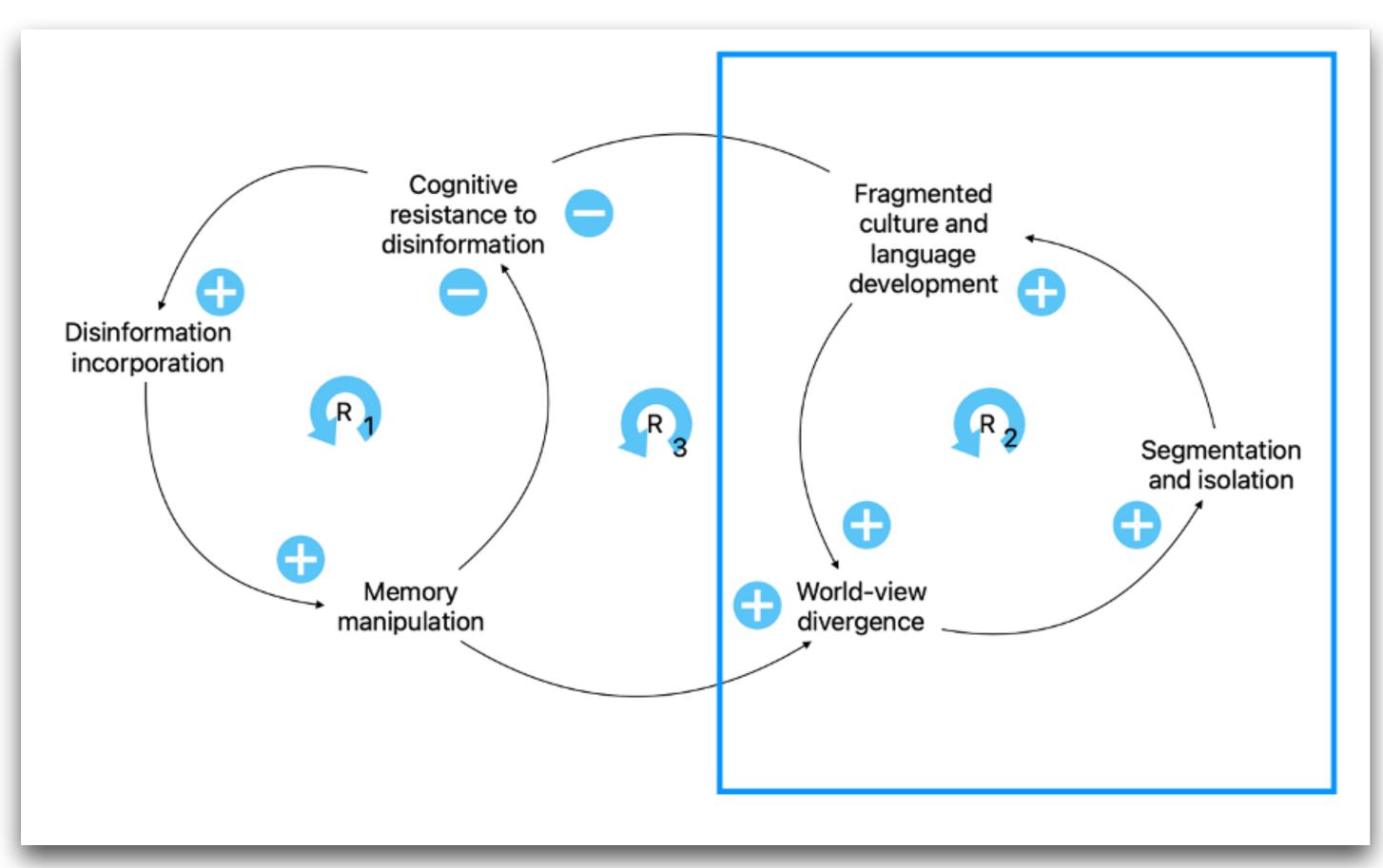


Figure 5

Casual loop diagram demonstrating the community level impacts and effects of disinformation on cultural divergence and reinforcement.

Methods

To assess public receptivity to future methods to build mental resilience against disinformation

- 1. Expert interviews and literature review on disinformation;
- 2. Driver and trends analysis;
- 3. Futures scenario creation and experiential futures; and
- 4. Data analysis and comparison to literature

Trend analysis From literature reviews and interviews

Category

Trend

Category	Trend
Social	Digital nationalism The generation of and reactions to disinformation reveal new sets of identities, communities, and social bonds outside of countries, ethnicities or current religions and secular institutions. For example, fans of the Korean pop music band "BTS" call themselves "ARMY" and will often join together to vote for the band in contests, influence business decisions and also swarm on disinformation about the band (Lee et al., 2022). Recent studies on media consumption and disinformation during the Ukrainian-Russian war discovered Europe and the UK were more likely to read and consume media from Russian sources, suggesting openness to ideas, disinformation and identity outside of their own national borders (Kling et al., 2022). This suggests that individuals, feeling part of a larger community, will participate in the generation, consumption and offensive against disinformation without direct government intervention.
Technology	Digital arm's race Given the challenges of disinformation, governments are actively involving the use of AI and machine learning in policy discussions (Berkowitz, 2020) and decision-making (Meissner & Keding, 2021) to better understand its impacts. This also helps governments better understand and sensitise to how these technologies can be used to generate information and influence decision-making. For example, the Finnish government has established a Committee of the Future that helps the government identify future themes and explore policy routes and alternatives in preparation (Parliament of Finland, 2018). The United States Department of Defense also continues to operate the Defense Information School which trains military personnel in public affairs, social media, journalism and public broadcasting (Department of Defense, 2020). This represents an interest from governments to better understand problems like disinformation and become more active in a regulatory, educational or advisory role.

,
a
se
nd
ot
ıs.
+
re
ect
ı
sers
nd
<i>'</i> .
he
nis
a econ — all realin y

Drivers and their influence

Drivers that more broadly affect the regulation of digital spaces, use and understand of technology and how information is created, shared and consumed and shape trends

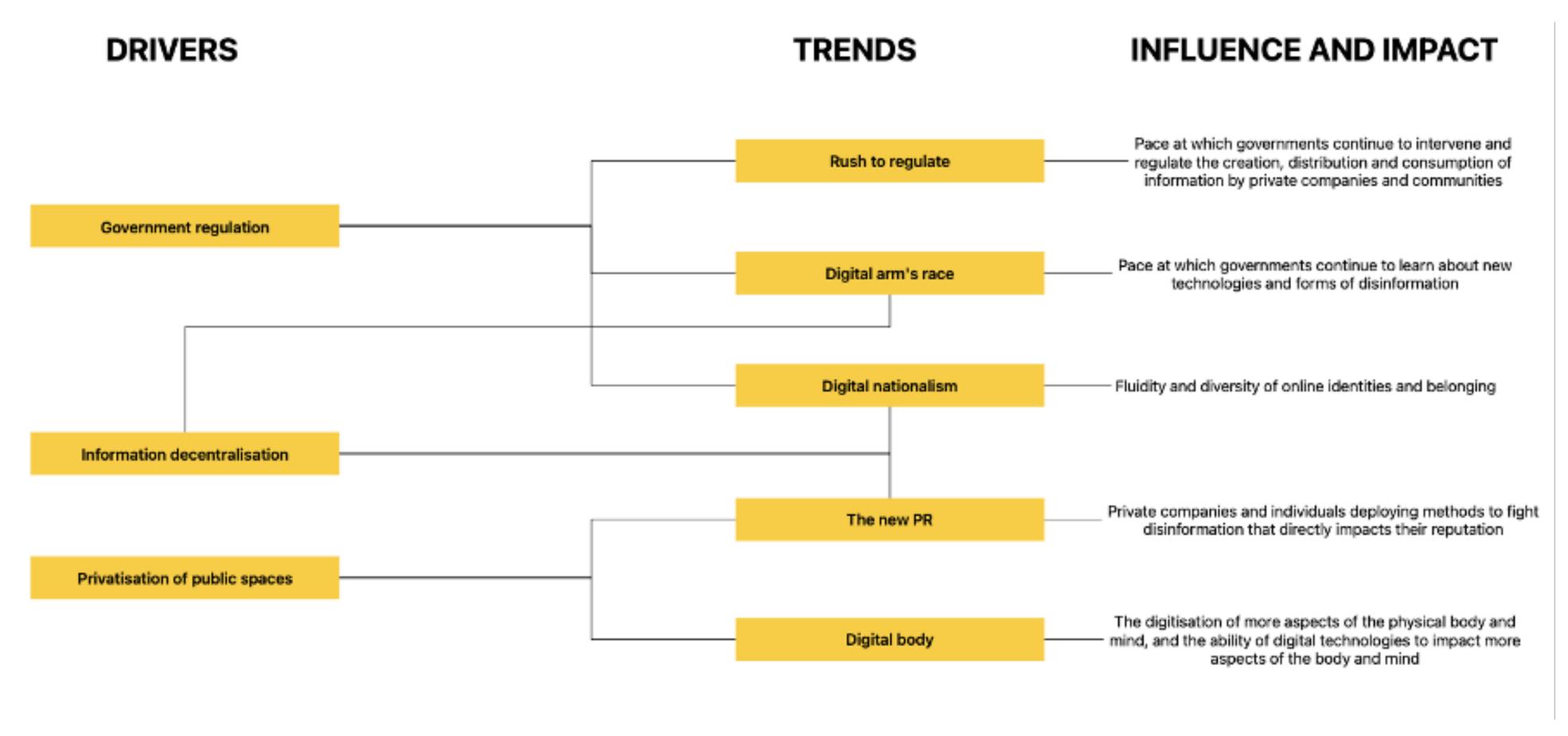


Figure 4

Drivers behind current trends and their influence on disinformation creation, spread and mitigation. Drivers can influence multiple trends and trends can be influenced by multiple drivers

Drivers analysisFrom literature reviews and interviews

Driver	Explanation
Government content regulations	This driver represents the on-going effort of governments to regulate access to the internet and content distributed on the internet. One polarity of government control includes disconnecting nations from the global network, eliminating internet connections and cellular signals or enforcing the use of a national intranet. On the other polarity, there is less government regulation of content and access; instead, the regulation of information is determined by the social media user base and culture of its users or its owner. Governments in these environments prefer to create educational initiatives in media literacy or rely on self-regulation by industry to government standards.
Information decentralisation	Communications and content-sharing via the internet has fragmented the way information is generated, shared and agreed-upon as truth in a highly digitalized world. Whereas information was previously declared as truth by newspapers, large media companies and academic journals, the emerging polarity follows a more networked approach. For example, users will conduct their own research or experiments and share the results via videos (e.g YouTube) or written content (e.g. Twitter, Reddit). Other users can comment directly or make content that analyses the veracity or methods of the original post – readers or users can make their own decisions about the quality of the original study and information as a result. A similar process can happen where content creators take journal articles or government reports and deconstruct them with their personal expertise and opinions. In both scenarios, users respond to the clout of the content creator versus their associations with more traditional institutions – this creates phenomena like widely publicised conflicts between YouTube content creators.
Privatisation of public spaces	There has been increasing pressure to privatise the ownership over newly created spaces, from the physical to the digital. Since the 1950s, there have been discussions and debates about how previously public spaces in urban areas have become more privatised or commercialised, despite still being open to the general public (Devereux & Littlefield, 2017). More recently, interactions that would typically take place in public spaces now occur in virtual ones – from online gaming in MMOs to virtual meetings over video conference calls or chat platforms, these new digital spaces originated as private and commercial spaces. However, these spaces also provide more personalisation as individuals can choose unique usernames, avatars or special video effects – expressions that were not possible or less accessible in the physical and public realms.

Mapping critical uncertainties Narrowing our field of impact

- To narrow the field of exploration for this study's experiential futures, the importance and uncertainty of these drivers of change were mapped to better understand their role in the futures of interventions in building mental resilience (Figure 6)
- Broadly, there was less certainty around how governments would regulate disinformation and develop policy or educational programs, but they are highly important and influential to individuals, companies and societies
- Boundaries and agency around the digital-self are also very important as individuals continue to live in increasingly digital societies, more aspects of their body and mind will be digitised
- However, there is high uncertainty about how individuals will be able to limit and control their exposure or resilience to disinformation as new digital channels continue to emerge before we can understand existing ones - for example, the eagerness to adopt crypto curries and virtual reality
- While the certainty information decentralisation and weaponisation in building mental resilience interventions is high, its impact is less clear
- With more decentralisation, it is more difficult for societies to establish a common agreement on which sources of information are most reliable this decentralisation would follow the trends and technologies established in the drivers and trends identified above
- Therefore, the two key drivers selected for futures creation were around government regulation and the speed and safety around increasing the scope of the digital self/technology adoption

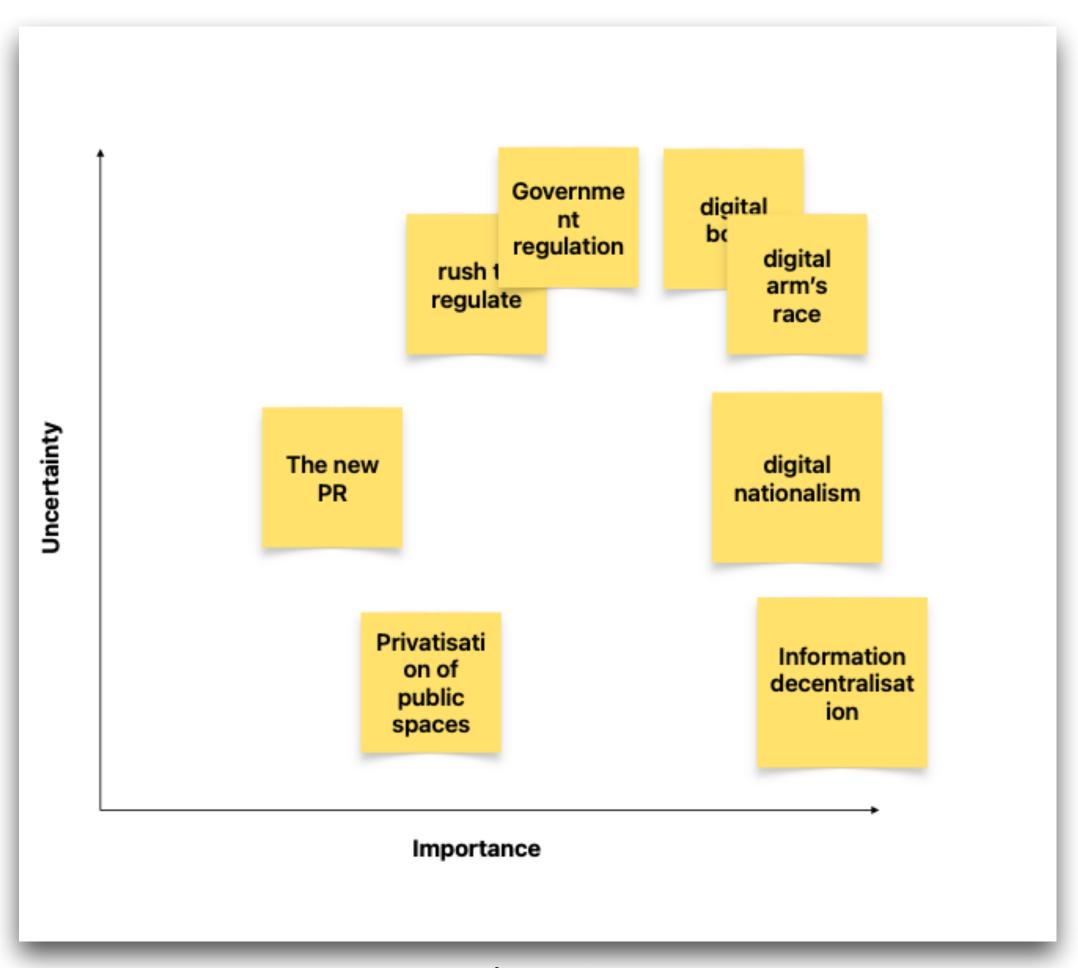


Figure 6

Mapping of the trends and drivers around mental resilience development by uncertainty of their future and importance to implementation of mental resilience intervention policies.

Futures scenarios Four quadrants

- A 2x2 was used to map our four futures:
 - The axis on government regulation explores a polarity between more heavy-handed, direct interventions to a more laissez-faire educational approach
 - The other axis focuses on the agency of individuals over their mind, body and data in digital spaces – from a more explicit ability to set boundaries in emerging and existing technologies to a less literate and cautious approach to embracing and using new tech
- From this, four scenarios were developed for this study and participant testing: Ministry of Technology, Digital community, Pay to protect and Correct and rehabilitate (Figure 7)
- Of these four scenarios, three (Ministry of Technology, Digital community and Pay to protect) were turned into physical installations experiential futures as they represented the largest delta from our current state environment

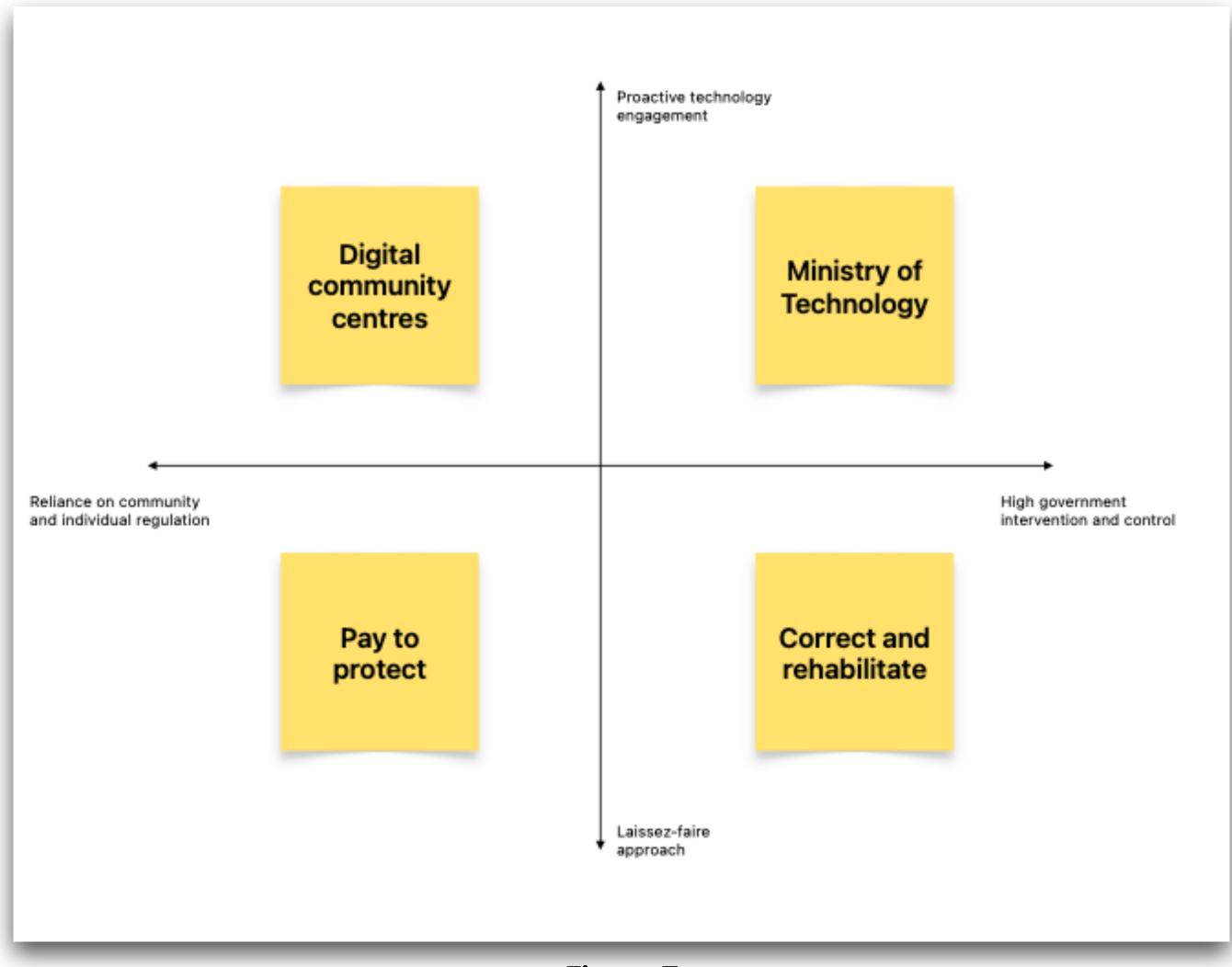


Figure 7

A 2x2 matrix exploring the polarities of government intervention and proactive understanding of emerging technologies to create futures scenarios.

Experiential futures

Audio-visual immersive environments

- The use of these installations and experiential futures is grounded in the philosophy of design fiction a way to telling speculative stories through objects that help stretch the mind beyond the practicalities of science, but still grounded in the realities of science (Bleecker, 2022)
- This method enables participants to explore scenarios beyond the limits of what they see and hear day-to-day about the future (e.g. via the news, websites), through guiding them to stretch their imagination through the use of concrete artefacts created to specifically represent futures in more extreme and distant horizons
- To effectively create experiential futures, the creation of these scenarios installations guided by the POEMS framework (Crawford, 2017), which involves carefully identifying how the objects, environments, messages and services used can reinforce the context and tone of each future - including a video component
- Visitors of the installations (Figure 8) were asked to complete a short survey to assess their receptivity, preference and belief in each of the futures presented.

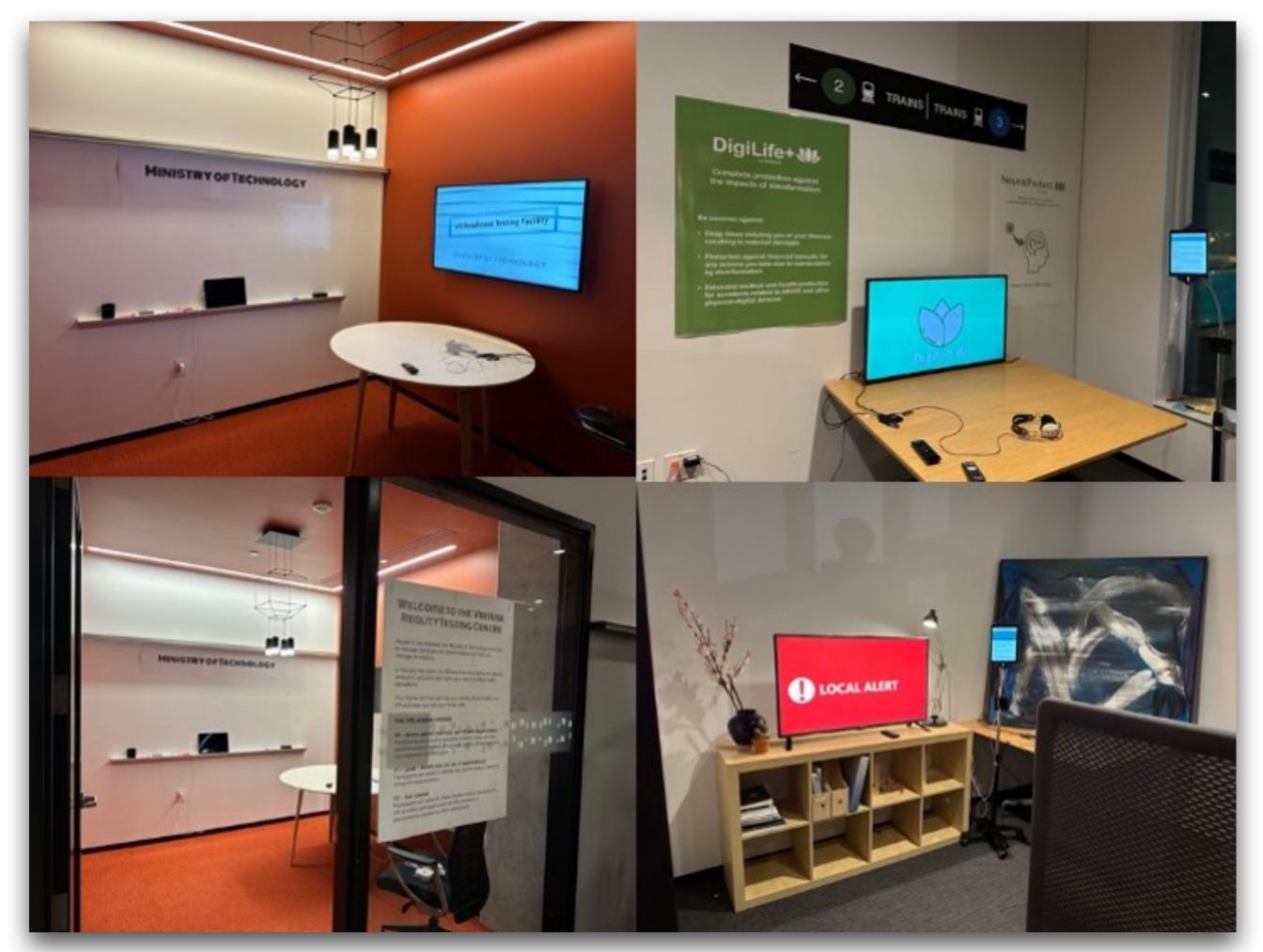


Figure 8

Photos of the experiential futures installations. The scenarios represented are: Ministry of Technology (top and bottom left), Pay to protect (top right) and Digital community (bottom right)

Ministry of technology High government intervention, proactive engagement

High government intervention, proactive engagement with new technology



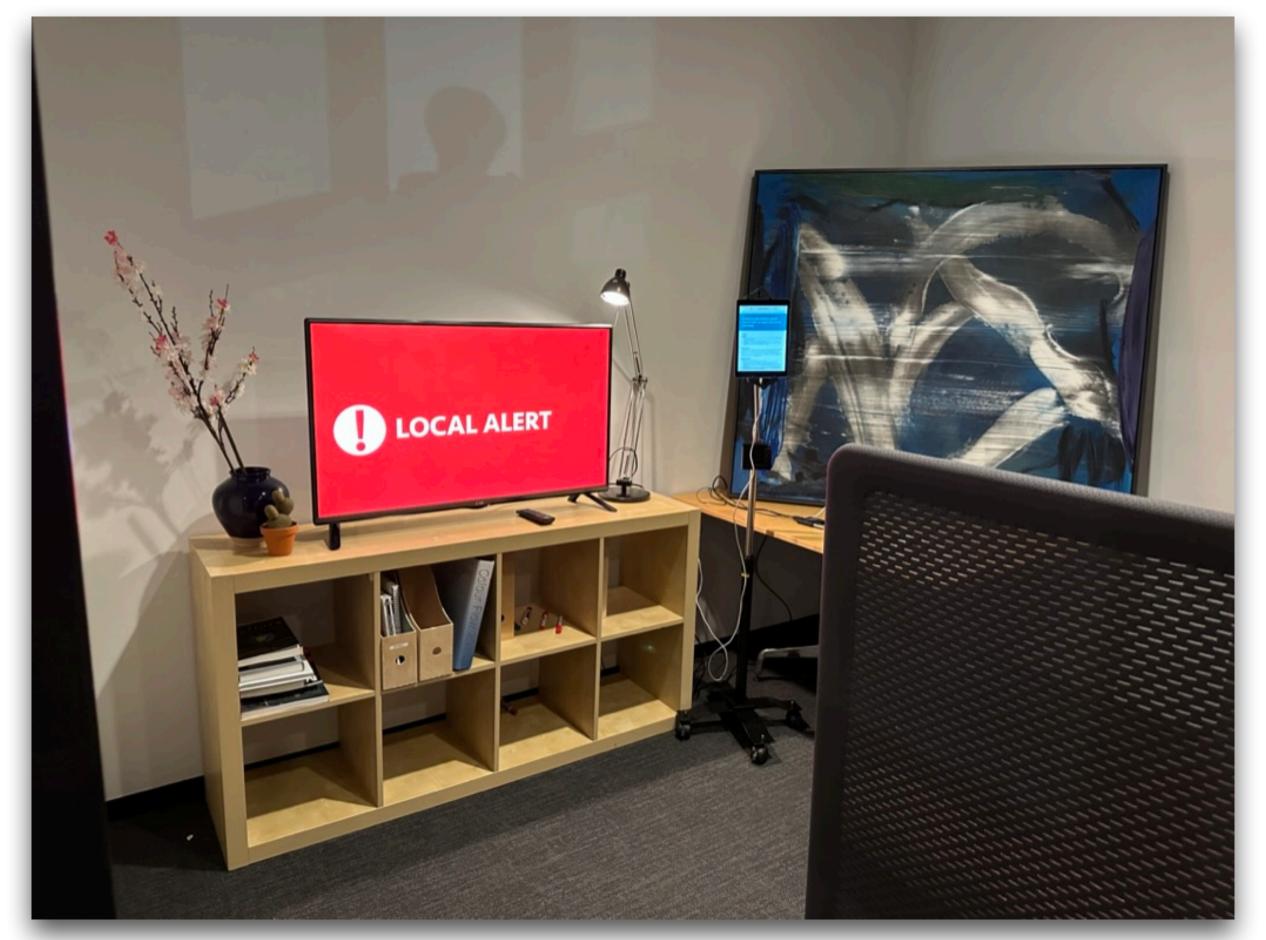
A VR testing centre - echoing current driver's licensing centres

This future is characterised by high direct intervention from the government to build mental resilience in a world more pro-active in exploring the impacts and effects of new technologies. In this future, the ways disinformation can spread in new technologies and disinformation is researched and understood before mass adoption is allowed to occur. As a result, an individual's mental resilience to these new channels must be verified by the government before they are allowed to use this new technology.

Artefact Category	Artefact
People	The participant is a candidate for a license to use new technology
Objects	Banner that thanks the candidate for their vigilance – signalling the community benefit of each individual taking responsibility for build resilience to disinformation
Environment	A government testing facility, like a driver's license testing centre
Messages	 Key messages include: New technology is useful, but can be risky Bad actors can use new technologies to spread disinformation Being able to identify disinformation is critical to community safety Testing and licensing helps protect individuals and
Services	Testing for a license to use new technologies
Video	https://www.youtube.com/watch? v=OjaEPUNTCPQ&list=PLcX2OehQVMBkfko6QnCHeSRl yPGBUWNZF

Digital community

Low government intervention, proactive engagement with new technology



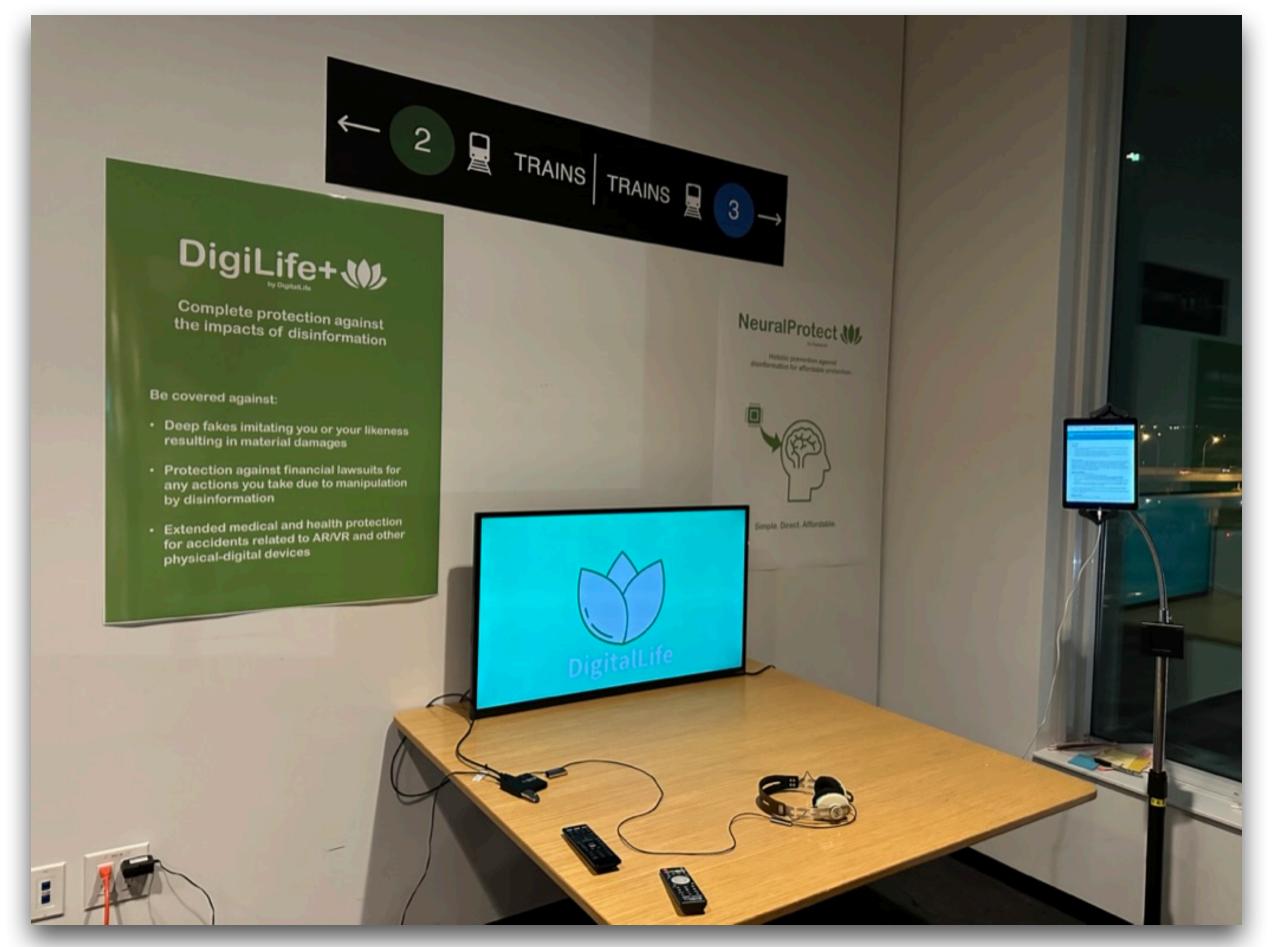
A living room environment with public disinformation alerts playing like current cable news channels

This future is characterised by educational and community-led approaches (e.g. NFP, CSO) to teaching and guiding communities through new and existing tech and general information literacy. While the government is concerned, they are less heavy-handed and prefer to fund research and interventions that are created and driven by communities and public institutions. Therefore, mental resilience in existing and new technologies is primarily done through education and advice, in public schools, community centres and across media outlets. Individuals and families are advised about disinformation events, but building mental resilience to these events is left up to individual proactivity.

Artefact Category	Artefact
People	The participant is an individual watching community news in their own living room
Objects	A television/streaming news feed of disinformation events globally and locally
Environment	A living room environment, in the comfort of the participants home
Messages	 Key messages include: Disinformation alerts globally and how to manage daily tasks against these events Local disinformation events and how to mitigate risks
Services	A news feed of alerts to warn individuals against disinformation
Video	https://www.youtube.com/watch? v=B11gpMxVOQk&list=PLcX2OehQVMBkfko6QnCHeSRly PGBUWNZF&index=2

Pay to protect Low government intervention, low engagement with new

Low government intervention, low engagement with new technology



Ads for a digital/disinformation insurance company playing on public transit (subway station)

This future is characterised by low government regulation and less pro-activity in trying to understand how disinformation is spread in new technologies. As a result, communities and individuals are trying to catch-up on how disinformation is spread in new technologies and protect themselves against mental resilience failures based on their personal interest and ability to pay. A key example in this future is the development of disinformation insurance – where individuals purchase insurance policies to protect themselves against the impacts of disinformation, but are offered discounts on their premiums if they elect to take educational classes or install neurological interventions so they themselves are more resistant to disinformation.

Artefact Category	Artefact
People	Disinformation insurance ads and advertisements for neural implants that prevent disinformation impacts
Objects	A subway/train station to signal the mass marketing of these interventions
Environment	 Key messages include: Disinformation is rampant and it is important to protect and insure individuals from its impacts Proactive intervention (e.g. neural implants) is one method individuals can take to be safer and reduce the cost of protection
Messages	Disinformation insurance against deep-fakes or bad actions taken as a result of disinformation Neural implant to make individuals more resilient to disinformation
Services	Disinformation insurance ads and advertisements for neural implants that prevent disinformation impacts
Video	https://www.youtube.com/watch? v=6LhqWBBdw4c&list=PLcX2OehQVMBkfko6QnCHeSRly

Survey resultsParticipant preferences and perceptions

- Visitor responses indicated some clear preferences and aversions to different futures
- The Pay to protect future, characterised by low government involvement and a lack of proactive technology engagement, triggered the most intense emotions and was identified to be the least preferred future, driven by concerns around thought surveillance, physical implants and thought monitoring even by private companies
- However, some participants did identify that this future felt realistic primarily referencing the potential for profit in insurance as the reason for its realism
- Interestingly, a few participants also saw the future of a pro-active government Ministry of Technology as over-regulation and invasive; it was also seen as the least likely, as participants were not convinced that government could keep recent with technology and regulation.
- The community-led and proactive future Digital community also triggered emotional intensity from some participants. In particular, the volume of disinformation and alarming format triggered anxiety about how to process all these events this was also considered to be both the most likely and preferred future
- Participants noted that this future was the simplest to ladder towards as it leverages existing infrastructure and aligns with how participants perceive community responses to disinformation warning about disinformation, instead of actively trying to regulate and prevent it

Results analysis Comparison to literature and areas for further research

- Overall, the preferences and sensitivities of participants to the different futures are aligned with research and findings in the broader discourse on managing the spread and impacts of disinformation even when it comes to building resilience against disinformation individuals are concerned with the amount of control government has over individual freedom of thought and expression
- The participants' general preference for the Digital community scenario also supports a more grassroots effort and exploration of ways to build resilience against disinformation like the more educational experiments and initiatives (e.g. disinformation games) explored by Roozenbeek and van der Linden (2019)
- There are fewer direct parallels for how private industry may participate in the prevention or insurance against disinformation; while North Americans are adopting some protection tools like password protection (as high as 55% in Canada (Anaya, 2021)); the broad adoption is limited by digital literacy and price sensitivity. (Ghobakhloo et al., 2011)
- This study measured public receptivity of these interventions, not their efficacy of the methods
- Also knowing that more community-focused futures are preferred and that more community-focused and educational initiatives are demonstrating efficacy, it is important to explore perceptions around how these methods can scale
- This study mainly used audio and visual elements to communicate the impact and environments of the future, primarily due to time and resource limitations further research to explore how immersive technologies like virtual reality, augmented reality and artefacts created from 3D printing may impact results will provide clarity in the fidelity required to create effective experiential futures

ConclusionsInsights and recommendations

- From the data and analysis gathered via the experiential futures installations, public reactions to different extremes in building resilience against disinformation are similar to current tensions around regulating disinformation production and social media usage
- Primarily, both government and private industry efforts are met with concerns over privacy, surveillance and control however, participants observed the profit motivations driving private industry investments, even if they are not guaranteed to succeed
- As a result, community-driven initiatives are generally considered the most positive and acceptable, but mostly due to their perceived ease of implementation or their less invasive nature there was little comment on their perceived efficacy.
- Therefore, governments may not have an easier path to regulation through investments in building resilience, but they may be able to leverage their own institutions (e.g. education system) rather than convincing private corporations (e.g. social media companies)
- If governments pursue partnerships with community groups to build resilience against disinformation, a balance between efficacy and perceived control may become a continuous balancing act initiatives that are more heavy-handed or direct to aim for efficacy may run into the same resistance and fear of control and surveillance.
- This project also explored the use of experiential futures in foresight work and gauge responses to different future scenarios the experiential futures primarily used audio and visual elements in the form of posters and videos on televisions and screens
- As technological advances help make new immersive and creative experiences more affordable and scalable (e.g. VR, AR, 3D printing), there is opportunity to understand whether investing time and resources in making experiential futures more immersive will generate different or more insightful responses and feedback from participants

References

- 1. Anaya, T. (2021, June 30). User authentication: 55% of Canadians reuse passwords across accounts. GetApp. https://www.getapp.ca/blog/2062/user-authentication-reuse-passwords
- 2. Bleecker, J. (2022). Design fiction: A short essay on design, science, fact, and fiction. *Machine Learning and the City: Applications in Architecture and Urban Design*, 561–578.
- 3. Casey, P., Baggili, I. M., & Yarramreddy, A. (2019). Immersive Virtual Reality Attacks and the Human Joystick. *IEEE Transactions on Dependable and Secure Computing*, 18, 550–562.
- 4. Crawford, K. (2017, September 20). *Design Thinking Toolkit, Activity 3—POEMS*. Atomic Spin. https://spin.atomicobject.com/2017/09/20/poems-template-user-observation/
- 5. Department of Defense. (2020, January 21). About. https://www.dinfos.dma.mil/About/
- 6. El-Masri, A., Riedl, M. J., & Woolley, S. (2022). Audio misinformation on WhatsApp: A case study from Lebanon. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-102
- 7. Fallis, D. (2014). The Varieties of Disinformation. In L. Floridi & P. Illari (Eds.), *The Philosophy of Information Quality* (pp. 135–161). Springer International Publishing. https://doi.org/10.1007/978-3-319-07121-3_8
- 8. Ferreira, R. G. (2008). The CIA and Jacobo Arbenz: History of a disinformation campaign. *Journal of Third World Studies*, 25(2), 59–81.
- 9. Ghobakhloo, M., Arias-Aranda, D., & Benitez-Amado, J. (2011). Adoption of e-commerce applications in SMEs. *Industrial Management & Data Systems*.
- 10. Greenspan, R. L., & Loftus, E. F. (2021). Pandemics and infodemics: Research on the effects of misinformation on memory. *Human Behavior and Emerging Technologies*, 3(1), 8–12. https://doi.org/10.1002/hbe2.228
- 11. Guzman, L. H. (2022, January 18). Chile: Pioneering the protection of neurorights [The UNESCO Courier]. Tan, D., & Sijie, J. (2020). Fake News, Free Speech and Finding Constitutional Congruence. Singapore Academy of Law Journal, 32.

- 12. Kling, J., Toepfl, F., Thurman, N., & Fletcher, R. (2022). Mapping the website and mobile app audiences of Russia's foreign communication outlets, RT and Sputnik, across 21 countries. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-110
- 13. Lee, J. H., Santero, N., Bhattacharya, A., May, E., & Spiro. (2022, September 28). Community-based strategies for combating misinformation: Learning from a popular culture fandom | HKS Misinformation Review. https://misinforeview.hks.harvard.edu/article/community-based-strategies-for-combating-misinformation-learning-from-a-popular-culture-fandom/
- 14. Liv, N., & Greenbaum, D. (2020). Deep Fakes and Memory Malleability: False Memories in the Service of Fake News. *AJOB Neuroscience*, 11(2), 96–104. https://doi.org/10.1080/21507740.2020.1740351
- 15. Lorenz, T. (2022, May 18). Disinformation Governance Board 'paused' after just 3 weeks—The Washington Post. https://www.washingtonpost.com/technology/2022/05/18/disinformation-board-dhs-nina-jankowicz/
- 16. Meissner, P., & Keding, C. (2021, October 12). *The Human Factor in AI-Based Decision-Making*. https://sloanreview.mit.edu/article/the-human-factor-in-ai-based-decision-making/
- 17. Pickard, V., Zimdars, M., & McLeod, K. (2020). Confronting the misinformation society: Facebook's "fake news" is a symptom of unaccountable monopoly power. Fake News: Understanding Media and Misinformation in the Digital Age, 123.
- 18. Roozenbeek, J., & van der Linden, S. (2019). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research*, 22(5), 570–580. https://doi.org/10.1080/13669877.2018.1443491
- 19. Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, 8(34), eabo6254. https://doi.org/10.1126/sciadv.abo6254
- 20.Southwell, B. G., Thorson, E. A., & Sheble, L. (2018). *Misinformation and Mass Audiences*. University of Texas Press. https://books.google.ca/books?id=jUw7DwAAQBAJ
- 21. Tan, D., & Sijie, J. (2020). Fake News, Free Speech and Finding Constitutional Congruence. *Singapore Academy of Law Journal*, 32.

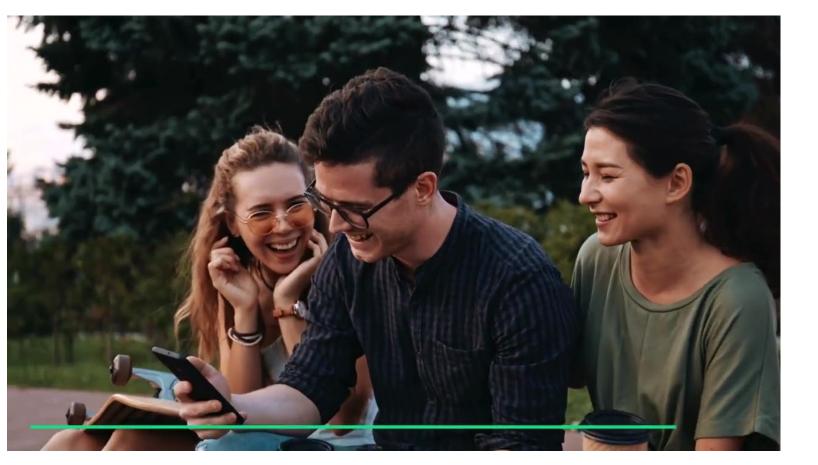
Video links Click thumbnails to view



Ministry of Technology



Digital community



Pay to protect