JUSTICE, HEALTH & DEMOCRACY
IMPACT INITIATIVE

# PLURAL PUBLICS

Shrey Jain
Divya Siddarth
E. Glen Weyl

March 2023

# Plural Publics

## AUTHORS

**Shrey Jain**
Microsoft Research
University of Toronto
v-jainshrey@microsoft.com
Our secrets will be our Turing Tests.

**Divya Siddarth**
Collective Intelligence Project
divya@cip.com

**E. Glen Weyl**
Microsoft Research
Plurality Institute
glenweyl@microsoft.com

Thanks to Stark Draper, Kim Laine, Josh Benaloh, Christian Paquin, Karen Easterbrook, Betül Durak, Matthew Prewitt, Santi Ruiz, Arthur Allshire, Puja Ohlhaver, Vitalik Buterin, Chris White, Enrico Bottazzi, Connor Spelliscy, and Zoë Hitzig for helpful discussions.

Statements and views expressed in this paper are solely those of the authors and do not imply endorsement by Harvard University or the Justice, Health & Democracy Impact Initiative.

**JUSTICE, HEALTH & DEMOCRACY**
IMPACT INITIATIVE

The Justice, Health, and Democracy Imapct Initiative is a multi-institutional collaboration between New America, the Brown University School of Public Health, and the Edmond & Lily Safra Center for Ethics at Harvard University.

> “To ensure continued functionality of digital information systems, as well as to correct their many failings, we need to develop tools that allow humans to communicate with unprecedented levels of digital context at scale.”

## 1. Context: The Foundation of Communication

The internet is often touted, not without justification, as the most revolutionary communication medium in human history. Yet while information packets now traverse any two points on the globe almost instantaneously, there is an important sense in which the internet has, paradoxically, reduced our capacity to communicate: it has imperiled our ability to establish and preserve the clarity of context on which human communication depends.

“Context” is used in a variety of ways, so we will aim to be precise about what we use it to mean: background data that is approximately “common knowledge” (known, known to be known, known to be known to be known, etc.). Communication is impossible without some such context. Languages themselves are the most basic example of context, providing a common set of symbols and grammar that allow otherwise unintelligible noises and symbols to convey meaning in a way that is understood by others who share the same language. Shared personal histories, secrets, cultural milieus, technical disciplines, national memories, oral traditions, and various other contexts empower us to convey far more meaning per packet of information than would be possible if we saw the world anew each day. Perception and understanding differs for every listener and depends on the context they share with the speaker. Context is central to defining the status of a relationship among the parties to the communication and the basis for the exchange they are engaged in.

## 2. The Threats to Context and Contextual Integrity

As leading technology scholars like danah boyd and Helen Nissenbaum have highlighted, it is precisely because the internet has made it possible to share so much information in such novel combinations that it has undermined our ability to establish, track, and protect context in two ways [1],[2]. First, as we talk to a wider range of people with less shared experience, it is increasingly hard to know what context we may assume in communicating. Without such clarity, communication is likely to be hampered or entirely ineffective: we may speak and others may hear, but they will typically lack the context to understand our meaning and may even misinterpret us, making us unwilling to speak clearly.

Second, the speed and ease with which information moves has made it hard to ensure that statements remain within the context in which they were made, often leading to “context collapse,” causing different audiences to misunderstand or misjudge the information based on their disparate contexts [1]. Avoiding such collapses requires maintaining “contextual integrity," as Nissenbaum labels it: ensuring that speakers can have a clear understanding of the contexts in which statements they make will be interpreted and ensuring that information does not leave these contexts. Absent such integrity, we may become unwilling to harness context in our communications for fear that what we say will be "taken out of context." Accelerating communicative capacity may therefore erode our ability to communicate effectively.

Of particular concern is the emergence of generative "artificial intelligence" (AI) systems, which are increasingly capable of producing highly realistic texts, images, and video [3]. These systems, also known as generative foundation models (GFMs), will proliferate persuasive machine content on our information networks at lower cost and typically higher profit than humans can more directly produce such data. This threatens the integrity of nearly all social systems, including the very internet corpus on which these foundation models are trained.

GFMs enable information to be easily transferred across tasks and domains, but this can cause context collapse.[1] While these considerations may seem abstract, recent exploits of low-context communication have made their consequences quite vivid. Scammers are harnessing foundation models to reuse voice clips posted online to mimic the voices of people victims know and love for confidence games [4]. Such extremely decontextualized reuse shows the potentially extreme risks of context collapse.

## 3. Mission: Plural Publics

The problem of context collapse is closely related to discussions around "privacy." In fact, Nissenbaum persuasively argues that much, if not most, of what we mean when we discuss "privacy" is really a concern for contextual integrity. However, nearly all formal work on privacy technology (much of it connected to the cryptography) and the great majority of privacy policy focuses on much more individualistic notions of privacy versus publicity, with facts being (stochastically) either private or public. We therefore find it useful to call out this problem of context collapse separately and provide some design criteria to help sharpen design and policy work in this direction.

Instead of focusing on "private" v. "public," we seek to protect and enable the emergence of a rich diversity of "publics." For us, a "public" is a digital communication channel in which individuals possess a confident understanding of the context surrounding their messages. In these publics, individuals have the confidence that their messages will be interpreted through the lens of that context. This notion of "public" owes much to political philosopher John Dewey's (1927) concept of "emergent publics" [5], polities that form to coordinate common action and democratic control over issues of shared concern, as well as to sociologist Georg Simmel's [6] emphasis on the role of shared secrets as a foundation for such common action. Internet pioneers like J. C. R. Licklider were partly inspired by these ideas (the "web" terminology may well have arisen from a mistranslation of a famous essay of Simmel's), and Licklider anticipated as early as 1968 [7] the challenges to effective communication the internet could create due to what Martin Gurri has called an "information tsunami" [8]. In tribute to that tradition, we label our goal set as promoting **plural publics**[2], following Dewey's use of "public" in the plural and emphasizing the need to create a diversity of such publics.

In a public, context must be (approximate) common knowledge among its participants. It is necessary for participants in a public to have confidence in the context they share, and that all participants have confidence in others' confidence and so on [9]. Today, communication continues, however hobbled, on digital

[1] The trust-eroding potential of foundation models is not all negative, however. Eroding trust also means that shared information will be distrusted if not verified, strengthening technical tools, discussed below, which use deniability to enhance contextual integrity.

[2] In traditional Mandarin, "digital" and "plural" have the same characters, so plural publics can also be understood as "digital publics."

platforms despite context collapse. However, the disruption caused by GFMs could significantly erode current communication platforms unless we adopt new forms of context-preserving technologies. To ensure continued functionality of digital information systems, as well as to correct their many failings, we need to develop tools that allow humans to communicate with unprecedented levels of digital context at scale.

## 4. Tools to Help Guide Us

Moving towards a contextually grounded information scheme will require new tools, governance, and sociotechnical advances. There are various tools available that can be utilized to build plural publics:

1. **Group-Chat Messaging** facilitates contextual communication within a public, as they provide a channel in which participants possess a shared understanding of the contextual landscape surrounding their messages [10]. However, group chats today make it relatively straightforward for members to share messages from the chat onward, out of context, either directly or through screenshots. Group chats also typically do a relatively poor job of ensuring common knowledge: it is hard to know how many others are paying attention to the messages posted and thus hard to know if they can be thought of as a shared context or just screams into a void.

2. **Deniable and Disappearing Messages** allow for only the intended recipient to be convinced by the authenticity of information.[3][11] Disappearing messages allow for context to be time-bound by preventing participants in the initial conversation from proving their prior claims. Even if information was attempted to be shared de-contextually, recipients of such information would not be able to prove its authenticity. These techniques are likely to be increasingly useful as persuasive machine-generated content of questionable validity proliferates and thus people come to rely more on verification to provide credibility.

3. **Distributed Ledger Technologies (DLTs)** enable context to be approximate common knowledge among all machines participating in a consensus protocol [12]. However, whether what is approximate common knowledge among machines is approximate common knowledge among the human operators of those machines is a sociotechnical rather than technical question, and requires much more study. Furthermore, much additional technical work would be needed to make DLTs serve the goal of plural publics, as they would have to operate much faster, at lower cost, and with much stronger privacy protections than do most current, publicly oriented DLTs.

4. **Identity Certificates** allow for individuals to cryptographically prove the possession of context to send or receive messages. Identity

[3] “Designated verifier signatures” are an example of a deniable signature where only the designated verifieris convinced of the authenticity of the sender’s digital signature [11].

certificates[4] are also subject to forgery or the theft/corruption of its keys. However, the intersection of many verified attestations could greatly improve both sender's and receiver's confidence about the contextual landscape they are communicating in. Advances in methods for community/social key recovery, non-transferability, and account abstraction may help reduce the threats of account corruption[13], [14], [15].

5. **End-to-End Encryption** prevents information from being shared across publics, but becomes more challenging to implement correctly with the increased use of GFMs. For example, proving that a participant in an end-to-end encrypted meeting is truly one of the intended participants requires that the public is able to maintain and broadcast a coherent view of who is supposed to be a part of it and that its members are able to reliably authenticate themselves to each other.

A combination of these technical tools would be useful for individuals who want to prevent GFMs from training on their data. Additionally, given that GFMs continue to pose a threat to individuals' financial security and reputation, plural publics communication will likely emerge as an effective solution. Incorporating deniable messages and identity certificates that privately authenticate a pre-established context, such as secrets, on a ledger between two parties can establish approximate common knowledge that the individual with whom one is communicating possesses the presented certificates. There are ongoing experiments with new forms of communication utilizing such tools.[5] This can help mitigate leading potential harms from GFMs: hyper-personalized disinformation campaigns, denial-of-information attacks, child sexual abuse material (CSAM), and phishing scams, among others [16].

While we believe that improved supporting technology can enhance the chances of achieving such informational pluralism, we do not suppose that any of the tools we discuss are necessary or sufficient conditions for achieving those ends. Furthermore, we are acutely aware of how these affordances may be abused by, for example, criminal organizations that need to form and protect secret context. The tools may even undermine their direct goals if their interface with sociotechnical practices does not match the technical affordances of the systems, a problem that will be especially acute for less technically sophisticated users.

## 5. Agenda on How to Advance

Privacy is widely seen as a fundamental human right grounding the survival of democracy. For example, at the 2021 Summit of Democracies, the then-United States Science Advisor Eric Lander highlighted "privacy-enhancing technologies" as a primary category of "Technology for Democracy" the administration hoped to support [17].

[4] The most widespread use of such a certificate system today is with X.509 Certificates. Other emergent credentialing solutions include BBS+ verified credentials (VCs), U-Prove, CL-signatures with anonymous credentials, selective-disclosure-JSON Web Tokens, and zero-knowledge soulbound tokens.

[5] Some emergent protocols experimenting with a combination of the tools above include Spritely, Mastodon, Nostr, AT Protocol, Farcaster, Lens, Status, and DSNP.

We believe that this emphasis, while clearly in the vicinity of something profound, is a bit imprecise. Democratic and pluralist societies do not depend on individual privacy primarily; in fact, as as Hannah Arendt's *The Origins of Totalitarianism* (1951) and other observers like Alexis de Tocqueville's *Democracy in America* (1835) before her have noted, authoritarian regimes are usually comfortable with or even promote *individual* privacy [18], [19]. Rather, what they fear is collective coordination outside the view and control of the state. As de Tocqueville said, "a despot easily forgives his subjects for not loving him, provided they do not love each other" [19]. As such, we believe that a central goal for technology that would enable it to support pluralistic societies would be to support what we call plural publics.

While the concept of plural publics provides a high-level design, many open questions remain. What is the interface between what a machine knows and what a human knows? To what extent and in what cases is deniability sufficient for preserving context? When, and in what amount, is common knowledge necessary or desired as humans begin to communicate alongside artificial agents? Can we measure the necessary context needed to enter a public? How do we facilitate communication and interoperation across many both nested and intersecting publics without undermining contextual integrity? What is the most efficient way to scale digital certificates on the internet today? What identification methods do we need so that members of a public can communicate securely among each other? What lessons can be drawn from the digital rights management ecosystem to the development of publics today? These are complex, sociotechnical questions, the kind that we would expect to take decades to work out. Unfortunately, we may not have decades: the rise of GFMs threaten, absent good answers to these questions, to undermine the very foundations of pluralism, trust, and identity. Work in this, ranging from theory quickly to this application, could hardly be more urgently needed.

# Endnotes

1 d. boyd, *It's Complicated: The Social Lives of Networked Teens*. New Haven, CT:Yale University Press, 2014.

2 H. Nissenbaum, "Privacy as Contextual Integrity," *Washington Law Review*, vol. 79, no. 1, pp. 119–158, 2004.

3 R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, *et al.*, "On the Opportunities and Risks of Foundation Models," *arXiv preprint arXiv:2108.07258*, 2021.

4 P. Verma, "They Thought Loved Ones Were Calling for help. It Was an AI Scam," *The Washington Post*, March 5, 2023.

5 J. Dewy, *The Public and Its Problems*, vol. 11 of *A Swallow paperbook*. H. Holt, 1927.

6 G. Simmel, "The Sociology of Secrecy and of Secret Societies," *American Journal of Sociology*, vol. 11, no. 4, pp. 441–498, 1906.

7 J. C. R. Licklider and R. W. Taylor, "The Computer as a Communication Device," *Science and Technology*, vol. 76, no. 2, pp. 21–38, 1968.

8 M. Gurri, *The Revolt of the Public and the Crisis of Authority in the New Millennium*. San Francisco: Stripe Press, 2014.

9 S. Morris and H. S. Shin, "Approximate Common Knowledge and Co-ordination: Recent Lessons from Game Theory," *Journal of Logic, Language and Information*, vol. 6, no. 2, pp. 171–190, 1997.

10 J. Lanier, "How to Fix Twitter—And All of Social Media," *The Atlantic*, May 26, 2022.

11 M. Jakobsson, K. Sako, and R. Impagliazzo, "Designated Verifier Proofs and Their Applications," in *Advances in Cryptology—EUROCRYPT '96. International Conference on the Theory and Applications of Cryptographic Techniques*, Proceedings, pp. 143–154, Berlin: Springer, 1996.

12 J. Y. Halpern and R. Pass, "A Knowledge-Based Analysis of the Blockchain Protocol," *arXiv preprint arXiv:1707.08751*, July 27, 2017.

13 V. Buterin, Y. Weiss, K. Gazso, et al., "ERC-4337 Account Abstraction Using Alt Mempool," *Ethereum Improvement Proposals*, September 29, 2021.

14 E. G. Weyl, P. Ohlhaver, and V. Buterin, "Decentralized Society: Finding Web3's Soul," *SSRN*, posted May 11, 2022.

15 S. Jain, L. Erichsen, and G. Weyl, "A Plural Decentralized Identity Frontier: Abstraction v. Composability Tradeoffs in web3," *arXiv preprint arXiv:2208.11443*, August 24, 2022.

16 M. Brundage, S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, P. Scharre, T. Zeitzoff, B. Filar, et al., "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," *arXiv preprint arXiv:1802.07228*, February 20, 2018.

17 White House, "Summit for Democracy Summary of Proceedings," December 23, 2021.

18 H. Arendt, *The Origins of Totalitarianism*. 1951; Reprint, New York: Houghton Mifflin Harcourt, 1973.

19 A. De Tocqueville, *Democracy in America*, vol. 2. New York: D. Appleton, 1899.