Civic OS: Toward a Human-Verified Digital Public Sphere

Shooresh Sufiye
Code Iland AB
August 2026

The Internet and social media have become important arenas for observing, expressing, and ultimately shaping public opinion. People increasingly encounter political ideas, social attitudes, and perceptions of what others believe through online networks. Research on the “spiral of silence” and related feedback effects shows that people’s willingness to express an opinion can depend on what they perceive to be the prevailing opinion around them (Sohn & Geidner, 2016; Chen, 2018). Social media therefore does more than simply record public opinion: it can create a feedback loop in which visible online reactions influence what people subsequently think, discuss, and express.

This creates a serious problem when the participants producing those signals are not necessarily genuine individuals. Bots, coordinated fake accounts, troll networks, and organized “cyber armies” can artificially amplify messages, hashtags, personalities, or political positions. Research on computational propaganda has documented how automated and coordinated accounts have been used to manipulate political communication and manufacture the appearance of popularity or consensus (Woolley & Howard, 2016; Bessi & Ferrara, 2016; Howard, Woolley, & Calo, 2018). More recent studies likewise show that bots can artificially inflate politicians’ apparent popularity and contribute to the dissemination and amplification of misleading information (Silva & Proksch, 2021; Tomassi, Falegnami, & Romano, 2024).

The consequence is fundamental: **the apparent voice of the Internet is not necessarily the voice of its human population**. Thousands of accounts supporting an idea may represent thousands of independent people—or a much smaller number of actors operating automated or coordinated identities. When journalists, governments, companies, researchers, and ordinary citizens interpret likes, comments, trends, and reposts as evidence of public sentiment, manipulated activity can therefore create a distorted picture of society.

As artificial intelligence makes synthetic identities and automated communication increasingly sophisticated, distinguishing genuine human participation from artificial participation will become even more important. Emerging work on “personhood credentials” argues that digital systems may need privacy-preserving mechanisms capable of establishing that an online participant corresponds to a real and unique human without necessarily revealing that person’s identity (Adler et al., 2024). Such systems could make large-scale Sybil attacks—where one actor creates many apparently independent identities—significantly more difficult.

This is the space in which **Civic OS** seeks to operate. Its ambition is to create a digital public sphere in which participation can be attributed to verified, unique human beings rather than an unknown mixture of people, bots, and fabricated identities. It is designed with strong protections

for privacy, freedom of expression, accessibility, and political neutrality; such an infrastructure could provide a more credible signal of what real people actually think.

Human verification alone, however, does not automatically make a platform statistically representative of society; demographic and participation biases would still need to be measured and addressed. Nevertheless, establishing that **one participant represents one genuine human** would solve an increasingly important part of the problem. Civic OS can therefore be understood not simply as another social network, but as an attempt to build a trusted **human layer for digital public opinion** - an infrastructure through which society can listen to itself with greater confidence.

ss@codeiland.com

Selected academic references

Sohn, D., & Geidner, N. (2016). *Collective dynamics of the spiral of silence: The role of ego-network size*. International Journal of Public Opinion Research, 28(1). Oxford Academic

Miyata, K., Yamamoto, H., & colleagues. (2015). *What affects the spiral of silence and the hard core on Twitter? An analysis of the nuclear power issue in Japan*. American Behavioral Scientist. SAGE

Chen, H.-T. (2018). *Spiral of silence on social media and the moderating role of disagreement and publicness in the network: Analyzing expressive and withdrawal behaviors*. New Media & Society. SAGE

Woolley, S. C., & Howard, P. N. (2016). *Political communication, computational propaganda, and autonomous agents—Introduction*. International Journal of Communication. IJoC

Bessi, A., & Ferrara, E. (2016). *Social bots distort the 2016 U.S. Presidential election online discussion*. First Monday. Article record

Howard, P. N., Woolley, S., & Calo, R. (2018). *Algorithms, bots, and political communication in the U.S. 2016 election: The challenge of automated political communication for election law and administration*. Journal of Information Technology & Politics. Taylor & Francis

Ferrara, E., Chang, H., Chen, E., Muric, G., & Patel, J. (2020). *Characterizing social media manipulation in the 2020 U.S. presidential election*. First Monday. First Monday

Silva, B. C., & Proksch, S.-O. (2021). *Fake it 'til you make it: A natural experiment to identify European politicians' benefit from Twitter bots*. American Political Science Review. Cambridge University Press

Tomassi, A., Falegnami, A., & Romano, E. (2024). *Mapping automatic social media information disorder: The role of bots and AI in spreading misleading information in society*. PLOS ONE. PLOS ONE

Adler, S., Hitzig, Z., Jain, S., Brewer, C., Chang, W., et al. (2024). *Personhood credentials: Artificial intelligence and the value of privacy-preserving tools to distinguish who is real online*. arXiv. arXiv