Lynnette Ng -- Research Statement

Cyber Social Agents: The Hidden Architects of Social Media

Across social media platforms, automated accounts operate as conduits of information. These Cyber Social Agents (CSAs) architect the information environment, constructing the space that shapes how millions of people form beliefs, choose sides and act. They seed narratives, coordinate amplification, exploit psychological vulnerabilities that guide belief formation. Yet for all their influence power, they remain largely invisible, operating beneath the threshold of everyday human perception.

My research program develops a computational science of these online human-agent social systems. I examine CSAs not as sources of manipulation to be detected and removed, but as socially embedded actors whose behavior and influence can only be understood within the relational fabric of social media. In my research, social media is treated not as a platform for human communication, but as a complex human-agent social system, in which automated and human participants interact, co-evolve and shape each other's behavior. In studying such communications, I take a deeply interdisciplinary approach, integrating graph-theoretic modeling, natural language processing, social psychology and agent-based simulation, grounded in datasets of over 200 million social media users and 5 billion posts.

I have established foundational contributions across three pillars:

1. Who Are the Architects? Detection and Differentiation of CSAs

A central challenge in studying CSAs is identifying them; that is, distinguishing automated accounts from humans, and differentiating nuanced CSA archetypes from one another. I develop several machine-learning based detection models to differentiate between bot and human at operational scale (Ng & Carley, *AAAI ICWSM*, 2023). These models evolve over time as agentic technologies evolve, providing updated analysis of the ecosystem (Ng & Carley, *Social Network & Analysis*, 2024). Globally, about 20% of the social media users affiliated to any country are automated users, suggesting the power and extent of information maneuvers that can take place online (Ng & Carley, *Scientific Reports,* 2025).

Beyond the binary agent/human distinction, I conceptualize CSAs as a heterogeneous population that employ a spectrum of operational tactics and rhetorical strategies. Drawing on large-scale qualitative annotation of behavioral traces across health pandemics, I produce a nuanced, behaviorally grounded typology and operationalize these distinctions using machine learning models, graph-theoretic models, and statistical models. This typology includes, among others, News Agents that posts or aggregates news stories, which are identified through the similarity of their posts to journalistic headlines; Bridging Agents that transmit information across disparate network communities, which are identified by users whose deletion increases the number of connected components in a graph; Social Influence Agents that strategically deploy information maneuvers, identifiable through narrative and network shifts (Ng et. al., *ACM WebSci*, 2025; Ng & Carley, *Online Trust and Safety*, 2026).

Architecturally, CSAs are exhibit distinctions from human users that reveal that automated origins. Compared to human users, CSAs post at approximately twice the rate and rely on shorter, semantically simpler linguistic constructions, consistent with the persuasive technique commonly known as "flooding the zone", which aims to drown out competing narratives. Structurally, CSAs tend to exhibit star-shaped interaction graphs to facilitate rapid broadcast for message amplification, compared to the hierarchical network patterns characteristic of human conversational threads (Ng & Carley, *Nature Scientific Reports,* 2025). This work is the first analysis and characterization of CSAs at such a scale. It connects the roles agents can take in the cyber social space into measurable characteristics, allowing persuasive influence to be studied as an emergent property of CSA behavior.

2. What are CSAs Architecting? Modeling Influence, Coordination and Belief Dynamics

The next pillar of my work investigates how coordinated agents shape narrative visibility, engagement, polarization, persuasion and belief change. A central contribution of my research is the idea that influence is not simply a property of individual messages, but is the emergent property of interactions between narratives, network structures and agent interactions. To study this, I develop network graph-based frameworks that quantify digital influence at multiple levels.

Modeling Belief Change. One of my obsession is on how coordinated CSA activity induces stance revisions at the individual level. Using a modified Friedkin-Johnsen model of social influence, I model stance flipping in online environments by integrating endogenous linguistic signals (i.e., historical stance expression, conviction) with exogenous network factors (i.e., neighbor stance distributions, reciprocity, coordination) to predict when individuals revise their express beliefs. This approach captures persuasion as a dynamic process that shifts opinion equilibria with networks, rather than isolated message effects. Empirically, belief change varies substantially across contexts: highly personal issues like vaccination exhibit strong resistance to stance revision, with only 1% of users changing their beliefs; but in geopolitical conflicts where users are farther removed (i.e. Russia-Ukraine conflict), up to 50% of them change their beliefs (Ng & Carley, *IEEE Transactions of Network Science and Engineering*, 2022). This work advances computational models of persuasion by treating beliefs as socially embedded and continually evolving within the digital ecosystem.

Coordination and Narrative Propagation. My obsession with belief change leads to the obsession with coordinated users, where user actions synchronize with each other, pulling other adjacent users into the same song. I develop coordination signature profiles that integrate semantic similarity, temporal synchronization, referral coordination and interaction structure (Ng & Carley, *WebSci*, 2022; Ng & Carley, *Applied Network Science*, 2023). In these, coordinated campaigns are not clusters of accounts, but are emergent collective behavior of information propagation. My research demonstrates that coordinated agents can systematically manipulate local information exposure, narrative salience, and perceived social consensus (Ilevia, Ng & Carley, *Nature Humanities & Social Science*, 2026). Coordination is therefore a fundamental mechanism of the influence architecture.

To quantitatively measure influence at scale, I develop graph-based influence metrics that go beyond raw engagement counts. The Appeal metric captures network-weighted popularity by weighting message engagement with the network centrality of its origin; the Scope metric captures structurally weighted narrative reach by integrating diffusion dynamics with structural connectivity. These metrics account for where influence originates and how far it propagates through the network, operationalizing influence as a joint property of content and network position (Ng, Zhou & Carley, *AMCIS*, 2025, under review *PLOS One*). Together, these contributions establish a unified framework for studying influence as an emergent property of coordinated socio-technical systems.

Cognitive Bias Triggers and Engagement. Finally, automated agents can only be influential if they manage to draw engagement from human users. I examine the psychological mechanisms that underlie CSA-driven engagement. Using computationally detected bias triggers (i.e., affect bias, availability bias, authority bias, cognitive dissonance), I quantify how narratives exploit psychological heuristics to drive interaction. Through a series of multiple regressions, I show that CSAs deploy these triggers more systematically than humans, and specific combinations of triggers are associated with distinct engagement outcomes (Ng, Zhou & Carley, under review, *Nature Scientific Reports)*. This work bridges observable behavioral signals with underlying cognitive mechanisms.

3. How can we Design Healthier Information Ecosystems? Simulating interventions in Human-AI Social Systems

The third pillar of my research asks a broader question: Can Cyber Social Agents themselves be used to improve the information ecosystems? The same coordination mechanisms and linguistic strategies that adversarial CSAs exploit for manipulation can, under the right conditions, be redirected towards constructive counter-architecture. I use agent-based simulation techniques to investigate this possibility under experimental setups that real-world platforms cannot provide.

My framework, AuraSight, combines graph-based social simulation, agent-based modeling and large language model narrative generation to produce a simulated network. Its parameters are empirically grounded interaction rules to produce synthetic but realistic social media environments populated by humans and CSAs (Ng, Kang & Carley, *CMU Technical Reports*, 2025). Simulation parameters are calibrated such that a 60:30:10 network preferential attachment: community leader: random interaction ratio produces graph centrality metrics closest to empirical distributions (Ng & Carley, *SBP-BRiMS,* 2025). This infrastructure allows me to experimentally vary network and linguistic conditions and analyze agent responses, which cannot be ethically or practically tested in real online populations.

Using BotSim, I investigate how different combinations of harmful messaging agents, good messaging agents and information correction agents shape the trajectory of network discourse. My results show that harmful narratives tend to dominate networks in the absence of interventions, but good messaging and information correction agents that are deployed in strategic network positions can significantly delay and partially mitigate harmful narrative cascades (Ng & Carley, *Journal of Artificial Societies and Social Simulation,* 2026). With this

line of work, I propose that rather than use defensive censorship models, we should use agentic counter-architectures, and use pro-social CSAs to reshape the information environment from within, gradually nudging collective discourse towards more balanced equilibria.

Broader Impacts: My research has been published in top venues like Nature Scientific Reports, Information Processing & Management and AAAI International Conference for Web and Social Media, to name a few. My work has won several awards, including the Best Paper Award for the Information Processing & Management and Online Social Networks and Media journals. My work has also been featured on mainstream channels like the Channel News Asia, the Pittsburgh Post-Gazette, New Scientist magazines, the NewsPress podcast, and has trended on Reddit and RedNote. My work on the description of CSAs and their activities have been contracted into a book by Cambridge Publishing Press, titled "Bots, Bias and Influence: The Hidden Architects of Social Media", forthcoming in 2026. My technical work has been contracted into a textbook by Cambridge Academic Press, titled "Generating Artificial Societies: Agent-Based Modeling with Large Language Models".

My simulation work has been presented for executive education courses, and my AuraSight system have been used as part of bi-annual training program for the US Office of Naval Research. Finally, my belief formation work has been used by the US Army, who have extended one of my datasets (Cruickshank, Soofi, Ng, *IEEE Big Data*, 2024) for the 2025 Military Operations Research Society data challenge.

Future Work: Toward Multi-Agent Social Systems

My future research aims to develop a comprehensive computational science of social systems – environments in which human and AI participants interact, co-evolve and collectively shape information landscapes at scale (Ng et al, *NeurIPS,* under review, 2026).

1. **A global account of Cyber Social Agents**

Most existing research on social media influence remains disproportionately Western-centric. I aim to build one of the first large scale global analyses of CSAs across Asia, the Global South, multilingual environments and cross-platform ecosystems. In this benchmark, I will examine how persuasive strategies, coordination dynamics and intervention effectiveness vary across cultural and political contexts. Beyond description, this work will test how the structure of online information spread predicts consequential offline behavior, such as product purchasing decisions or political alignment, across diverse societal conditions. My goal is to build a taxonomy of global differences, and a generalizable theory of which architectural mechanisms are universal and which are cultural-dependent.

2. **Generative Persuasion and Narrative Engineering**

As generative AI becomes increasingly integrated into online communication, persuasive narratives are themselves becoming computationally engineered. I aim to study how language, images, video, and network structure jointly shape persuasion in multimodal environments (i.e. Instagram, TikTok). This will model how narratives are generated, adapted and amplified in response to social feedback. Therefore, these studies include the development of archetype-

specific narrative generation systems for different CSA types, integrating linguistic features, multimodal features and network graph properties to model how persuasive content is produced and socially calibrated through interaction.

3. Multi-Agent Social Systems

Current AI systems are typically evaluated as isolated agents solving static tasks. My long-term research vision instead investigates AI systems like Cyber Social Agents embedded within dynamic social ecosystems populated by humans and other agents. This work will study co-evolution between human and AI behavior, distributional instability in belief landscapes, strategic heterogeneity across agent types, and network-constrained interaction dynamics. Ultimately, I aim to establish a computational framework for studying how collective behavior (i.e., consensus formation, polarization, narrative dominance) emerges within large-scale human-AI social systems.

4. Designing Responsible Interventions

Finally, I will use my simulation line of work to drive policy analysis, narrative intervention system designs, communication strategies and the development of socially beneficial CSAs. A central challenge of this is to build responsible counter-interventions that are not manipulation, which requires transparency safeguards and governance. I intend to work closely with governments, civil society organizations and industry partners to design such simulations for evidence-based interventions towards healthier information environments.

Conclusion: Every day, billions of social media users navigate an information environment that is also quietly shaped by agents they don't discern. Understanding the hidden architects of social media is not just a technical problem, but is a prerequisite to building information systems that serve humans. My research program aims to develop the conceptual vocabulary, empirical methods and simulation infrastructure to understand, measure and design these agents in a multi-agent social system.

References

Ng, L. H. X., Carley, K. M. (2025) A global comparison of social media bot and human characteristics. *Scientific Reports, 15(1),* 10973

Ng, L. H. X., Zhou, W., Carley, K. M. (2025) Appeal and Scope of Misinformation Spread by AI Agents and Humans. *AMCIS 2025 Proceedings. 6.*

Ng, L. H. X., Carley, K. M. (2025) Are LLM-Powered Social Media Bots Realistic? In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation* (pp. 14-23). Cham: Springer Nature Swizterland.

Ng, L. H. X., Kang, B. N. Y., Carley, K. M. (2025) AuraSight: Generating Realistic Social Media Data. *Carnegie Mellon University Technical Reports. CMU-S3D-25-109*

Ng, L. H. X., Zhou, W., Carley, K. M. Bots exploit cognitive bias triggers to shape misinformation engagement. Under Review at Scientific Reports.

Ng, L. H. X., Cruikshank, I. J., Lim, X. W. A., Carley, K. M. Bots exploit cognitive bias triggers to shape misinformation engagement. Under Review at NeurIPS.

Ng, L. H. X., & Carley, K. M. (2026). BotSim: Mitigating The Formation Of Conspiratorial Societies with Useful Bots. *Journal of Artificial Societies and Social Simulation*, *29*(1), 4.

Ng, L. H. X., & Carley, K. M. (2026). 35 The Dual Personas of Social Media Bots. In *Online Trust and Safety: Tools to Combat Online Harms, Misinformation and Malicious Content* (pp. 35-54). Boca Raton and London: CRC Press.

Ng, L. H. X., Carley, K. M. (2022) Pro or anti? A social influence model of online stance flipping. *IEEE Transactions on Network Science and Engineering, 10(1),* 3-19

Cruickshank, I. J., **Ng, L. H. X.,** & Soofi, A. (2024, December). DIVERSE: A Dataset of YouTube Video Comment Stances with a Data Programming Model. In *2024 IEEE International Conference on Big Data (BigData)* (pp. 2080-2089). IEEE.

Ng, L. H. X., & Carley, K. M. (2022, June). Online coordination: methods and comparative case studies of coordinated groups across four events in the united states. In *Proceedings of the 14th ACM Web Science Conference 2022* (pp. 12-21).

Ng, L. H. X., & Carley, K. M. (2023). A combined synchronization index for evaluating collective action social media. *Applied network science*, *8*(1), 1.