Study identifies human–AI interaction scenarios that lead to information cocoons

October 31 2023, by Ingrid Fadelli

Users are confined to information cocoons. Credit: Piao et al.

The widespread use of AI algorithms, particularly algorithms designed to recommend content and products to users based on their previous activity online, has led to the rise of new phenomena known as social media echo chambers and information cocoons. These phenomena pose limitations in the information presented to users online, prompting them to only read content that resonates with their outlook on life, views and opinions.

A recent study from an interdisciplinary team led by Prof. Yong Li at Tsinghua University investigates the dynamics underpinning the formation of so-called information cocoons, online spaces in which users only encounter voices or users that re-enforce their own views. Their paper, published in *Nature Machine Intelligence*, outlines two different scenarios that can prompt the formation of information cocoons, along with possible strategies to prevent it.

"Artificial intelligence has permeated all kinds of human activities and catapulted the presence of algorithms in every aspect of modern life," Jinghua Piao, first author of the paper, told Tech Xplore. "However, the wide adoption of AI-driven algorithms creates a new set of challenges, for example, reducing exposure to ideologically diverse news, opinions, political views and friends.

"In particular, recommendation algorithms, some of the most widely adopted AI-driven technologies, are infamous for isolating humans from diverse information and eventually trapping them in a single topic or viewpoint, known as information cocoons."

Information cocoons can have far-reaching adverse consequences, as they can exacerbate prejudice and social polarization, prevent growth, creativity and innovation, accentuate misinformation, and obstruct

efforts aimed at creating a more inclusive world.

While their existence has been widely documented, the mechanisms underlying the emergence of these online "information bubbles" remain poorly understood.

"The purpose of our recent paper was to uncover the origin of information cocoons in the complex human-AI interaction system," Piao explained. "The notion of information cocoons is adopted to describe a widely observed phenomenon that with the increase in interactions with intelligent algorithms, people are isolated from diverse information and eventually trapped in a single topic or viewpoint."

The key assumption underpinning the recent work is that the creation of information cocoons cannot be only attributed to humans or to recommendation algorithms. Instead, the team believes that they arise as the result of complex interactions and exchanges of information between multiple entities.

"Through an empirical and theoretical investigation, we reveal that information cocoons emerge from the adaptive information dynamics in the interaction feedback loop between humans and AI-driven recommendation algorithms," Piao said. "Such a feedback loop is characterized by essential ingredients: (i) similarity-based matching, (ii) positive feedback, (iii) negative feedback, and (iv) random self-exploration."

Similarity-based matching is the process through which recommendation algorithms match people with online content, products and other users that most closely resemble those they consumed or interacted with in the past. Prof. Li's team found that this tendency to make recommendations based on similarities is a key force driving online social medias and networks towards information cocoons.

"Positive feedback further amplifies this effect, resulting in a decrease in information entropy (i.e., information diversity)," Piao said. "Negative feedback and random self-exploration promote information diversity by resisting the effect of the effective force field and introducing perturbation to the system. This resistance drives the system from information cocoons to diversification."

The researchers were ultimately able to identify two key processes that play a key role in the emergence of information cocoons in complex systems involving interactions between AI recommendation algorithms and humans. These processes include an imbalance between positive and negative feedback, along with the continuous reinforcement of similarity-based matching.

"Our findings suggest two practical ways to mitigate real-world information cocoons," Piao said. "The first is the effective utilization of negative feedback, which offers a new perspective on users' preferences by identifying their dislikes. The second is the promotion of self-exploration, which diversifies the available information by empowering users to exercise greater autonomy over the algorithm."

Overall, Prof, Li's team gathered valuable insight about the mechanisms underpinning the formation of information cocoons online, which could soon inform the creation of alternative AI tools and strategies addressing these mechanisms. Their paper was the result of a one-year collaboration involving researchers with expertise in various disciplines, including statistical physics, computational science and public policy.

"From the perspective of statistical physics, we now plan to extend our theoretical model to more real-world scenarios and explain important phenomena with similar governing laws, for example, segregation and polarization," Piao added. "From the perspective of computational science, our team plans to do more research on complex systems for AI

and AI for social good. From the perspective of public policy, we plan to promote the design of corresponding public policies with our findings."

More information: Jinghua Piao et al, Human–AI adaptive dynamics drives the emergence of information cocoons, *Nature Machine Intelligence* (2023). DOI: 10.1038/s42256-023-00731-4.

© 2023 Science X Network