EDITORS
NORITA AHMAD American University of Sharjah; nahmad@aus.edu
PREETI CHAUHAN IEEE Reliability Society; preeti.chauhan@ieee.org

Algorithmic Nudge: An Approach to Designing Human-Centered Generative Artificial Intelligence

Donghee Shin, Texas Tech University
Norita Ahmad, American University of Sharjah

An algorithmic nudge is any form of choice architecture in artificial intelligence systems that changes users' behavior in a predictable way without forfeiting any options or significantly altering their technological choices.

HARNESSING NUDGE FRAMEWORKS TO DESIGN HUMAN-CENTERED GENERATIVE ALGORITHMS

A human-like artificial intelligence (AI) chatbot, such as the Chat Generative Pretrained Transformer (ChatGPT), a type of generative AI (GAI) application, has rapidly become a widely used tool for information search and generative services. ChatGPT is very versatile as it can answer questions and interact in a conversational manner with humans. This viral AI-powered chatbot can write essays, poems, theses, and even computer code. The dialogue format allows ChatGPT to answer follow-up questions, admit mistakes, challenge incorrect premises, and make appropriate suggestions.[1]

The human-centered GAI approach involves users throughout the algorithm development and testing processes, providing an effective experience between humans and the GAI. The need for a human-centered framework for AI has emerged to address

Digital Object Identifier 10.1109/MC.2023.3278156
Date of current version: 26 July 2023

the ethical, practical, and legal issues with GAI and make it sustainable so that it augments, empowers, and enriches human experiences instead of substituting human capacity.[2] The framework could lead to a fairer, more transparent, accountable, and explainable GAI that supports human values, preserves human rights, and promotes user control to steer future GAI in the right direction. Important questions include how algorithms fit within a social context, how they can enable meaningful control, and how users can manage algorithm systems effectively. The answers to these questions will guide the development of GAI systems that reshape the relationship between humans and machines. Meaningful human control can play a key role in paving a practical way to develop human-centered algorithms in GAI as well as in developing extended AI by providing theoretical underpinnings of ethical reflection.

> Because the results have natural language modality, users tend to trust these results, which have been algorithmically constructed for them.

HUMANNESS AND GAI

The strength of GAI is often described as its ability to produce "human-like" creative outputs, including images and text.[3] Current widespread debates focus on the potential impacts of these systems on existing systems of trust, evaluation, and creative value and broader social issues regarding potential bias. Two issues have been identified as being directly connected in human-centricity in GAI development. First, the datasets that are used to train GAI systems typically encode dominant cultural views, begging the question of which humans are being modeled by these systems and in what ways a normative view of humanness in AI reinforces discriminatory and hegemonic views.[4] The second issue involves the ways in which humanism has utilized a particular view of "human nature" to establish and maintain normative, moral, cultural, and legal claims that elevate some individuals to the status of being moral agents, while relegating those deemed to be nonhuman to instrumental roles.[5] This includes the long history of antihumanist and posthumanism thought to support mastery over others.[5] It is thus essential to examine how humanness is being defined and operationalized within AI. To advance our understanding of humanness in the context of GAI, we should focus on three aspects: 1) What kinds of data serve as the models for AI? 2) What human traits are considered appropriate and appropriable? 3) How are these traits operationalized within specific AI systems?

These questions are essentially related to algorithmic experience (AX), namely, how people experience AI, what people know about algorithms, and how these user perceptions influence their interactions with AI. AX can be a subset of user experience (UX), but AX is focused on the domain of AI as its features, and performances are often significantly different from those of other technological objects.[5] AX is a conceptual lens that considers the user's perspective to understand more effectively how users perceive these algorithms and how they experience AI overall. Despite GAI's holistic impact on reality, it remains to be defined how people experience or enjoy GAI, and how automated processes may advance their experience with algorithms. Algorithm services are conceived and designed to advance UX, but how users improve their experience through algorithms remains unanswered. The AX framework is an effective tool for humanizing AI. AX includes users' ability to control algorithmic decision making by increasing awareness of how the system works so that users can manage algorithmic bias and negative influence consciously.[4] The concept of meaningful user control over AI has been proposed as a key component of the AX, which can also address issues concerning fair, transparent, and accountable AI.

ALGORITHMIC NUDGING VIA GAI

The use of GAI was initially limited to research settings but has now extended to diverse domains and everyday scenarios.[6] As such, algorithmic nudging via GAI is becoming an effective practice. Nudge principles have been applied to algorithms so that they maneuver or manipulate search results through allusive search recommendations and targeted ads, steer users toward recommendations, and mix commercials with information in social media feeds.[3] For example, data from Facebook's feed exemplifies an algorithmic nudge as Facebook's algorithms empower these feeds to function as choice architects for the users. The algorithms curate news posts based on what it predicts will maximize a user's click-through rate, that is, the number of posts a user clicks on, given the number of posts that a user is shown. The algorithm selects a user's feed automatically and invisibly. By using algorithms that work invisibly, nudges can be personalized to individuals, and their efficacy can be traced and attuned as the algorithm improves from data generated from users' feedback.[4]

Platform providers use nudges within different dimensions of AI algorithms, such as news-recommending services, content-purchasing suggestions, and prescriptive decision-making tools. One example of algorithmic nudges in GAI is explanatory cues in generated information. Explanatory cues in GAI can influence users' behavior by gently steering them toward preferred choices through visual cues, push notifications, and alert alarms. With rapid developments in

AI and machine learning, algorithmic nudges have become much more influential than traditional nudges because algorithms can effectively identify target segments toward which to nudge users to make behavioral changes (Figure 1). Curating personalized models is a well-honed skill, with large amounts of online data about users' behavioral patterns available. Algorithmic nudging helps users evaluate and interact with algorithmic systems on the basis that informed judgments lead to wiser decisions. If used correctly, algorithmic nudges can help people assess how platforms, firms, and the government use these technologies, enabling them to advocate for responsible technology design and use that avoid biases and protect privacy at the same time.[7] Algorithmic nudges can involve meaningful efforts to empower more users to impact data flows and perceive whether or when they or others are being sidelined.

Another example of using algorithmic nudge in GAI can be misinformation alerts in the system. As a form of indirect intervention, accuracy alerts have been used to nudge people to discern truthfulness, which can improve their discernment of what to share on social media. While the effectiveness of such nudges has not been confirmed, some of them, such as Full Fact, PolitiFact, FeedReflect, and Associated Press, have been growing in popularity, becoming a major trend in social media and algorithmic platforms.[3]

Although ChatGPT has become increasingly popular and easily accessible, it is prone to inaccuracies and misinformation. Despite its popularity, users have discovered that ChatGPT is vulnerable to misinformation as it lacks a judgment of the boundary between fact and fiction.[8] With the emergence of GAI, the misinformation dilemma escalates. ChatGPT has attracted enormous traction for its ability to quickly and confidently present users with information.[9] However, this confidence and speed could pose a risk as GAI often generates false narratives. The efficient and quick information generation by GAI could have wide-ranging implications for the diffusion of misinformation online. Although GAI applications offer tremendous benefits, awareness of the potential dangers of generating and spreading misinformation is important. For example, ChatGPT cannot discern the boundary between fact and likelihood.[8] Thus far, GAI technologies are designed not for factual accuracy but for quick and eloquent conversations. Most GAI cannot verify whether the results they generate are accurate, and users are limited in their ability to check the source of the

> The key premise of human-centered GAI is to consider who will interact with AI instead of designing services only because they are technically feasible.

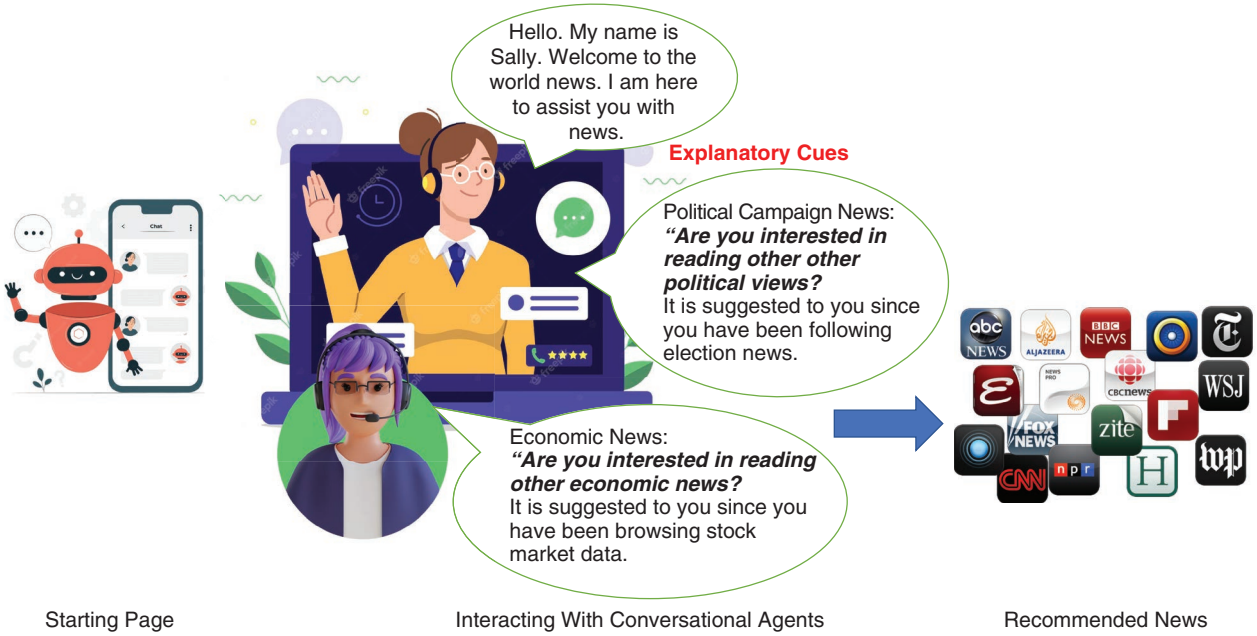

FIGURE 1. Explanatory cues as an algorithmic nudge.

information and its processing by the algorithms to generate content.[9] Owing to this inability, ChatGPT tends to fabricate information to satisfy users' inquiries with a quick response time. Misinformation can flow into AI models as well as from them, which means that some GAI will be vulnerable to "injection attacks," where malicious users input lies to the tools and train them; the tools in turn spread them. ChatGPT, for example, is susceptible to being used as a platform for misinformation generation and amplification.[1] In many cases, misinformation is presented as a fact and truth. Because the results have natural language modality, users tend to trust these results, which have been algorithmically constructed for them. This concern is related to a lack of transparency and explainability. One of the biggest challenges for ChatGPT-generated answers is explaining to users where the information comes from and how the answers are generated, constructed, and presented to them.

Although the consequences remain debatable and elusive, researchers have examined the relationship between AI and nudges, arguing that algorithmically personalized results can influence users and often lead to unintended consequences and unwanted habits.[10] The relationship between AI and nudges illustrates how personalized, tailored algorithms can use persuasion and psychometrics to affect individual and collective behavior in unintended ways.[10] Many recent discussions concur with concerns about algorithmic nudges' ability to predict user tastes.[10,11] The ongoing debate is on how algorithmic nudges are managed to lead to better consequences and whether regulations should be enforced on liability assigned for negative nudges that lead to bad influences. Considering that nudging intermediaries can amplify the severity of public-related harm, it has been contended that any rules and regulations should respond to unethical nudges with varying guidelines for deciding cases of intermediary liability. Thus, the question of how to design AI nudges that guide people toward better decisions ethically and responsibly remains open.

> Understanding the use of human-centricity is critical given the ways in which humanness has previously served to maintain technology-driven development.

Algorithmic nudges could be systematically deployed as a tool to enhance the never-ending cycle of extracting data, potentially manipulating consumer choice, and creating an optimum advertisement system controlled by AI. Algorithmic nudges can be used in the form of prebunking, debunking, and inoculation misinformation treatment. Appropriate nudges can combat misinformation by warning people about fake news before they see it. Also, nudges can be used in identifying disinformation after people have consumed information.

HUMAN-CENTERED APPROACH

The human-centered GAI approach involves users throughout algorithm development and testing processes, providing an effective experience between humans and AI (Figure 2).[12] The human-centered GAI system continuously advances user interaction while offering effective interaction between AI and humans.[10] The need for a human-centered AI framework has emerged to address the ethical, practical, and legal issues with AI and make it sustainable so that it augments, empowers, and enriches human experiences, instead of replacing human capacity.[1] The framework could lead to fairer, more transparent, accountable, and explainable AI that supports human values, preserves human rights, and promotes user control to steer future AI in the right direction. Important questions include how algorithms fit within a social context, how they can enable meaningful control, and how users can manage algorithm systems effectively. The answers to these questions will guide the development of AI systems that allow humans to see, perceive, create, and behave with confidence and trust. Meaningful human control will play a key role in paving a practical path to realizing meaningful human control

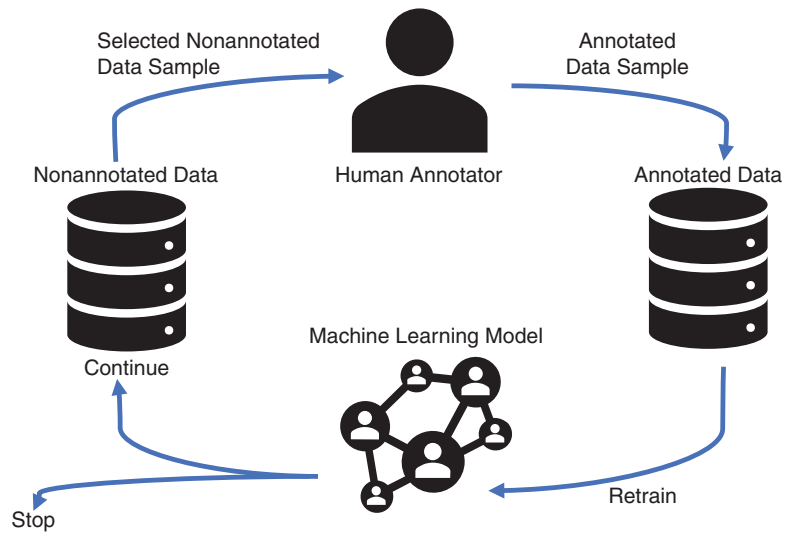

FIGURE 2. Human-centered AI.

over AI algorithms. Extended AI can be designed and should be developed in a human-centered and meaningfully controllable way to contribute to a fairer and more transparent design to forge key positive effects with clear accountability.

The key premise of human-centered GAI is to consider who will interact with AI instead of designing services only because they are technically feasible. The underlying assumption is that AI systems should be available and communicate in a way that normal nontechnical users can understand. Human-centered AI aims to develop AI so that it can understand how humans think, perceive, communicate, and interact, instead of compelling humans to learn how AI systems perform and function. This adds two important parameters to human-centered AI systems: 1) they should be able to understand humans, and 2) they should help humans trust them through fair, transparent, and explainable processes. These kinds of AI principles are mechanisms that make autonomous AI systems more sustainable because they will not commit common sense errors, infringe on human rights intentionally, or carelessly create situations that can lead to harm and conflict.[13]

Understanding the use of human-centricity is critical given the ways in which humanness has previously served to maintain technology-driven development. Algorithmic nudging via GAI is an emerging practice that deserves attention from industry and academia. With continuing advances in GAI and machine learning, algorithmic nudging is much more potent than its nonalgorithmic counterpart. Nudge principles have been applied to algorithms, and increasingly more algorithms are used in nudging. While convenient and useful, these nudges have elicited a series of ethical concerns about privacy, information disclosure, manipulation, and tweaking.

Practically, the nudge model provides useful suggestions for policymakers and AI industry on how to manage and operate accuracy nudges. Given the identified importance of accuracy nudges, AI firms and GAI providers should make fundamental changes to the social media ecosystem to redirect attention to accuracy systematically. In particular, given user attention's changeable nature, firms should figure out the interventions that would last long enough to influence user behavior on a long-term basis.

REFERENCES

1. G. Eysenbach, "The role of ChatGPT, generative language models, and artificial intelligence in medical education: A conversation with ChatGPT and a call for papers," *JMIR Med. Educ.*, vol. 9, no. 1, Mar. 2023, Art. no. e46885, doi: 10.2196/46885.
2. M. Jovanović and M. Campbell, "Generative artificial intelligence: Trends and prospects," *Computer*, vol. 55, no. 10, pp. 107–112, Oct. 2022, doi: 10.1109/MC.2022.3192720.
3. D. Shin, "The perception of humanness in conversational journalism: An algorithmic information-processing perspective," *New Media Soc.*, vol. 24, no. 12, pp. 2680–2704, Dec. 2022, doi: 10.1177/1461444821993801.
4. D. Shin, *Algorithms, Humans, and Interactions: How Do Algorithms Interact With People? Designing Meaningful AI Experiences*. New York, NY, USA: Routledge, 2023.
5. N. Mellamphy, "Re-thinking 'human-centric' AI: An introduction to posthumanist critique," *Europe Now*, vol. 45, Nov. 2021.
6. Ö. Aydın and E. Karaarslan. *Is ChatGPT Leading Generative AI? What is Beyond Expectations?* (Jan. 2023). SSRN. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4341500
7. S. Akter et al., "Algorithmic bias in data-driven innovation in the age of AI," *Int. J. Inf. Manage.*, vol. 60, Oct. 2021, Art. no. 102387, doi: 10.1016/j.ijinfomgt.2021.102387.
8. E. Van Dis, J. Bollen, W. Zuidema, R. Rooij, and C. Bockting, "ChatGPT: Five priorities for research," *Nature*, vol. 614, no. 7947, pp. 224–226, Feb. 2023, doi: 10.1038/d41586-023-00288-7.
9. P. Korzynski et al., "Generative artificial intelligence as a new context for management theories: Analysis of ChatGPT," *Central Eur. Manage. J.*, early access, 2023, doi: 10.1108/CEMJ-02-2023-0091.
10. C. Burr, N. Cristianini, and J. Ladyman, "An analysis of the interaction between intelligent software agents and human users," *Minds Mach.*, vol. 28, no. 4, pp. 735–774, Dec. 2018, doi: 10.1007/s11023-018-9479-0.
11. Z. Tufekci, *Twitter and Tear Gas*. New Haven, CT, USA: Yale Univ. Press, 2017.
12. B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe and trustworthy," *Int. J. Human-Comput. Interact.*, vol. 36, no. 6, pp. 495–504, 2020, doi: 10.1080/10447318.2020.1741118.
13. A. Rai, "Explainable AI: From black box to glass box," *J. Acad. Marketing Sci.*, vol. 48, no. 1, pp. 137–141, Jan. 2020, doi: 10.1007/s11747-019-00710-5.

DONGHEE SHIN is a professor of digital media and professional communication at the College of Media and Communication at Texas Tech University, Lubbock, TX 79409 USA. Contact him at: don.h.shin@ttu.edu.

NORITA AHMAD is a professor of management information systems at the American University of Sharjah, Sharjah, United Arab Emirates. She is a Member of IEEE. Contact her at nahmad@aus.edu.