Khi đưa trẻ chơi cờ vua: Cảm xúc của trí tuệ nhân tạo

14:27 | 07/07/2021

Cuốn sách "Affect and Artificial Intelligence" của Wilson xứng đáng nhận được sự quan tâm của các độc giả ngành khoa học máy tính và quan trọng hơn là các học giả của ngành khoa học nhân văn.

Trong cuốn sách "*Affect and Artificial Intelligence*" (Lược dịch: Xúc cảm và Trí tuệ nhân tạo), GS. Elizabeth A. Wilson (Đại học Emory, Mỹ) đã phân tích về khía cạnh cảm xúc liên quan tới những mạng lưới mà con người đã dự đoán và tạo thành các thiết bị tính toán ở giữa thế kỉ XX. Bằng cách nghiên cứu về những yếu tố cảm xúc (sự tò mò, khinh thường, tức giận, buồn bã...) được thể hiện trong những công trình nghiên cứu của ba nhà khoa học và toán học xuất sắc thời kì đầu của máy tính gồm Alan Turing (1912-1954), Walter Pitts (1912-1969) và Joseph Weizenbaum (1923-2008), Wilson đã lập luận cảm xúc là một phần nền tảng để tạo nên những cỗ máy thông minh. Qua đó, cuốn sách góp phần giúp các đối tượng tính toán như máy tính, thuật toán... trở nên gần gũi hơn đối với các học giả của ngành nhân văn.

Krzywoszynska (2012) đã ngắn gọn chỉ lập luận thú vị nhất trong cuốn sách, đó là: Bằng cách nào mà khái niệm AI thường luôn được hiểu với hàm ý vô cảm. Ngay từ những phần đầu, Wilson đã đưa ra sự đối chiếu giữa hai mô hình tư duy về AI bằng cách trích dẫn từ bài viết nổi tiếng Computational Machinery and Intelligence (*Máy móc tính toán và Trí tuệ*) của Alan Turing (1950): mô hình chơi cờ (*a chess-playing model*) và mô hình đứa trẻ (*a child-like model*). Wilson cho thấy rằng, hầu hết các bài bình luận sơ khai về AI đều bị chi phối bởi mô hình chơi cờ, mô hình liên quan mật thiết đến sự trừu tượng, xử lý số và thao tác ký hiệu. Và đây là vì sự miễn cưỡng của chúng ta trong việc thừa nhận vai trò của cảm xúc đối với sự vận hành của những cỗ máy thông minh. Cũng chính sự miễn cưỡng này đã hạn chế tiềm năng phát triển của ngành trí tuệ nhân tạo. Thông qua dữ liệu tiểu sử, Wilson cho thấy các công trình của Alan Turing có dấu ấn của sự nhiệt tình quá mức, trong khi công trình của nhà toán học ở MIT, Walter Pitts, người đồng sáng tạo ra mạng lưới thần kinh (neural nets) mang dấu ấn của sự kiềm chế và tránh né cảm xúc. Trong khi đó, các công trình của Joseph Weizenbaum, cha đẻ của Eliza, một chương trình máy tính được lập trình để có khả năng đối thoại như những nhà trị liệu tâm lí, cho thấy sự phức tạp trong việc lượng hóa hay tự động cơ quá trình trị liệu tâm lí và tầm quan trọng của cảm xúc trong việc tạo nên khả năng đối thoại cho AI. Elizabeth Wilson đã truyền đạt thành công lập luận này đến với độc giả thông qua những dẫn chứng dẫn chứng tiểu sử (*biographical data*) đến từ cuộc đời của ba nhà tiên phong trong lĩnh vực máy tính. Đầu vậy, *Affect and Artificial Intelligence* cũng còn để ngỏ nhiều khía cạnh liên quan tới trí tuệ nhân tạo và cảm xúc.

Trước hết, Wilson đã bỏ lỡ cơ hội đề cập đến lĩnh vực điện toán cảm xúc (affective computing), lĩnh vực được Rosalind Picard phát minh vào cuối những năm 1990 (Picard, 2000). Lĩnh vực này được tạo ra nhằm dạy cho máy móc khả năng đọc, theo dõi, phân loại và thậm chí là thể hiện cảm xúc như con người. Có lẽ đó chính là dự án trực tiếp nhất nhằm tạo nên "khả năng cảm nhận" (*ability to feel*) cho một cỗ máy. Đến năm 2003, lĩnh vực này đạt được tỉ lệ chính xác là 81% khi nhận ra được tám cảm xúc cơ bản được quy định bởi Ekman (1999) (Picard, 2003). Trong những năm gần đây, các công ty sản xuất và phân phối các sản phẩm trí tuệ nhân tạo cảm xúc, dựa trên cách tiếp cận này, đã tuyên bố tỷ lệ nhận biết cảm xúc chính xác là 95%. Dù vậy, tuyên bố này đã vấp phải rất nhiều tranh luận do sự gia tăng của các lý thuyết mới và kết quả thực nghiệm về cảm xúc chống lại lý thuyết của Paul Ekman (Heaven, 2020). Dựa trên các lập luận của Wilson trong cuốn sách, có lẽ tác giả sẽ phản đối những tuyên ngôn của ngành điện toán cảm xúc.

Cuốn sách góp phần giúp các đối tượng tính toán như máy tính, thuật toán... gần gũi hơn đối với các học giả của ngành nhân văn

Các lý thuyết khác nhau về cảm xúc, về cách cảm xúc được hình thành, biểu đạt và suy luận, sẽ có những tác động khác nhau trong phương thức viết thuật toán

Rõ ràng, cho đến nay vẫn còn nhiều bất đồng về bản thể của cảm xúc: Cảm xúc là tự nhiên và đồng nhất, bất kể văn hóa hay cảm xúc bị chi phối bởi xã hội? Các lý thuyết khác nhau về cảm xúc, về cách cảm xúc được hình thành, biểu đạt và suy luận, sẽ có những tác động khác nhau trong phương thức viết thuật toán và xây dựng bộ dữ liệu giúp máy móc nhận biết và biểu hiện cảm xúc. Phương pháp tiếp cận của điện toán cảm xúc có thể đạt được những thành công nhất định với sự bùng nổ của dữ liệu lớn và sức mạnh tính toán, tuy nhiên, nó thường đi kèm với định kiến của các nhà thiết kế thuật toán cũng như trong dữ liệu.

Có hai ví dụ thể hiện rõ ràng về vấn đề này. Một số nghiên cứu gần đây về độ chính xác của việc đọc cảm xúc của con người thuộc các chủng tộc khác nhau đã cho thấy rằng, các thuật toán hiện tại chỉ ra người da đen hay tức giận hơn (McStay, 2018; Rhue, 2019). Và trong trường hợp tập dữ liệu ban đầu bị chi phối bởi một giới tính, chẳng hạn trong lực lượng cảnh sát, việc sử dụng phương pháp AI hiện tại để theo dõi cảm xúc của các nữ cảnh sát có thể dẫn đến những lỗi vô ý không đáng có (Purdy, Zealley và Maseli, 2019).

Một vấn đề khác mà chúng tôi mong rằng, Wilson nên khám phá là sự tiếp biến văn hóa về tình cảm và cảm xúc. Tức là cách mọi người học và thay đổi những phản ứng cảm xúc khi tiếp cận văn hóa mới, cũng như cách điều đó tác động đến sự phát triển AI. Wilson có lẽ đã truyền đạt thành công rằng, tất cả những người tiên phong về AI đều nghĩ, chúng ta cần hiểu hơn về cảm xúc để xây dựng AI thông minh, hoặc có những cách mà cảm xúc gây nên ảnh hưởng tới công việc của họ. Tuy nhiên, cuốn sách chưa từng đề cập đến vấn đề thay đổi cảm xúc. Đây là một vấn đề quan trọng, bởi lẽ đời sống cảm xúc của chúng ta đang ngày càng bị ảnh hưởng từ sự gia tăng kết nối với truyền thông và các mạng xã hội, tạo nên sự xuất hiện của rất nhiều nền văn hóa và hệ tư tưởng và làm nảy sinh các phương thức biểu đạt cảm xúc mới. Vấn đề tiếp biến văn hóa này (Vuong et al., 2018a; Vuong et al., 2020; Vuong & Napier, 2015) sẽ tiếp tục xuất hiện, và phương pháp tiếp cận theo chủ nghĩa hành vi đang thịnh hành trong trí tuệ nhân tạo cảm xúc sẽ tạo nên cản trở cho sự phát triển của ngành.

Cuối cùng, có lẽ dữ liệu tiểu sử (*biographical data*) của cuốn sách có thể được hưởng lợi từ một phương pháp tổ chức dữ liệu tốt hơn (Vuong et al., 2018b). Ví dụ, quá trình tư duy và công trình của ba nhà tiên phong có thể được phân tích dựa theo ba góc độ: cảm xúc, trực giác và lý trí, trong đó trực giác là cầu nối từ cảm xúc đến lý trí. Trực giác là một tập hợp các giá trị cốt lõi cơ bản, các tiên đề hoặc các quy tắc chung, mà từ đó định hướng những hành vi, cảm xúc và nhận thức của chúng ta. Trực giác của chúng ta có thể được cập nhật và thay đổi thông qua rèn luyện và học hỏi. Wilson có thể phân tích sự cải biến trực giác của các nhà tiên phong về vai trò của cảm xúc trong AI qua những công trình khoa học, thí nghiệm và kinh nghiệm xây dựng các thuật toán. Ở đây cũng là một góc độ mà quan điểm về sự tiếp biến văn hóa của cảm xúc có thể sẽ đóng một vai trò nào đó.

Rõ ràng, việc tổ chức dữ liệu một cách trật tự có thể giúp định hình các mạch tường thuật trong cuốn sách của Wilson, cũng như giúp tác giả đưa ra được kết luận chung, điều mà nhiều nhà bình sách cho rằng, đã thiếu sót. Thậm chí, việc tổ chức dữ liệu tiểu sử tốt có thể giúp cải thiện khả năng tái thiết lập kết quả nghiên cứu trong các lĩnh vực nhân văn khác (Peels & Bouter, 2018; Vuong, 2020). Bất chấp những thiếu sót như vậy, chúng tôi đánh giá, cuốn sách của Wilson vẫn xứng đáng có được sự quan tâm của độc giả của lịch sử ngành khoa học máy tính và quan trọng hơn là các học giả của ngành khoa học nhân văn, những người muốn nghiên cứu cách mà cảm xúc đã tác động đến công việc của những nhà lý thuyết và kỹ sư AI tiên phong như thế nào.

References

- Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of Cognition and Emotion*. Sussex, UK: John Wiley & Sons.
- Heaven, D. (2020). Why faces don't always tell the truth about feelings. *Nature*, 578, 502-504.
- McStay, A. (2018). *Emotional AI: The rise of empathic media*. London: Sage.
- Peels, R., & Bouter, L. (2018). The possibility and desirability of replication in the humanities. *Palgrave Communications*, 4(1), 95.

Picard, R. W. (2000). Affective computing. Cambridge, Massachusetts: MIT Press.

Picard, R. W. (2003). Affective computing: challenges. International Journal of Human-Computer Studies, 59(1), 55-64.

Purdy, M., Zealley, J., & Maseli, O. (2019). The risks of using AI to interpret human emotions. Harvard Business Review. Retrieved from <https://hbr.org/2019/11/the-risks-of-using-ai-to-interpret-human-emotions>

Rhue, L. (2019). Anchored to Bias: How AI-Human Scoring Can Induce and Reduce Bias Due to the Anchoring Effect. Available at SSRN 3492129.

Vuong, Q. H. (2020). Reform retractions to make them more transparent. Nature, 582(7811), 149.

Vuong, Q. H., & Napier, N. K. (2015). Acculturation and global mindsponge: an emerging market perspective. International Journal of Intercultural Relations, 49, 354-367.

Vuong, Q.-H., Bui, Q.-K., La, V.-P., et al. (2018a). Cultural additivity: behavioural insights from the interaction of Confucianism, Buddhism and Taoism in folktales. Palgrave Communications, 4(1), 143.

Vuong, Q.-H., Ho, M.-T., Nguyen, H.-K. T., et al. (2020). On how religions could accidentally incite lies and violence: folktales as a cultural transmitter. Palgrave Communications, 6(1), 82.

Vuong, Q.-H., La, V.-P., Vuong, T.-T., Ho, M.-T., et al. (2018b). An open database of productivity in Vietnam's social sciences and humanities for public use. Scientific Data, 5(1), 180188.

Wilson, E. A. (2011). Affect and artificial intelligence. Washington: University of Washington Press.

Hồ Mạnh Tùng,

Trung tâm Nghiên cứu Xã hội liên ngành, Đại học Phenikaa, Hà Nội

Viện Triết học, Viện Hàn lâm Khoa học Xã hội, Hà Nội, Việt Nam

Trần Đức Hưng Long,

Ritsumeikan Asia Pacific University, Oita, Beppu, Nhật Bản

URL: <https://kinhtevadubao.vn/khi-dua-tre-choi-co-vua-cam-xuc-cu-a-tri-tue-nhan-tao-18222.html>

© Kinh tế và Dự báo - Bộ Kế hoạch và Đầu tư