Kinh tế và Dư báo

Online
CƠ QUAN CỦA BỘ KẾ HOẠCH VÀ ĐẦU TƯ

Luận điểm về AI nhìn từ “Cỗ máy xúc cảm” của DeLancey

11:17 | 13/07/2021

Cỗ máy xúc cảm: Cảm xúc hé lộ điều gì về tâm trí và trí tuệ nhân tạo (Tạm dịch từ tên sách: *The Passionate engines: What emotions reveal about the mind and artificial intelligence*) của Craig Delaney (Đại học Oswego, Hoa Kỳ) đã khắc họa 1 bức tranh tổng thể về ngụ ý của những cảm xúc cơ bản đối với các vấn đề trọng tâm của triết học tinh...

Cuốn sách bàn về năm vấn đề chính: Học thuyết về cơ chế của cảm xúc, tính chủ đích, tâm thức (*consciousness*) và trí tuệ nhân tạo (AI). Kể từ lần xuất bản đầu tiên vào năm 2001, cuốn sách đã nhận được nhiều đánh giá, chẳng hạn từ Graham (2002), Radden (2003) và Scarantino (2004). Tất cả đều ca ngợi cuốn sách, vì những đóng góp của nó cho hệ thống nghiên cứu triết học xoay quanh cảm xúc cùng với lối hành văn rõ ràng và chín chu.

Bài viết này hướng tới nhận xét ngắn gọn về các ý chính trong cuốn sách và sau đó tập trung nhiều hơn vào các luận điểm về AI mà vốn chưa được đánh giá chi tiết. Trước tiên phải công nhận rằng, vì thời gian xuất bản của cuốn sách mà tác giả không được hưởng lợi từ việc quan sát những phát triển mới nhất của khoa học thần kinh về cảm xúc (Barrett, 2017) và điện toán cảm xúc (*affective computing*), tức là những nghiên cứu AI về cách nhận dạng cảm xúc của con người (McStay, 2018; Picard, 2000). Dù vậy, nhiều quan điểm trong cuốn sách có liên quan mật thiết với trí tuệ nhân tạo cảm xúc (*Emotional AI*) và hoàn toàn có thể giúp chúng ta hiểu biết thêm về ngành khoa học đang có những phát triển ứng dụng sôi động này.

Trước tiên, chiến lược chung của DeLancey là, con người có nhiều loại cảm xúc và mỗi cảm xúc đều được biểu hiện theo một cách riêng biệt, ta vẫn có thể xác định được sự tồn tại của những cảm xúc cơ bản, chẳng hạn như sợ hãi hay tức giận. Theo ông, một cảm xúc cơ bản là cảm xúc mà xuất hiện ở nhiều nền văn hóa khác nhau, có hành vi tương quan giống nhau, có xu hướng định hướng hành động, tạo động lực và đường như, chúng tồn tại để giúp con người ta thích nghi trong quá trình tiến hóa. Ông đã sắp xếp lại những quan điểm của mình về cảm xúc và đặt tên là lý thuyết về cơ chế cảm xúc (*affect program theory*), thuộc về trường phái bản chất luận về cảm xúc (*essentialist view of emotion*) (Barrett, 2017). Lý thuyết này được các học giả như Paul Ekman (1999), với lý thuyết cảm xúc phổ quát là đại diện tiêu biểu. Dưới lăng kính này, DeLancey tiếp tục chỉ trích một số triết lý về cảm xúc: tư duy luận (*cognitivism*), học thuyết coi cảm xúc là một dạng nhận định về giá trị; thuyết cho rằng “một số trạng thái tinh thần lệ thuộc vào lập trường quan điểm của người quan sát” (tr.50); thuyết kiến tạo xã hội (*social constructionism*), trong đó trạng thái cảm xúc là sản phẩm của các chuẩn mực văn hóa, nó không có hoặc ít liên quan đến các vấn đề sinh học của chúng ta.

Sau khi bàn về các triết lý đương đại về cảm xúc, DeLancey tập trung vào “lập trình cảm xúc (*affect engineering*)”, với nỗ lực có thể lập trình được cảm xúc của con người vào trong các hệ thống AI. Ông cho rằng, việc lập trình AI để có thể biểu lộ và cảm nhận cảm xúc là một thử nghiệm mẫu mực cho bất kỳ lý thuyết cảm xúc nào. Ông phân biệt rõ “lập trình cảm xúc nông” và “kỹ thuật cảm xúc sâu”. Lập trình cảm xúc nông sẽ không thực sự cho AI cảm xúc thực sự, nó chỉ tạo ra vẻ bề ngoài như thể có khả năng đọc hay hiển thị cảm xúc cho AI. Trái lại, lập trình cảm xúc sâu cho rằng, thực sự tạo ra khả năng có cảm xúc cho các hệ thống AI mới là một chiến lược cốt lõi (tr.204). DeLancey lập luận rằng, việc thực hiện lập trình cảm xúc sâu có thể rất có lợi cho việc nghiên cứu AI.

Tiếp tới, ông đã đưa ra những lập luận dựa theo sinh học (*biomorphic argument*) và cho rằng, quá trình tiến hóa sinh học đã tạo ra những sinh vật tự chủ tốt hơn nhiều so với trong phòng thí nghiệm AI và cảm xúc đóng một vai trò quan trọng trong các hành vi tự chủ đó. Do đó, nếu tập trung hơn vào sinh học thì chúng ta có thể cải thiện được việc lập trình sự chủ đích (*intentionality*) cho các hệ thống AI. DeLancey đã đưa ra 6 bài học để tạo nên cỗ máy cảm xúc và đối chiếu cỗ máy này với những mối quan tâm hiện tại của ngành nghiên cứu AI, ví như việc xử lý ký hiệu, thao tác, làm việc với các con số... Trước hết, việc nghiên cứu AI nên tập trung vào các chuyển động trước, rồi tới cảm xúc, cũng như những thứ hữu hình trước những thứ trừu tượng. Thứ hai, trong các hệ thống sinh học, những cảm xúc là thành phần quan trọng của quá trình đưa ra quyết định. Thứ ba, hệ thống cảm xúc bản nhận thức của chúng ta có thể chính xác hơn các hệ thống thuần nhận thức khác.

Thứ tư, sự hiện thân (*embodiment*) nên được cân nhắc một cách kỹ lưỡng hơn trong việc nghiên cứu AI. Điều thứ năm và thứ sáu chỉ ra tầm quan trọng của quá trình xử lý song song (*parallel processing*) của các hệ thống cảm xúc bản nhận

Nhiều quan điểm trong cuốn sách có liên quan mật thiết với trí tuệ nhân tạo cảm xúc

“Tác phẩm của DeLancey thực sự là cầu nối giữa các học giả khoa học nhân văn và cộng đồng kỹ thuật, giúp cho các hệ thống AI trở nên minh bạch hơn so với trước đây.”

thức và các ứng dụng theo chiều sâu của các quá trình này để xây dựng hành vi thông minh và có chủ đích. Những bài học này xuất phát từ chính trực giác của DeLancey và ông cho rằng, những cảm xúc cơ bản vốn có tính thúc đẩy và định hướng các hành động, những thứ đã tiến hóa và thích ứng trong lịch sử sinh học của chúng ta.

Về cơ bản, những quan điểm của DeLancey về tiềm năng của việc lấy cảm hứng từ sinh học để nghiên cứu AI và cảm xúc là nền tảng quan trọng cho trí tuệ của chúng ta là rất đáng tán đồng. Tuy nhiên, vẫn còn đó những thiếu sót. Đầu tiên, DeLancey đã không giải thích việc tái tạo bản thể cảm xúc trên ngôn ngữ lập trình (*instantiate*) là như thế nào. Liệu việc đó có đòi hỏi máy móc cũng phải có cảm giác không? Với con người và các loài động vật khác, những sinh vật tự nhiên, cảm xúc đóng vai trò quan trọng trong việc thực hiện các hành vi có chủ đích. Tuy nhiên, nếu không có những lý giải đột phá về ý thức nhân tạo (*artificial consciousness*), ta không thể cứ quy chụp rằng, cảm xúc cũng quan trọng đối với các hệ thống máy móc, vốn là những thứ không thuộc về tự nhiên. Hơn nữa, ai có thể chắc rằng cảm xúc của máy móc sẽ giống như những cảm xúc của con người?

Một điều nữa có thể làm cuộc thảo luận của DeLancey phong phú hơn là vai trò của cảm xúc đối với những suy luận phản thực tại (*counterfactual*) của chúng ta. Nhà thống kê Judea Pearl, “cha đẻ” của lưới Bayesian - một hướng tiếp cận thông dụng trong việc xây dựng AI (Vuong et al., 2020), lập luận rằng, nghiên cứu AI chưa có nhiều tiến bộ về lý thuyết do họ đã không quan tâm tới việc suy luận phản đề một cách nghiêm túc (Bereinboim & Pearl, 2016; Pearl, 2019). Nghiên cứu về cách trẻ nhỏ hình thành các mô hình và hiểu biết sơ khai về thế giới cho thấy khả năng suy luận phản thực tại (*counterfactual reasoning*) đóng vai trò quan trọng từ rất sớm (Gopnik, 2012). Như vậy, để vượt ra khỏi phương pháp tiếp cận máy tự học dựa trên các mô hình về sự tương quan của phản ứng và kích thích vốn còn sơ khai (Deustsch, 2020), cách lập trình suy luận phản thực tại hiện đang được các nhà nghiên cứu AI đề xuất và tìm cách triển khai. Dù vấn đề về suy luận phản thực tại có thể đã là một trong những bài học mà DeLancey đưa ra trong sách của mình, cần phải thừa nhận rằng những lập luận dựa trên sinh học của ông vẫn là những đóng góp quan trọng cho các luận thuyết về việc xây dựng các máy móc và phần mềm thông minh có khả năng tự chủ.

Trong thời đại mà máy móc và các thuật toán chịu trách nhiệm đưa ra các quyết định ngày càng nhiều, việc nghiên cứu sơ đồ thiết kế của những thuật toán này - lĩnh vực “Trí tuệ nhân tạo minh bạch” (*transparent AI*), không chỉ giúp cải thiện chất lượng các quyết định của chúng ta mà còn là vấn đề niềm tin vào khoa học và công nghệ (Vuong, 2018; Vuong, 2020). Về phương diện đó, tác phẩm của DeLancey thực sự là cầu nối giữa các học giả khoa học nhân văn và cộng đồng kỹ thuật, giúp cho các hệ thống AI trở nên minh bạch hơn so với trước đây.

References

- Bereinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27), 7345-7352.
- Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. London: Houghton Mifflin Harcourt.
- DeLancey, C. (2001). *Passionate engines: What emotions reveal about the mind and artificial intelligence*. Oxford: Oxford University Press.
- Deutsch, D. (2020). *Beyond reward and punishment*. Brockman, J. (Ed.). *Possible Minds: Twenty-Five Ways of Looking at AI*, 113-124. Penguin Books, UK.
- Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of Cognition and Emotion*. Sussex, UK: John Wiley & Sons.
- Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, 337(6102), 1623-1627.
- Graham, G. (2002). Review of *Passionate Engines: What Emotions Reveal about Mind and Artificial Intelligence* by Craig DeLancey. *Notre Dame Philosophical Reviews*. Retrieved from <https://ndpr.nd.edu/news/passionate-engines-what-emotions-reveal-about-mind-and-artificial-intelligence/>
- McStay, A. (2018). *Emotional AI: The rise of empathic media*. London: Sage.
- Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3), 54-60.
- Picard, R. W. (2000). *Affective computing*. Cambridge, Massachusetts: MIT Press.
- Radden, J. (2003). Review of “Passionate engines: what emotions reveal about mind and artificial intelligence” by Craig DeLancey. *Consciousness & Emotion*, 4(1), 143-148. doi:<https://doi.org/10.1075/ce.4.1.13rad>
- Scarantino, A. (2004). Craig DeLancey: *Passionate Engines: What Emotions Reveal about the Mind and Artificial Intelligence*. *Philosophy of Science*, 71(2), 227-230. doi:10.1086/381422
- Schneider, S. (2020). *Artificial You: AI and the future of your mind*. Princeton: Princeton University Press.
- Vuong, Q.-H. (2018). The (ir)rational consideration of the cost of science in transition economies. *Nature Human Behaviour*, 2(1), 5-5. doi:10.1038/s41562-017-0281-4
- Vuong, Q. H. (2020). Reform retractions to make them more transparent. *Nature*, 582(149). doi:<https://doi.org/10.1038/d41586-020-01694-x>
- 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. doi:10.1057/s41599-020-0442-3.

Hồ Mạnh Tùng, Viện Triết học, Viện Hàn lâm Khoa học Xã hội Việt Nam

Centre for Interdisciplinary Social Research, Phenikaa University, Hanoi

Nguyễn Quốc Hùng, Đại học Ritsumeikan Asia Pacific University, Beppu, Oita, Nhật Bản

Hồ Mạnh Tùng, Đại học Ritsumeikan Asia Pacific, Beppu, Oita, Nhật Bản

URL: <https://kinhtevadubao.vn/luan-diem-ve-ai-nhin-tu-co-may-xuc-cam-cua-delancey-18311.html>

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